Volume 48 Number 2 June 2024 ISSN 0350-5596

Informatica

An International Journal of Computing and Informatics

Special Issue:

Deep Learning Assisted Intelligent Human Computer Interaction for Next Generation Internet Applications

Guest Editors:

Al Mehedi Hasan,
Yong Seok Hwang,
Jungpil Shin

Editorial Boards

Informatica is a journal primarily covering intelligent systems in the European computer science, informatics and cognitive community; scientific and educational as well as technical, commercial and industrial. Its basic aim is to enhance communications between different European structures on the basis of equal rights and international refereeing. It publishes scientific papers accepted by at least two referees outside the author's country. In addition, it contains information about conferences, opinions, critical examinations of existing publications and news. Finally, major practical achievements and innovations in the computer and information industry are presented through commercial publications as well as through independent evaluations.

Editing and refereeing are distributed. Each editor from the Editorial Board can conduct the refereeing process by appointing two new referees or referees from the Board of Referees or Editorial Board. Referees should not be from the author's country. If new referees are appointed, their names will appear in the list of referees. Each paper bears the name of the editor who appointed the referees. Each editor can propose new members for the Editorial Board or referees. Editors and referees inactive for a longer period can be automatically replaced. Changes in the Editorial Board are confirmed by the Executive Editors.

The coordination necessary is made through the Executive Editors who examine the reviews, sort the accepted articles and maintain appropriate international distribution. The Executive Board is appointed by the Society Informatika. Informatica is partially supported by the Slovenian Ministry of Higher Education, Science and Technology.

Each author is guaranteed to receive the reviews of his article. When accepted, publication in Informatica is guaranteed in less than one year after the Executive Editors receive the corrected version of the article.

Guest Editorial Preface

Deep Learning Assisted Intelligent Human Computer Interaction for Next Generation Internet Applications

Human-computer interaction (HCI) is a field that focuses on how people use technology, focusing on creating user-friendly, interactive systems tailored to their needs. One of the main challenges in HCI is creating intuitive and natural user interfaces, especially for non-technical users. As technology becomes more widespread, there is a growing need to create interfaces that are usable by individuals with various skills and disabilities. Designing interfaces that enable multitasking and multiple interaction modalities is another challenge. HCI also needs to address security and privacy concerns. Deep learning can be used to enhance interface usability by creating intelligent interfaces that can understand and respond to users in a natural and intuitive way. One of the main advantages of intelligent HCI with deep learning assistance is the development of new interaction modalities, such as voice, touch, and gesture, which can be recognized and responded to by the interface. This makes the interface more versatile and easier to use. Another area of research in deep learning-assisted intelligent HCI is the development of new security and privacy methods. By encrypting and safeguarding user data, deep learning can be used to develop interfaces that can recognize and respond to potential security concerns, such as hacking attempts. After a careful review process, a total of five papers were accepted for publication, and their major contributions are highlighted below.

In the first article is entitled "Vulnerability Detection in Cyber-Physical system using machine learning," the author proposed the cyber-physical system is a specific type of IOT communication environment that deals with communication through innovative healthcare devices. The intelligent Health care industry utilizes wireless medical sensors to gather patient health information and send it to a distant server for diagnosis or treatment. A method for intelligent threat recognition based on machine learning that enables run-time risk assessment for better situational awareness in CPS security monitoring. When used in industrial reference applications, the model creates a safe environment where the patient is only made aware of risks when categorization optimism exceeds a specific limit, minimizing security managers' pressure and efficiently assisting their choices.

In the second article is entitled "Cardiacnet: Cardiac Arrhythmia Detection and Classification Using Unsupervised Learning-Based Optimal Feature Selection with Custom CNN Model," the author describes the irregularity in the heartbeats caused cardiac arrhythmia, which resulted in serious health problems. This cardiac arrhythmia is monitored by electrocardiogram (ECG) signals. An accurate and timely analysis of ECG data can prevent serious health problems. CardiacNet, which is an AI tool for identifying and classifying cardiac arrhythmia from MIT-BIH based dataset. Unsupervised machine learning algorithm-based principal component analysis (UML-PCA) is used to extract the features of the pre-processed dataset.

In the third article is entitled "Parkinson Net: Convolutional Neural Network model for Parkinson Disease Detection from image and voice data," this article demonstrates the Parkinson's disease is a critical dopaminergic neuron problem that causes brain disorders. The early prediction of PD can save human lives. CAD with artificial intelligence models can predict PD in a quick time as compared to manual prediction. ParkinsonNet focused on the implementation of a deep learning mechanism for PD identification from both voice and image datasets. The simulation results show that the proposed ParkinsonNet achieved 99.67% accuracy on image data and 98.21% accuracy on voice data.

In the fourth article is entitled "DRG-Net: Diabetic Retinopathy Grading Network using Graph Learning with Extreme Gradient Boosting classifier," the author proposed the Diabetic retinopathy is a leading cause of blindness that occurs in different age groups. The early detection of DR can save millions of people from blindness issues. The manual analysis of DR requires more processing time and experienced doctors. CAD based artificial intelligence models have been developed for an early DR prediction. We have implemented the DR grading network using graph learning properties, the synthesis minority over-sampling technique (SMOTE) is applied to the EyePACS and Messidor dataset to balance the instances of each DR class into uniform levels.

In the final article is entitled "ELM-Based imbalanced Data Classification-A Review," the author describes the imbalance issues occur in Machine Learning when there is high distortion in the class distributions. A great challenging task in ML is the imbalance of data classification. The most prominently adopted technique to deal with data having an imbalance class distribution is the Extreme Learning Machine (ELM). Unwanted class boundaries as of data with unbalanced classes may be learned by ELM similar to various other classification algorithms. An augmentation in the ELM framework is done for efficient imbalanced classification.

We would like to thank all the authors for their valuable contributions. A special thanks to the reviewers for their timely comments and suggestions on the research articles. We hope this special issue will add significant value to the research community.

Guest Editors

Al Mehedi Hasan
(mehedi_ru@yahoo.com), Dept. of Computer Science and Engineering
Rajshahi University of Engineering and Technology, Bangladesh

Yong Seok Hwang
(thestone@kw.ac.kr), Holodigilog Human Media Research Center (HoloDigilog), Nano Device Application Center (NDAC)
Kwangwoon University, Seoul, Korea

Jungpil Shin
(jpshin.uoa@gmail.com), School of Computer Science and Engineering
The University of Aizu, Aizuwakamatsu, Japan

Cardiacnet: Cardiac Arrhythmia Detection and Classification Using Unsupervised Learning Based Optimal Feature Selection with Custom CNN Model

Kalyanapu Srinivas[1], Vijayalakshmi Ch[2], Subba Reddy Borra[3], K. Srujan Raju[4], Ganga Rama Koteswara Rao[5], K. N. V. Satyanarayana[6], Pala Mahesh Kumar[7*]
[1]Head of the Department, Department of Artificial Intelligence and Data Science, Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh, India
[2]Assistant Professor, Department of Electronics and Communication Engineering, St. Peter's Engineering College, Maisammaguda, Hyderabad, Telangana
[3]Professor and Head, Department of Information Technology, Malla Reddy Engineering College for Women (A), Maisammaguda, Hyderabad, Telangana
[4]Professor and Head, Department of Computer Science and Engineering, CMR Technical Campus Kandlakoya, Medchal Road, Hyderabad, Telangana
[5]Professor, Department of Computer Science and Information Technology, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India
[6]Assistant Professor, Department of Electronics and Communication Engineering, Sagi Rama Krishnam Raju Engineering College, Bhimavaram, Andhra Pradesh, India
[7]AI Engineer, SAK Informatics, Hyderabad, Telangana, India
Email: mahesh@sakinformatics.com
*Corresponding author

Keywords: ECG dataset, cardiac arrhythmia, convolutional neural network, improved harris hawk's optimization, unsupervised machine learning

Received: August 6, 2023

The irregularity in the heartbeats caused cardiac arrhythmia, which resulted in serious health problems. This cardiac arrhythmia is monitored by electrocardiogram (ECG) signals. As a result, an accurate and timely analysis of ECG data can prevent serious health problems. However, the conventional manual prediction systems and artificial intelligence (AI) methods failed to detect cardiac arrhythmia because they failed to extract the deep salient features from the ECG dataset. So, this research work implements a model named CardiacNet, which is used to identify and classify cardiac arrhythmias from an MIT-BIH-based dataset. Initially, the pre-processing operation is performed to remove the non-linearities from the dataset. Then, unsupervised machine learning algorithm-based principal component analysis (UML-PCA) is used to extract the features of the pre-processed dataset. Further, the optimal feature selection operation is carried out using improved Harris Hawk's optimization (IHHO), which is a naturally inspired model. Moreover, a customized convolutional neural network (CCNN) model performs the classification of various cardiac arrhythmia diseases using IHHO features. The simulation results show that the proposed CardiacNet resulted in an accuracy of 97.57%, sensitivity of 98.29%, specificity of 97.97%, F-measure of 97.40%, precision of 98.66%, Matthew's correlation coefficient (MCC) of 98.17%, dice of 98.96%, and Jaccard of 97.12%. The performance comparisons show that the proposed CardiacNet resulted in improved metrics over all existing methods.

Povzetek: Razvit je sistem CardiacNet za učenje brez nadzora za optimalno izbiro lastnosti s prilagojenim CNN modelom za natančno zaznavanje in klasifikacijo srčnih aritmij iz podatkov EKG.

1 Introduction

Abnormal heart rhythm is known as cardiac arrhythmia. According to the World Health Organization, cardiovascular diseases claim over 17 million lives each year [1]. About 31% of all fatalities may be attributed to this. The American Heart Association reports that heart disease is responsible for one in every three fatalities in the United States. Every year, cardiovascular disease kills more individuals than all forms of cancer and chronic lower respiratory disorders combined [2]. The 2022 research indicated that 2–4% of people in North America and European nations were living with AF. In adults, a heart rate of more than 100 beats per minute is considered tachycardia, whereas a heart rate of fewer than 60 beats per minute is considered bradycardia (less than 60 beats per minute). When the heart begins to contract before it should, this is called a premature contraction. An erratic heartbeat is also known as fluttering or fibrillation. Cardiac arrhythmias are classified not just by the patient's

heart rate but also by several other factors. In addition to being utilized as a diagnostic criterion [3], the location of the irregular heart rate may also be used to classify the condition. In most cases, problems with the atrioventricular node are the root cause of arrhythmias in the heart's atrium. The atrioventricular node is situated in the heart, halfway between the two upper chambers (the atria) and the two lower chambers (the ventricles) [4]. During cardiac contraction, blood is pushed from the atria to the ventricles. Some examples of atrial arrhythmias include atrial fibrillation (AF), atrial flutter, atrial tachycardia, premature atrial contractions, and sinus bradycardia. The AF and atrial flutter are types of arrhythmias that, if left untreated, may be life-threatening. AF is a cardiac rhythm disorder characterized by abnormally fast and disorganized atrial contractions and electrical impulses that originate from places other than the sino-atrial node [5].

This causes an atrial rhythm that is fast and irregular. As a result of the atria's wall fibrillation, the heart is unable to pump blood properly to the ventricles. AF may lead to several potentially deadly consequences, including stroke and heart failure. AF has been connected to a wide variety of medical conditions, including hypertension, an overactive thyroid, coronary heart disease, and rheumatic heart disease. Atrial flutter, like AF [6], may have serious consequences. Atrial flutter is characterized by the fast and regular movement of cardiac electrical impulses throughout the atria, in contrast to the disorganized passage of electrical impulses seen in AF. Premature heartbeats, or ventricular arrhythmias, may originate from an ectopic ventricular focus. Ventricular fibrillation, ventricular tachycardia, and premature ventricular contractions are all examples of ventricular arrhythmias. It's important to note that certain forms of arrhythmia cause no obvious symptoms and don't put the patient's life at risk [7]. However, there are forms of arrhythmia that show no symptoms but may lead to serious complications, including a clot, a stroke, heart failure, or even sudden cardiac death. Arrhythmia is a disorder that may arise when the heart's electrical impulses, which serve to coordinate beats, are not functioning properly [8]. The initial stages in making a diagnosis of this illness are to interpret the ECG and confirm that the ECG shows no symptoms of cardiac arrhythmia [9]. To alter and analyze cardiac arrhythmia ECG data, methods including feature extraction, engineering, pattern analysis, and deep learning are utilized, which increases the diagnostic precision and comprehension of arrhythmia situations. Through a variety of techniques, such as visual inspection, automated pattern recognition, and deep learning, arrhythmia diseases may be identified using amplitude changes in ECG readings, increasing the accuracy and efficiency of identifying cardiac abnormalities. Using ECG data from the MIT-BIH database and another database, several researchers were able to successfully categorize all four types of arrhythmias. The classification of cardiac arrhythmias has received a lot of attention, with ECG data serving as the gold standard [10]. Myocardial infarction is one of the most prevalent and dangerous forms of cardiac arrhythmia documented so far. When the regular beat of the heart is disturbed, a diagnosis of cardiac arrhythmia is made. However, the heart itself is the site of the abnormalities that give rise to this illness. Heart abnormalities lead to abnormal activation, depolarization, and repolarization. The ECG will show a shift in waveform reflective of these alterations. The waveform of an ECG will change depending on the kind of cardiac arrhythmia that is being suspected. So, the conventional AI [11], machine learning [12], and deep learning [13] methods failed to result in better sub-class classification. So, the novel contributions of this work are illustrated as follows:

- Implementation of CardiacNet for classification of Sinus Bradycardia, Right Bundle Branch Block, Old Anterior Myocardial Infarction, Ischemic Changes, Coronary Artery Diseases, and the Normal Heart from the MIT-BIH dataset
- Adoption of both feature extraction and optimal feature selection methods such as UML-PCA and IHHO is necessary for the effective analysis of ECG data.
- The development of a CCNN model for multiple classification classes from IHHO features, which also performs the prediction of disease from test data, Improved Harris Hawk's Optimisation features are used by the CNN model for multiple-class arrhythmia grouping, which calls for the gathering of the data, feature extraction, model design, training, assessment, fine-tuning, installation, and analysis.

The rest of the paper is organized as follows: Section 2 contains the literature survey with existing drawbacks. Section 3 contains a detailed analysis of the proposed method with sub-block explanations. Section 4 contains the detailed simulation analysis. Section 5 contains the conclusion.

2 Literature survey

In [14], the authors implemented the T-wave integral and the total integral from a single cycle of normal and patient ECG data to detect and localize myocardial infarction (MI) in the left ventricle. Through identifying ischemia-related alterations and infarct patterns, the T-wave integral in electrocardiography aids in the assessment of repolarization anomalies in myocardial infarction patients. It aids in diagnosis and prognosis, assisting with therapy selection with objectivity. A whole ECG cycle was used to extract these features. The T-wave integral was used because of the significance of this property to the T-wave in the MI. T-wave integration in machine learning (ML) helps with biometric identity, research and development, ischemia risk assessment, ECG signal quality, therapeutic efficacy, and personalized medicine. The second part of this study considers the integral of a single ECG cycle [15]. This happens because the shape of the ECG signal is altered after an MI alters the total integral. By including additional variables in the classification process, this body of work may increase accuracy. It was shown that MIs may be identified with a high degree of accuracy using a

simple technique based on the extraction of just two characteristics from a regular ECG. The diagnostic properties of multiscale wavelet energies and eigenvalues of multiscale covariance matrices were investigated by [16]. Classifiers can take many forms; the K-nearest neighbor technique, and support vector machines (SVM) with linear and RBF kernels [17], to name a few. The diagnostic ECG database provides the data utilized in these tests. There is a wide variety of situations included in the datasets, from healthy controls through anterior, lateral, and septal MI as well as inferior, lateral, and posterior-lateral MI. The findings validated the efficacy of the suggested method in identifying MI disorders.

The authors [18] presented a novel approach to the automated identification and localization of MI using cardiac electrical activity analysis. The 12-lead ECG data from 200 participants is segmented in this research; this contains 125,652 "normal" beats and 485,753 "indicative of myocardial infarction" beats. Initially, a signal was discretely wavelet transformed (DWT) [19] at depths of up to four using data collected from 12 cardiac electrodes. The DWT coefficients are then used to derive a set of twelve nonlinear characteristics. The collected data is then evaluated using a t-score. By using 47 characteristics extracted from lead 11, the suggested technique was able to obtain the greatest average accuracy (91.8%), sensitivity (91.4%), and specificity (91.27%) in identifying normal and MI ECG (two classes). Using an artificial neural network classifier [20], we were able to automatically categorize cardiac arrhythmias into one of five groups. An ANN classifier is trained using the linear and nonlinear parameters retrieved from a heart rate variability (HRV) signal. The suggested method was validated by testing it on the MIT-BIH arrhythmia database, where it achieved a success rate of 92.38 percent.

In [21], the authors proposed an automatic cardiac arrhythmia classification technique using probabilistic neural networks (PNN) trained on multi-channel ECG data. Using the results of this research, we were able to reliably categorize arrhythmias for use in diagnostic decision support systems. In [22], the authors conducted a study on the various approaches for evaluating the HRV signal retrieved from the ECG waveform. After the HRV signals have been described in terms of these characteristics, they are submitted to classifiers, which further classify them (for healthy persons and diseases like heart disease and diabetes) to aid in diagnosis.

To improve the efficiency and accuracy with which AF may be detected in displayed ECG traces, authors investigated and implemented a DL technique, including a recurrent neural network (RNN) and gated recurrent units (GRUs) [23]. This research made use of a public dataset from Physionet at MIT-BIH. As far as we are aware, this is the first time DL has been used to produce a real-time diagnosis of AF. This article presents experimental findings showing that RNN and GRU each attain an accuracy of 0.920 and 0.900, respectively. However, there is a need for further filtering, pre-processing, or de-noising when using this method. Through methods including missing data management, feature selection, and dimension reduction, preprocessing processes normalize dataset dimensions, improving data quality and making it more suitable for statistical analysis and machine learning. Data is normalized to a standard range, resulting in increased algorithm efficiency and assuring consistency and equivalence.

The DL Algorithm used a long short-term memory (LSTM) [24]-RNN to determine the presence and severity of cardiac diseases using ECG data. The model's excellent accuracy was validated using many different cardiac disease datasets. We tested the model on an arrhythmia dataset to see how well it could distinguish between cardiac diseases and arrhythmia (which has a high likelihood of mimicking the associated heart rate variability).

Further, the PhysioBank's [25] goal is to provide researchers with easy access to high-quality digital recordings of physiological signals and accompanying data. So far, it has accumulated databases of multiparameter cardiopulmonary, brain, and other biological signals from both healthy individuals and patients suffering from a broad variety of diseases with serious implications for public health. Some examples of these conditions are life-threatening arrhythmias, congestive heart failure, cardiac diseases, neurological disorders, and old age. The PhysioToolkit is a free and open-source software library for a wide variety of physiological signal processing and analysis tasks, including but not limited to displaying and characterising signals interactively [26]; creating new databases; simulating physiological and other signals; quantitatively comparing analysis methods; and detecting physiologically significant events using both conventional and novel methods based on statistical physics and nonlinear dynamics. PhysioToolkit was developed by the UCSF Physiological Signal Processing and Analysis Group. When it comes to sharing and analyzing recorded biological signals, PhysioNet is the community-driven infrastructure you need. This data bank is available via the World Wide Web. It included functions that encouraged collaborative data analysis and algorithmic improvement by teams.

3 Proposed system

Cardiac arrhythmia is the medical term for an irregular heartbeat. Therefore, preventing significant health issues through an accurate and automated examination of ECG data is essential. However, traditional manual prediction systems and conventional AI, ML, and DL approaches were unable to identify cardiac arrhythmias because they were unable to extract characteristics from ECG datasets. By learning hierarchical features from ECG datasets, deep learning algorithms might enhance the identification process while overcoming obstacles such as signal complexity, noise, and model complexity. This can increase the accuracy of recognizing cardiac arrhythmia situations. This is the reason they were unable to detect cardiac arrhythmia. This research aims to apply optimal

feature selection, UML, and DL techniques to the problem of identifying heart arrhythmias from raw ECG data. Figure 1 shows the block diagram of the proposed CardiacNet. Here, the MIT-BIH dataset is considered for evaluating the performance of the overall system. Initially, the dataset is split into 80% for training and 20% for testing. Then, the entire operation is going to be performed on both training and testing datasets. To ensure consistent dimensions, maintain data integrity, and facilitate correct analysis, normalization entails activities including preprocessing, managing missing data, encoding, scaling, dimension reduction, feature selection, and outlier treatment. Further, pre-processing operations are carried out to remove the missing symbols and unknown characters, including special characters. To ensure transparency and repeatable findings, text analysis pre-processing comprises managing missing symbols, unknown characters, and special characters using methods including replacement, normalization, regular expressions, spell checking, human inspection, and domain-specific handling. The pre-processing operation also normalizes the number of rows and columns present in the dataset. After that, UML-PCA is applied to the preprocessed dataset to extract its features. By concentrating on critical data patterns, Principal Component Analysis (PCA), a vital data analysis approach, decreases dimensionality, extracts key features, improves model performance, and increases efficiency. In addition, the IHHO model, which is inspired by nature, is used in the process of carrying out the optimum feature selection operation. By imitating the cooperative hunting techniques of Harris's hawks, employing adaptive strategies, reproduction, mutation, and learning to efficiently explore feature space, the Improved Harris Hawk's Optimisation model enhances feature selection. In addition, the CCNN model can categorize the many different cardiac arrhythmia illnesses by using IHHO properties. The Better A more sophisticated version of the Harris Hawk's Optimisation method, the Harris Hawk's Optimisation (IHHO) model was created for the best feature selection in data analysis and machine learning. Iteratively refines feature subsets for better performance, using inspiration from cooperative hawk hunting. Based on ECG signals, CNN models may categorize cardiac arrhythmias, with accuracy varied depending on system design, database effectiveness, and medical knowledge. Finally, CardiacNet classifies sinus bradycardia, right bundle branch block, old anterior myocardial infarction, ischemic changes, coronary artery diseases, and the normal heart. Finally, CardiacNet classifies sinus bradycardia, right bundle branch block, old anterior myocardial infarction, ischemic changes, coronary artery diseases, and the normal heart.

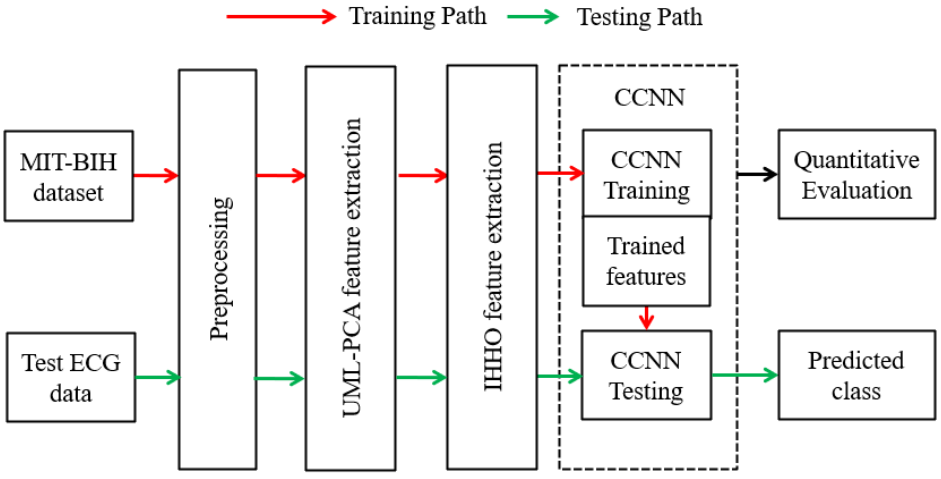
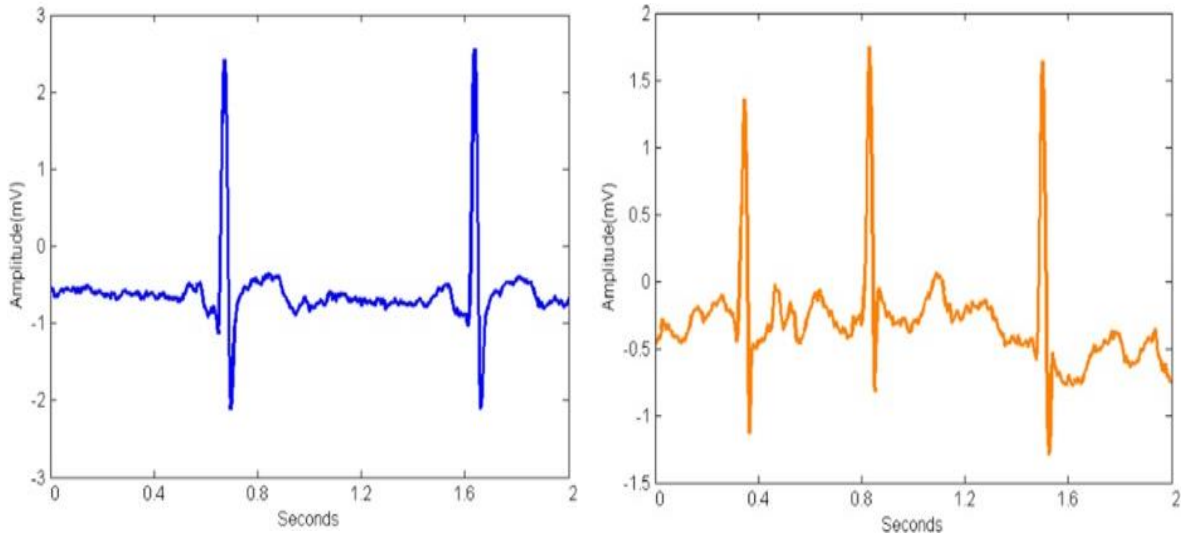


Figure 1: Proposed CardiacNet block diagram.

3.1 MIT-BIH dataset

All the MIT-BIH Arrhythmia Database includes 48 unique 30-minute recordings of two-channel ambulatory ECGs. Between 1975 and 1979, 47 subjects were studied with the use of these recordings by the BIH Arrhythmia Laboratory. Key components such as the dataset, feature extraction, pre-processing, machine learning algorithm, findings, discussion, performance assessment, and conclusion are used to evaluate the efficiency of the MIT-BIH Arrhythmia Database system and make recommendations for further research. Twenty-three recordings were randomly selected from four thousand 24-hour ambulatory ECG recordings taken from inpatients (roughly 60%) and outpatients (roughly 40%) at Boston's Beth Israel Hospital; the remaining 25 recordings were

selected from the same set to include less common but clinically significant arrhythmias. Figure 2 shows the sample ECG signals from the MIT-BIG dataset; those are normal ECG signals and cardiac arrhythmia ECG signals. The cardiac arrhythmia ECG signal contains high amplitude fluctuations that can aid in the diagnosis of arrhythmia diseases. High amplitude cardiac arrhythmia fluctuations ECG signals help with illness diagnosis, aberrant rhythm detection, pattern creation, visual identification, clinical decision-support, advancement tracking, and research comprehension.



(a) (b)

Figure 2: Sample signals from MIT-BIH dataset [1]. (a) normal ECG signal, (b) cardiac arrhythmia ECG signal.

3.2 UML-PCA feature extraction

Potential applications of UML-PCA span numerous disciplines, from population genetics to studies of the microbiome to studies of the atmosphere and beyond. A technique called UMAP-PCA reduces dataset dimensions by combining UMAP and PCA, combining global and local structures, and offering a useful lower-dimensional approximation for dealing with complicated high-dimensional data. In a real coordinate system, a collection of points is represented by a list of unit vectors. Each of these scalars represents an orthogonal line segment that best matches the data. For this definition, the best-fitting line is the line with the smallest mean perpendicular distance between the data points and the line. In linear regression, the "best-fitting line" is found by minimizing the sum of squared residuals between the data points and the line, attempting to reduce standard deviation and offer the most accurate approximation. These directions make up an orthonormal basis that ensures the different data dimensions are not linearly connected. UML-PCA is used to execute a change of basis on data by identifying the principal components and then acting on that insight. While using UML-PCA to reduce the number of dimensions might increase efficiency, make interpretation easier, and improve model performance, it also comes with hazards including information loss and incorrect interpretation. Even if it doesn't always happen, UML-PCA usually ignores everything except the most important factors. So, it is often used for assessing massive datasets with several variables or dimensions for each observation. This technology allows for the visualization of data in several dimensions and improves data interpretability without losing any relevant details. To reduce the number of dimensions in a dataset, statisticians might use UML-PCA. Utilizing a transformation matrix, linear transformations change the original coordinate system while preserving data linkages. They are essential in mathematics for data interpretation and analysis and are employed in domains including graphics, image processing, and physics. To achieve this goal, a linear transformation is applied to the data to relocate it to a new coordinate system where the data's variance may be explained using fewer variables. Because of this, we may now further simplify matters. To make it simpler to identify clusters of numbers, many studies simply display the data in two dimensions, depending on the first two primary components. The first two components are critical for doing this and identifying the disease-specific features.

3.3 IHHO algorithm

In recent years, there has been a surge in interest and understanding of how to implement evolutionary algorithms and swarm intelligence algorithms effectively, cheaply, and efficiently. The No Free Lunch (NFL) theorem states that no algorithm is a universally optimal optimizer for any issue. According to the NFL theorem,

no one optimization method can be the best in all application domains. To overcome these constraints, new optimizers that employ population-based strategies, hybrid methods, adaptive techniques, parallel computing, and transfer learning are being developed. The NFL theorem has led to new optimizers with specific local and global searching strategies. Strategies for local and global searches are essential for fine-tuning a new optimizer and boosting productivity and effectiveness. Efficiency is improved by combining both approaches in a new optimizer, encouraging speedier convergence and robust exploring. Figure 3 shows the flowchart of the IHHO algorithm for optimal feature selection. The unique nature-inspired optimization approach was inspired by Harris' Hawks, one of the world's most intelligent birds. Because of their outstanding team-chasing skills, flexibility, energy efficiency, and social cohesiveness, Harris's hawks excel at cooperative hunting. Their communal feeding and specialized tasks increase their success on the hunt. Unlike other raptors, the Harris's hawk forages cooperatively with other members of its stable group, whereas other raptors are more likely to attack and take prey alone. Due to its highly evolved inventive team-chasing abilities, this desert predator can hunt down potential prey, encircle them, flush them out, and then attack them. Predators in the desert use complex team-chasing tactics to identify and catch prey. These tactics include coordinated attempts, disorientation, stress, herding behavior, strategic placement, and group learning. The following steps show the optimal feature selection process.

Step 1: Set T, and N as the HHO parameters, which hold the properties of IHHO.

Step 2: Initialize the location of the population of Harris Hawks X of N random particles with n, t dimensions.

Step 3: The objective function for each element of X is evaluated. The average location of hawks

$$X_m(t) = \frac{1}{N}\sum_{i=1}^{N} X_i(t) \tag{1}$$

where $X_i(t)$ signifies each hawk position in iteration t, where N is the total number of hawks in the iteration.

Step 4: Set the location of the rabbit (best location) as Xrabbit.

Step 5: For each hawk Xi calculate E. The prey's energy is represented as:

$$E = 2E_0\left(1 - \frac{t}{T}\right) \tag{2}$$

Here T signifies the maximum number of repeats and E_0 is energy at its starting point.

Step 6: Depending on the energy value of his prey update the location of Harris Hawks X.

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1|X_{rand}(t) - 2r_2X(t)| & q \geq 0.5 \\ (X_{rabbit}(t) - X_m(t)) - r_3(LB + r_4(UB - LB)) & q < 0.5 \end{cases} \tag{3}$$

Here, $X(t+1)$ denotes the position vector of hawks in the next iteration, $X_{rabbit}(t)$ denotes the rabbit position, $X(t)$ denotes the hawks' current position vector, $X_{rabbit}(t)$ denotes random numbers between (0,1) that update in each iteration, and the number of iterations is R. The variables' upper and lower limits are represented by the letters LB and UB, $X_{rand}(t)$ represents any randomly picked hawk in the current members population, and X_m represents the current population hawks' average location.

Step 7: The t index is increased in 1, if the stop criteria ($t \geq T$) are not satisfied jump to step 5. Finally, the optimal features are selected from the available data.

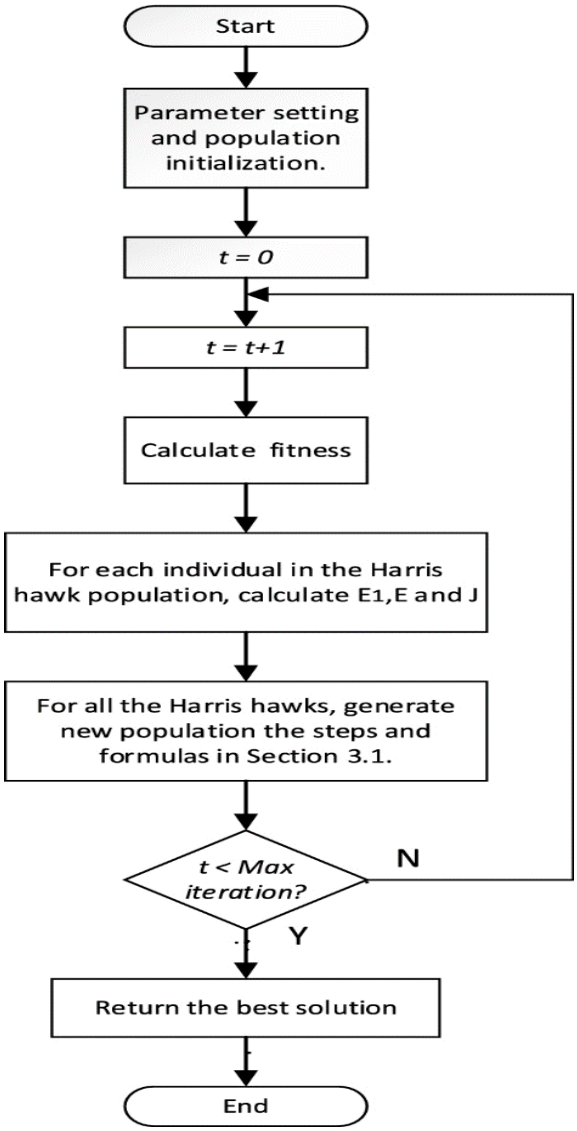


Figure 3: Flowchart of IHHO.

3.4 Proposed CCNN

The ECG data was categorized according to cardiac arrhythmia via the development of a multi-layer deep convolutional network. Figure 4 shows the proposed CCNN model enables automatic input ECG classification without requiring any manual feature extraction or selection. The suggested CNN model for classifying cardiac arrhythmias makes use of a specialized architecture, ECG signal processing, convolutional layers, pooling, non-linearity, regularisation, optimization, tuning hyperparameters, early stopping validation, Class Activation Mapping, and interpretability metrics. This is feasible because of the holistic design. The CCNN structure is at the heart of the deep network model, which is built on the layers of a CCNN. Input, convolutional, and activation functions are used in the construction of CCNN layers, which are impacted by the properties of the input data, the needs of the issue, and the selection of the hyperparameters. Feature maps, which are representations of ECG segments, are concatenated using weights of varying sizes in multi-dimensional convolution layers. A convolution with 128 weight vectors is used on the ECG data in the first layer of the model. To avoid gradient problems, assist with uniform weight initialization, promote faster convergence, and improve generalization, normalizing activation outputs is an essential preprocessing step in deep learning. Before processing each batch, the activation outputs of this layer are normalized by the batch normalization layer. By preventing division by zero, batch normalization improves training, regularisation, and neural network efficiency by ensuring uniform allocation of activating responses within every batch. The max pooling layer takes the maximum value from each region specified on the feature maps gathered in the layers below it to generate new feature maps. To make better use of the available feature space,

the feature maps from the previous layer have been reduced in size in this one. Reducing the size of feature maps is a crucial step in making DL systems faster and more resource-friendly. Decreased feature map size in deep learning methods increases computational speed, resilience, and efficiency, making models more scalable, deployable, and quick on devices with limited resources. To achieve this, we use methods like averaging the max pooling layer's maximum values. In the fourth layer, the convolution technique is repeated many times on the input feature maps using weights that are 32 by 7. With the pooling approach applied to the layer, the feature maps with a region width of two are decreased by a factor of half. This is done by completing yet another round of batch normalization. The next layer is a convolution, and the layer after that is a max pooling, both of which are repeats of the prior layers' algorithms. The feature maps collected at the previous layer will be downscaled in this step so that they may be used as input at higher network levels. When dealing with multidimensional feature vectors, we "flatten" them into a single dimension. Following their transit through the flattened layer, the features are sent onto a 512-unit, strongly linked neural network layer. A SoftMax layer serves as the network's last layer. This layer determines the total number of output classes. For multi-class classification problems, the Softmax layer is an essential last layer in neural networks. It generates normalized probability distributions, facilitates training, stable optimization, and understanding, handles one-hot encoded labels, and supports model assessment and inference. To forecast what category, the input data belongs to, the SoftMax layer is used. To prevent the network from becoming too specialized as it learns, several layers have a dropout parameter. Because of this, the layer's precision is limited to prevent problems. By periodically deactivating neurons during training, increasing resilience, and limiting co-adaptation, the dropout parameter is a regularisation strategy for neural networks that reduces overfitting and improves generalization.

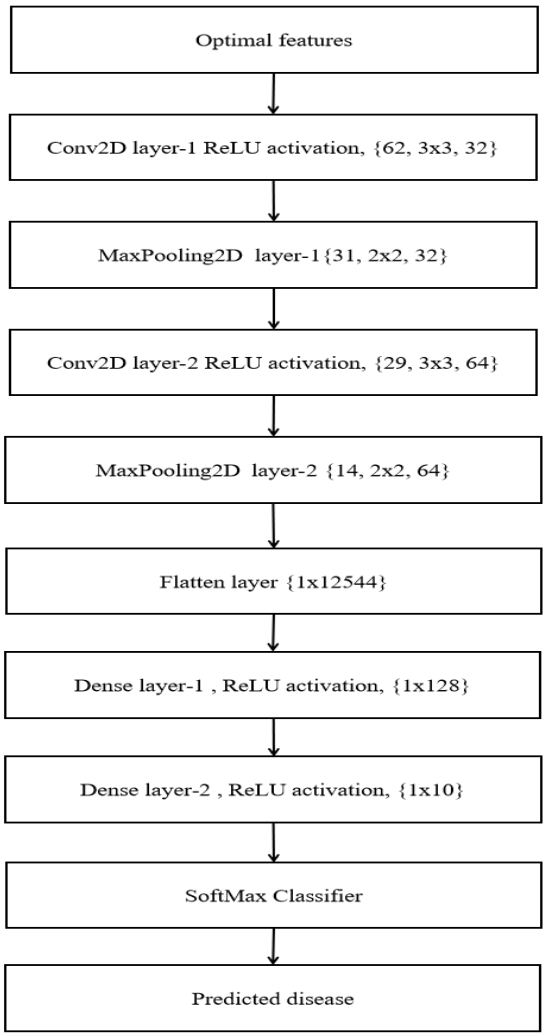


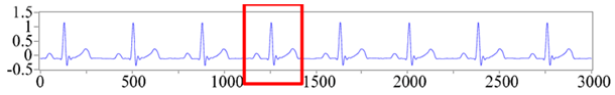
Figure 4: Block diagram of the proposed CCNN model.

4 Results and discussion

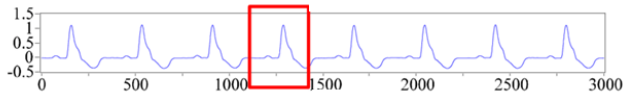
This section gives a detailed performance analysis of the proposed CardiacNet. The performance of the proposed method is measured using several performance metrics, such as accuracy, sensitivity, specificity, F-measure, precision, MCC, dice, and Jaccard. All these metrics are measured for proposed methods as well as existing methods. Then, all the methods use the same MIT-BIH dataset for performance estimations.

4.1 Subjective analysis

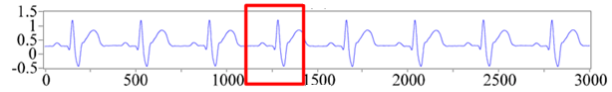
Figure 5 shows the classified outcomes using CardiacNet. The test input is considered several ECG signals, and its data is converted into a binary domain. Now, the CardiacNet analyses these binary signals, yielding a classification result. Here, the red-colored box shows the optimal feature zone. Figure 5(a) shows the classified outcome as sinus bradycardia; Figure 5(b) shows the classified outcome as right bundle branch block; Figure 5(c) shows the classified outcome as anterior myocardial infarction; Figure 5(d) shows the classified outcome as ischemic changes; Figure 5(e) shows the classified outcome as coronary artery diseases; and Figure 5(f) shows the classified outcome as a normal heart.



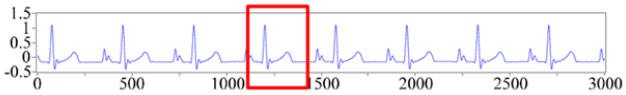
(a) sinus bradycardy



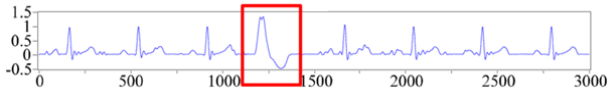
(b) Right bundle branch block



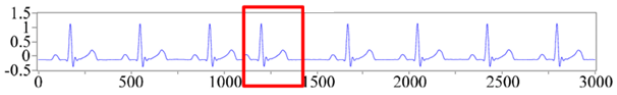
(c) Old Anterior Myocardial Infarction



(d) Ischemic changes



(e) Coronary artery diseases



(f) normal heart

Figure 5: Classification results from CardiacNet

4.2 Performance comparison

Table 1 compares the classification performance of various approaches to the proposed CardiacNet. The first column contains performance metrics. The second column contains the performance estimation during the SVM [17] method. The third column contains the estimated performance of the PNN [20] method. For assessing the generalization, model selection, hyperparameter tweaking, overfitting identification, prediction confidence, task applicability, and iterative improvements of the PNN approach, performance is essential. In the fourth column, the performance estimation during GRU [23] is presented. Here, the proposed CardiacNet resulted in improved classification performance as compared to SVM [17], PNN [20], and GRU [23].

Table 2 presents the percentage of improvements over Table 1. Here, the CardiacNet has increased accuracy by 4.21%, sensitivity by 6.38%, specificity by 2.24%, F-measure by 4.88%, precision by 6.08%, MCC by 2.37%, dice by 3.79%, and Jaccard by 3.77% as compared to the SVM [17]. In the third column, CardiacNet increased accuracy by 4.21%, sensitivity by 6.38%, specificity by 2.24%, F-measure by 4.88%, precision by 6.08%, MCC by 4.52%, dice by 3.79%, and Jaccard by 3.77% as compared to PNN [20]. In the last column, CardiacNet increased accuracy by 3.66%, sensitivity by 6.20%, specificity by 3.70%, F-measure by 5.45%, precision by 7.82%, MCC by 7.59%, dice by 4.48%, Jaccard by 1.55% as compared to GRU [23].

Table 1: Performance comparison of various methods.

Metric	SVM [17]	PNN [20]	GRU [23]	Proposed CardiacNet
Accuracy (%)	93.62	93.62	94.12	97.57
Sensitivity (%)	92.39	92.39	92.55	98.29
Specificity (%)	95.82	95.82	94.47	97.97
F-measure (%)	92.86	92.86	92.36	97.40
Precision (%)	93.00	93.00	91.50	98.66
MCC (%)	95.89	93.92	91.24	98.17
Dice (%)	95.34	95.34	94.71	98.96
Jaccard (%)	93.59	93.59	95.63	97.12

Table 3 presents the ablation study for the proposed CardiacNet. Here, the first column contains the different performance metrics. The second column contains the performance estimation using only CCNN, where UML-PCA and IHHO are absent. Then, the third column contains the performance estimation during UML-PCA with the CCNN method, where IHHO feature selection is absent. Finally, the last column contains the performance of CardiacNet with all modules presented. Here, the performance of the proposed method is improved when all modules are present in comparison to the absence of any individual module. So, this ablation study shows the significance of UML-PCA feature extraction and IHHO feature selection.

Table 2: Percentage of improvement in proposed CadiacNet model as compared to state-of-the-art methods.

Metric	SVM [17]	PNN [20]	GRU [23]
Accuracy (%)	4.219184	4.219184	3.665533
Sensitivity (%)	6.385973	6.385973	6.202053
Specificity (%)	2.24379	2.24379	3.70488
F-measure (%)	4.88908	4.88908	5.456908
Precision (%)	6.086022	6.086022	7.825137
MCC (%)	2.377724	4.525128	7.595353
Dice (%)	3.796937	3.796937	4.487383
Jaccard (%)	3.77177	3.77177	1.558088

Table 4 presents the percentage of improvements over Table 3. Here, the CardiacNet has increased accuracy by 5.34%, sensitivity by 6.32%, specificity by 4.60%, F-measure by 2.18%, precision by 5.76%, MCC by 1.76%, dice by 1.39%, and Jaccard by 4.25% as compared to the only CCNN case. Performance measures such as accuracy of 5.04%, sensitivity of 3.46%, specificity of 1.70%, F-measure of 4.78%, precision of 4.68%, MCC of 2.23%, dice of 5.23%, and Jaccard of 5.53% are compared to the UML-PCA+ CCNN presented case in the third column. In the final column, accuracy by 6.50%, sensitivity by 2.79%, specificity by 6.33%, F-measure by 4.69%, precision by 6.82%, MCC by 5.14%, dice by 5.99%, and Jaccard by 2.81% are compared to the IHHO+ CCNN presented case.

Table 3: Ablation study of the proposed method.

Metric	CCNN	UML-PCA+ CCNN	IHHO + CCNN	Proposed CardiacNet
Accuracy (%)	92.65	92.91	91.64	97.57
Sensitivity (%)	91.54	94.07	94.68	98.29
Specificity (%)	93.97	96.65	92.44	97.97
F-measure (%)	96.43	94.04	94.12	97.40
Precision	92.90	93.86	91.98	98.66

(%)				
MCC (%)	96.40	95.96	93.30	98.17
Dice (%)	96.01	92.51	91.84	98.96
Jaccard (%)	94.89	93.74	96.22	97.12

Table 4: Percentage of improvements in ablation study as compared to state-of-the-art methods.

Metric	CCNN	UML-PCA+ CCNN	IHHO+ CCNN
Accuracy (%)	5.342688	5.047896	6.50371
Sensitivity (%)	6.325104	3.465504	2.798902
Specificity (%)	4.607854	1.707191	6.339247
F-measure (%)	2.188116	4.785198	4.696133
Precision (%)	5.769645	4.687833	6.827571
MCC (%)	1.763485	2.230096	5.144695
Dice (%)	1.395688	5.231867	5.999564
Jaccard (%)	4.257561	5.536591	2.816462

5 Conclusion

This work effort developed CardiacNet, which is an AI tool for identifying cardiac arrhythmias using a dataset based on MIT and BIH. To begin, a pre-processing procedure is carried out on the dataset to eliminate any non-linearities that it may include. After that, UML-PCA is applied to the pre-processed dataset to extract its features. In addition, the IHHO model, which is inspired by nature, is used in the process of carrying out the optimum feature selection operation. In addition, the CCNN model can categorize the many different cardiac arrhythmia illnesses by using IHHO properties. Finally, CardiacNet classifies sinus bradycardia, right bundle branch block, old anterior myocardial infarction, ischemic changes, coronary artery diseases, and the normal heart. Here, the CardiacNet has increased accuracy by 4.21%, sensitivity by 6.38%, specificity by 2.24%, F-measure by 4.88%, precision by 6.08%, MCC by 2.37%, dice by 3.79%, and Jaccard by 3.77% as compared to existing methods. Further, this work can be extended with improved transfer learning methods and class-specific performance estimation.

Acknowledgements

Funding: The authors did not receive any funding.

Conflicts of Interests: Authors do not have any conflicts.

Data Availability Statement: No datasets were generated or analyzed during the current study.

Code availability: Not applicable.

Authors' Contributions: All author is contributed to the design and methodology of this study, the assessment of the outcomes and the writing of the manuscript.

References

[1] Pipawala, R., Paul, A., & Mukherjee, T., (2023). Prediction of heart disease using modified hybrid classifier. *Informatica*, 65-72. https://doi.org/10.31449/inf.v47i1.3629

[2] Liu, Y., Qin, C., Liu, J., Jin, Y., Li, Z., & Liu, C (2022). An efficient neural network-based method for patient-specific information involved arrhythmia detection. *Knowledge-Based Systems*, 109021. https://doi.org/10.1016/j.knosys.2022.109021.

[3] Serhal, H., Abdallah, N., Marion, J. M., Chauvet, P., Oueidat, M., & Humeau-Heurtier, A (2022). Overview on prediction, detection, and classification of atrial fibrillation using wavelets and AI on ECG. *Computers in Biology and Medicine*, 105168. https://doi.org/10.1016/j.compbiomed.2021.105168

[4] Chen, Y., Zhang, C., Liu, C., Wang, Y., & Wan, X (2022). Atrial fibrillation detection using a feedforward neural network. *Journal of Medical and Biological Engineering*, 63-73. https://doi.org/10.1007/s40846-022-00681-z.

[5] Mohonta, S. C., Motin, M. A., & Kumar, D. K (2022). Electrocardiogram based arrhythmia classification using wavelet transform with deep learning model. *Sensing and Bio-Sensing Research*, 100502. http://dx.doi.org/10.1016/j.sbsr.2022.100502.

[6] Rahul, J., & Sharma, L. D (2022). Automatic cardiac arrhythmia classification based on hybrid 1-D CNN and Bi-LSTM model. *Biocybernetics and Biomedical Engineering*, 312-324. https://doi.org/10.1016/j.bbe.2022.02.006.

[7] Rahul, J., & Sharma, L. D (2022). Artificial intelligence-based approach for atrial fibrillation detection using normalised and short-duration time-frequency ECG. *Biomedical Signal Processing and Control*, 103270. https://doi.org/10.1016/j.bspc.2021.103270.

[8] Li, H., Lin, Z., An, Z., Zuo, S., Zhu, W., Zhang, Z., ... & García, J. D. P (2022). Automatic electrocardiogram detection and classification using bidirectional long short-term memory network

improved by Bayesian optimization. *Biomedical Signal Processing and Control*, 103424. https://doi.org/10.1016/j.bspc.2021.103424.

[9] Ivora, A., Viscor, I., Nejedly, P., Smisek, R., Koscova, Z., Bulkova, V., ... & Plesinger, F (2022). QRS detection and classification in Holter ECG data in one inference step. *Scientific Reports*, 12641. https://doi.org/10.1038/s41598-022-16517-4.

[10] Hammad, M., Abd El-Latif, A. A., Hussain, A., Abd El-Samie, F. E., Gupta, B. B., Ugail, H., & Sedik, A (2022). Deep learning models for arrhythmia detection in IoT healthcare applications. *Computers and Electrical Engineering*, 108011. https://doi.org/10.1016/j.compbiomed.2022.105325.

[11] Srivastava, R., Kumar, B., Alenezi, F., Alhudhaif, A., Althubiti, S. A., & Polat, K (2022). Automatic arrhythmia detection based on the probabilistic neural network with FPGA implementation. *Mathematical Problems in Engineering*. https://doi.org/10.1155/2022/7564036.

[12] Al-Jammali, K (2023). Prediction of heart diseases using data mining algorithms. *Informatica*, 57-62. https://doi.org/10.31449/inf.v47i5.4467

[13] Kim, Y. K., Lee, M., Song, H. S., & Lee, S. W (2022). Automatic cardiac arrhythmia classification using residual network combined with long short-term memory. *IEEE Transactions on Instrumentation and Measurement*, 1-17. https://doi.org/10.1109/TIM.2022.3181276.

[14] Castillo-Atoche, A., Caamal-Herrera, K., Atoche-Enseñat, R., Estrada-López, J.J., Vázquez-Castillo, J., Castillo-Atoche, A.C., Palma-Marrufo, O. and Espinoza-Ruiz, A (2022). Energy efficient framework for a AIoT cardiac arrhythmia detection system wearable during sport. *Applied Sciences*, p.2716. https://doi.org/10.3390/app12052716.

[15] Król-Józaga, B (2022). Atrial fibrillation detection using convolutional neural networks on 2-dimensional representation of ECG signal. *Biomedical Signal Processing and Control*, 103470. https://doi.org/10.1016/j.bspc.2021.103470.

[16] Kim, J. K., Jung, S., Park, J., & Han, S. W (2022). Arrhythmia detection model using modified DenseNet for comprehensible Grad-CAM visualization. *Biomedical Signal Processing and Control*, 103408. https://doi.org/10.1016/j.bspc.2021.103408.

[17] Guess, M., Zavanelli, N., & Yeo, W. H (2022). Recent advances in materials and flexible sensors for arrhythmia detection. *Materials*, 724. https://doi.org/10.3390%2Fma15030724.

[18] Madhavi, K. R., Kora, P., Reddy, L. V., Avanija, J., Soujanya, K. L. S., & Telagarapu, P (2022). Cardiac arrhythmia detection using dual-tree wavelet transform and convolutional neural network. *Soft Computing*, 3561-3571. https://doi.org/10.1007/s00500-021-06653-w.

[19] Liu, Z., Chen, Y., Zhang, Y., Ran, S., Cheng, C., & Yang, G (2023). Diagnosis of arrhythmias with few abnormal ECG samples using metric-based meta learning. *Computers in Biology and Medicine*, 106465. https://doi.org/10.1016/j.compbiomed.2022.10646.

[20] Chen, S. W., Wang, S. L., Qi, X. Z., Samuri, S. M., & Yang, C (2022). *Review of ECG detection and classification based on deep learning: coherent taxonomy, motivation, open challenges and recommendations.* Biomedical Signal Processing and Control, 103493. https://doi.org/10.1016/j.bspc.2022.103493.

[21] Wesselius, F. J., van Schie, M. S., De Groot, N. M., & Hendriks, R. C (2022). An accurate and efficient method to train classifiers for atrial fibrillation detection in ECGs: Learning by asking better questions. *Computers in Biology and Medicine*, 105331. https://doi.org/10.1016/j.compbiomed.2022.105331

[22] Hu, R., Chen, J., & Zhou, L (2022). A transformer-based deep neural network for arrhythmia detection using continuous ECG signals. *Computers in Biology and Medicine*, 105325. https://doi.org/10.1016/j.compbiomed.2022.105325

[23] Kuila, S., Dhanda, N., & Joardar, S (2022). ECG signal classification and arrhythmia detection using ELM-RNN. *Multimedia Tools and Applications*, 25233-25249. https://doi.org/10.1007/s11042-022-11957-6.

[24] Rizqyawan, M. I., Siradj, Y., Amri, M. F., & Pratondo, A (2022). Re-implementation of convolutional neural network for arrhythmia detection. *International Journal on Advanced Science, Engineering and Information Technology*, 1319-1326. http://dx.doi.org/10.18517/ijaseit.12.4.13435.

[25] Liu, Z., Zhou, B., Jiang, Z., Chen, X., Li, Y., Tang, M., & Miao, F (2022). Multiclass Arrhythmia Detection and Classification from Photoplethysmography Signals Using a Deep Convolutional Neural Network. *Journal of the American Heart Association*, e023555. https://doi.org/10.1161/jaha.121.023555.

[26] Shibly, F. H. A & Lakshmana Kumar. R (2023). Image Processing for Automatic Cell Nucleus Segmentation Using Super pixel and Clustering Methods on Histopathological Images. *Tamjeed Journal of Healthcare Engineering and Science Technology*, 54-63. http://dx.doi.org/10.59785/tjhest.v1i1.6

Parkinson Net: Convolutional Neural Network Model for Parkinson Disease Detection from Image and Voice Data

Subba Reddy Borra[1]*, K. Ritika[2], N. Akshaya Reddy[2], Neela Shanvitha[2], T. Anusha[2], O. Sri Sai Raja Rajeswari[2]
[1]Department of Information Technology, Malla Reddy Engineering College for Women (UGC-Autonomous), Maisammaguda, Hyderabad, India.
[2]Department of Information Technology, Malla Reddy Engineering College for Women (UGC Autonomous), Hyderabad, India.
E-mail: subba_reddyborra67@outlook.com
*Corresponding author

Keywords: parkinson's disease, deep learning, categorical cross entropy, convolutional neural network, image and voice data.

Received: August 6, 2023

Parkinson's disease (PD) is a critical dopaminergic neuron problem that causes brain disorders. The early prediction of PD can save human lives. So, computer-aided detection (CAD) with artificial intelligence (AI) models can predict PD in a quick time as compared to manual prediction. Traditional machine learning (ML) methods, on the other hand, identified PD using either voice or image datasets. However, they resulted in poor PD detection performance, which caused misclassification. So, this work focused on the implementation of a deep learning (DL) mechanism for PD identification from both voice and image datasets, which is named ParkinsonNet. Initially, a combined dataset is considered, which contains the voice and image samples. Then, a data processing operation is performed to normalize the images to a uniform size, which also performs the data balancing operation in the voice dataset. Then, a voice-image ensemble-based convolutional neural network (VIE-CNN) model is trained with the pre-processed voice-image data. Here, categorical cross entropy (CCE) is used to optimize the losses generated during the training. Then, the VIE-CNN model predicts the normal and abnormal classes from the test data. The simulation results show that the proposed ParkinsonNet achieved 99.67% accuracy on image data and 98.21% accuracy on voice data. The simulation results show that the proposed ParkinsonNet resulted in improved accuracy over conventional methods.

Povzetek: Predstavljen je sistem ParkinsonNet, ki uporablja globoko učenje za identifikacijo Parkinsonove bolezni iz glasovnih in slikovnih podatkov. Pri tem dosega 99,67 % točnost pri slikovnih podatkih in 98,21 % pri glasovnih podatkih.

1 Introduction

PD is a neurological disorder that primarily affects people in their s and older [1]. It is often regarded as one of the ailments that are most difficult to cure in the current world. PD is characterized by tremors, bradykinesia, and postural instability. Parkinson's syndrome, which affects most sufferers, is a form of movement disorder [2]. The movement disorder's subtype associated with PD is by far its most prevalent symptom. The condition known as PD is the most prevalent cause of movement problems and the one that affects the most people. PD is a neurological ailment and a chronic sickness that mostly affects older people today. The symptoms include slowness of movement, stiffness, and tremors. Parkinson's disease's main clinical symptoms are tremors, rigidity, and slowness of movement (bradykinesia). Bradykinesia causes a noticeable slowing down of movement, rigidity causes resistance to motion in the muscles, and tremors are rhythmic shaking that most often affects the hands or fingers while at rest. An approach that is both efficient and strong in its automation is required as a tool to be able to diagnose PD accurately in its early stages. Limited and noisy data, interpretability of complicated models, potential overfitting due to model complexity, and the requirement for expert expertise to properly preprocess and interpret results are some of the difficulties faced while utilizing deep learning to diagnose Parkinson's disease. One of the conditions that must be met before continuing is this. It is feasible to address these demands thanks to recent advancements in technical capabilities [3]. A harsh and breathy voice, reduced intensity, monotony of pitch and loudness, decreased tension, inappropriate silences, brief rushes of speech, fluctuating tempo, faulty consonant articulation, and inappropriate silences are all symptoms associated with PD (dysphonia). Because collecting speech data does not require any form of invasion of privacy and can simply be carried out using portable electronic equipment, there are grounds for hope for the construction of a prospective diagnostic tool that is used to treat a broad variety of voice-related ailments. When treating voice-related conditions, using speech data as a diagnostic tool has benefits over more conventional approaches. Unlike frequently subjective and less accurate traditional evaluations, it offers objective, quantitative analysis, enabling more accurate assessments and early

detection of errors. Obtaining high-quality, well-labeled data, selecting appropriate model architectures, optimizing hyperparameters, implementing regularization to prevent overfitting, prioritizing model interpretability, carrying out rigorous validation and testing, and adhering to ethical considerations, particularly about patient privacy and data transparency, are key principles when using Deep Learning techniques for diagnosis.

As the illness advances, more symptoms appear, making it increasingly difficult to discover a therapy for PD. As a direct result of this, diagnostic instruments that have a higher level of sensitivity are necessary to make an accurate diagnosis of PD. The usage of AI-based CAD systems [4,5]. for illness, diagnosis has significantly increased in recent years, sometimes even in the early stages of the ailment. Because of the technical advancements that have been achieved in this field, it is now theoretically possible for this to become a practical reality. This pattern is projected to continue at least into the not-too-distant future, according to expectations. ML and DL algorithms have been developed to diagnose PD and provide answers to its diagnostic problems [6,7]. These diagnostic algorithms were developed to help in the diagnosis of PD, and they are based on several diagnostic techniques that are now being used in clinical practice. Integrating algorithm-based diagnostic techniques into Parkinson's disease clinical practice may result in earlier and more accurate diagnoses, personalized treatment, improved monitoring, and cost-efficiency. However, it also raises ethical and privacy issues, demanding careful data management and regulatory oversight. These diagnostic algorithms were created to identify PD. This study's goal is to analyze the diagnostic techniques that rely on differential diagnosis to find PD. Examples of the sorts of things that are discussed under this topic are the pre-processing of PD datasets [8], the extraction and selection of features [9], and classification. These and other strategies might be used. In addition, this selection of features may potentially include additional categories of items. Additionally, the use of such CAD systems for the diagnosis of PD from several modalities, such as speech signals, gait signals, magnetic resonance imaging, positron emission tomography, and single-photon emission computed tomography, has increased, as have the Dopamine Transporter Scan, a tremor signal, a handwriting signal, handwritten images, and a variety of other clinical characteristics. Voice, movement, and handwriting [10] are all examples of modalities that are used to send messages. In addition to evaluating existing literature, this effort looked at current and emerging research problems and potential answers to those problems. The findings of this research have uncovered several emerging tendencies as well as research functions that, when further investigated, will contribute to the advancement of automatic disease recognition. Modern trends in autonomous illness recognition are heavily influenced by technological developments since they make it possible for more sophisticated algorithms, better sensors, quicker data processing, and greater computer capacity. These advancements improve illness recognition's precision and effectiveness, which improves patient care and outcomes. These findings will aid in the diagnosis of PD as well as its incorporation into electronic healthcare systems. The novel contributions of this work are illustrated as follows:

- The development of ParkinsonNet for the identification of PD from both voice and image datasets.
- The design of the VIE-CNN model for normal and abnormal disease classification using pre-processed datasets.
- The adoption of the CCE model for reducing losses during training, which also performs label-specific training for ~~classificatiom.~~ Classification.

The rest of the paper is organized as follows: Section 2 contains the literature survey and existing drawbacks. Section 3 contains a detailed analysis of the proposed method with sub-block explanations. Section 4 contains the detailed simulation analysis. Section 5 contains the conclusion.

2 Literature survey

[11] presented the several datasets that were used in the evaluation of the proposed PD identification algorithms. This review also looked into the model assessment metrics and cross-validation procedures that have been used in various studies on this topic. The authors [12] created the fundamental CNNs for learning characteristics from pictures formed by handwriting dynamics. These images capture various information relevant to the person being evaluated. In addition, to promote research that is associated with computer-assisted PD diagnosis, this study provides a dataset that is comprised of images and data based on signals that are available to the public. [13] conducted a comprehensive review of previous studies on the PD diagnosis and its several subtypes. They did this by analyzing the data using a computer program. Artificial neural networks (ANNs) are artificial versions of the brain's natural neural networks for the identification of PD. Dash [14] developed the ML approaches to solve these issues and to develop the processes for diagnosing and assessing PD. The categorization of individuals with PD, healthy controls, and patients whose clinical presentations were like those of PD patients were all able to be studied successfully thanks to the use of these approaches. Parkinson's disease mainly presents as motor symptoms in patients, such as tremors, bradykinesia (slowness of movement), rigidity, and postural instability. The condition is characterized by these motor symptoms, which are frequently the most obvious. In [15], the PD is identified using an incremental support vector machine (ISVM). An Incremental Support Vector Machine (ISVM) is used to accurately distinguish between cases of Parkinson's disease (PD) and cases without it by training the model on data samples, progressively updating it with fresh data, and modifying the decision boundary. In the framework of this inquiry, self-organizing maps and non-linear iterative partial least squares were used to, first,

decrease the dimensionality of the data and, second, carry out the task of clustering. Both methods were found to be effective in accomplishing these goals. In [16], the authors developed a novel concept for an innovative PD detection system, and they based its construction on techniques of recurrent neural networks to analyze gait data. In [17], the authors obtained their forecasts for total prognoses regarding the progression of the disease (UPDRS) and motor-UPDRS with the use of SVM. The UPDRS (Unified Parkinson's Disease Rating Scale) is a comprehensive assessment tool that evaluates various aspects of Parkinson's disease, including motor and non-motor symptoms. The motor-UPDRS specifically focuses on assessing motor symptoms such as tremors, rigidity, bradykinesia, and posture/gait issues. The accuracy of the results was therefore increased because of this. Within the scope of this inquiry, non-linear iterative partial least squares and self-organizing maps were used to, for starters, reduce the number of dimensions that the data had.

The Voice Impairment Classifier is the name of two neural network-based models that were presented by [18] to assist medical professionals and patients in the process of illness diagnosis at an earlier stage. To effectively predict the illness, a comprehensive empirical evaluation of CNNs was performed on large-scale image classification of gait signals that were transformed into spectrogram images. In [19], the authors explored three distinct kinds of classifiers for the benchmark (voice) dataset. The multilayer perceptron, the SVM, and the K-Nearest Neighbor (KNN) were the classifiers used to identify the PD. With a classification accuracy of 95.89%, it was found that the most successful classifier was KNN paired with the Levenberg–Marquardt method. The authors [20] created an ANN-SVM model for detecting Parkinson's disease from voice data. However, this method resulted in higher computational complexity. The authors [21] suggested using a random undersampling strategy to introduce more equilibrium into the training process. To achieve a higher level of precision when identifying PDs, the researchers who carried out this study came up with the idea of a cascaded learning system. The Chi2 model [22] and the adaptive boosting (Adaboost) [23] model is both included in this solution. The Chi2 is a model that analyses the feature space, ranks the important properties, and selects a subset of those characteristics to use in the Adaboost model's prediction of PD. The Adaboost model bases its evaluation of the subset of features on the evaluations, rankings, and selections made by the Chi2 model. [24] presented a technique for the diagnosis of PD that makes use of vowels with extended phonation and an architecture of ResNet that was originally dedicated to the classification of pictures. The accuracy that was attained on the validation set is more than 90%, which is very low. Static and dynamic speech features were explored in connection to PD detection [25-27]. They suggested using a bidirectional long-short-term memory (LSTM) model to identify PD by capturing the time-series dynamic properties of a speech signal.

3 Proposed methodology

The use of DL, a subfield of AI, is quickly expanding to include a wide range of diagnostic activities in the medical field. Deep Learning has demonstrated effectiveness in the diagnosis of skin issues, the detection of retinal diseases, the prediction of cardiac and genetic disorders, and the detection of malignancies in images, all of which improve diagnostic precision and patient care. The diagnosis of a diverse range of disorders and conditions is included in these responsibilities. This chapter provided an overview of the applications of DL techniques and discussed several key principles necessary for the diagnosis of Parkinson's disease. The integration of speech and picture samples in ParkinsonNet supports the program's goals by enabling a more thorough and precise evaluation of Parkinson's disease patients. The holistic data approach enhances the monitoring, diagnosis, and customization of treatment plans, ultimately advancing ParkinsonNet's mission to improve the quality of life for people living with Parkinson's disease. Figure 1 shows the proposed ParkinsonNet block diagram. In the beginning, a combined dataset that includes both the speech and picture samples is taken into consideration. The decision to utilize a combined dataset comprising both speech and picture samples aims to offer a more comprehensive and multi-modal perspective of the data. This approach can enhance accuracy and provide deeper insights, particularly in fields like medical diagnostics, where it allows for a more holistic assessment of a subject's condition or characteristics. Here, the handwritten patterns are considered the image dataset. Further, voice problems were shown to be related to symptoms in 90% of PD patients who were in the early stages of the illness. The progressive nature of Parkinson's disease (PD) makes it challenging to discover effective therapies because advancing symptoms diversify and intensify, making it harder to assess treatment impact. Moreover, variations in symptom patterns among individuals complicate the development of universally effective treatments. Parkinson's disease (PD) sufferers' voice issues are an important early warning indication. They frequently manifest before motor symptoms, allowing for early diagnosis and management, which can enhance patients' quality of life through prompt treatment and speech therapy. Through precise diagnostics, early warnings, individualized therapies, and data-driven assistance, advanced technology has significantly improved the accuracy and efficiency of identifying and treating illnesses in their early stages, improving patient outcomes and reducing costs. Because of this, there is an increasing amount of interest in the incorporation of speech characteristics into computer-assisted diagnosis and remote monitoring of people who are in the early stages of PD. Advanced machine learning models that evaluate vocal variables like pitch, tremor, and articulation are used to incorporate speech characteristics in computer-assisted diagnosis and remote monitoring of early-stage Parkinson's disease. These models employ speech data to identify small alterations suggestive of Parkinson's disease, providing a handy and non-invasive tool to

monitor disease development and aid in early diagnosis and treatment planning.

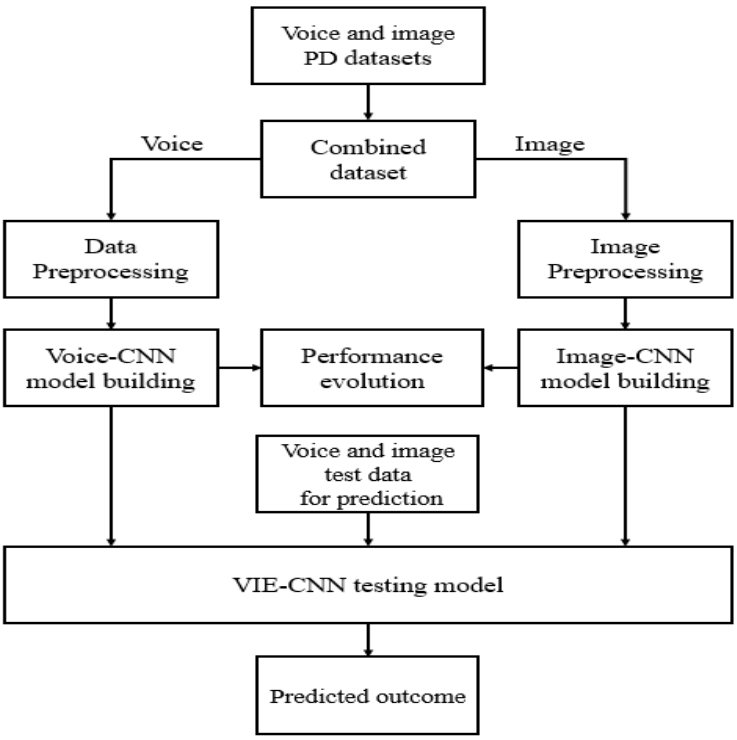


Figure 1: Proposed ParkinsonNet block diagram.

The contribution made by this research is an improvement in accuracy as well as a decrease in the number of vocal characteristics that are used in the process of PD identification. Next, a data processing operation is carried out to normalize the pictures to a size that is consistent throughout the board. This data processing operation also carries out a data balancing operation inside the voice dataset. It can be difficult to normalize images in a voice collection because of differences in lighting, backgrounds, and camera settings. Techniques including histogram equalization, color correction, background removal, and data augmentation are used in mitigation strategies to maintain consistent image quality and increase model robustness. Then, the dataset splitting operation is carried out, where 80% of the dataset is used for training and 20% of the dataset is used for testing. Following that, a VIE-CNN model is trained using the pre-processed voice-image data. Here, the VIE-CNN contains the two standalone multi-dimension models, namely the Voice-CNN and the Image-CNN. Specifically, the Voice-CNN model is used to train the pre-processed voice data, and the Image-CNN model is used to train the pre-processed image data. Here, the Voice-CNN and Image-CNN models are named based on their kernel size, such as 1x1, 3x3, etc. In this instance, CCE is used to optimize the losses that are sustained during the training. Categorical Cross-Entropy (CCE) improves training by acting as the objective function for gradient descent-style optimization techniques. As a way to reduce the difference between expected and actual class probabilities, it directs parameter modifications. The approach computes gradients to repeatedly update parameters, lowering the CCE loss and improving the model's classification accuracy. Multi-class classification using machine learning employs the loss function known as categorical cross-entropy (CCE). It quantifies the disparity between anticipated class probability and actual class labels. It is determined mathematically by adding the negative logarithm of the anticipated probabilities for the true classes. CCE directs model training by penalizing greater differences between expected and actual probabilities, encouraging more precise predictions. In the output layer of neural networks, it is frequently used in conjunction with the softmax activation. After that, the VIE-CNN model uses the test data to make predictions about the normal and abnormal classes.

3.1 Dataset

This work considered both voice and image datasets. Further, the image-based PD dataset is also considered because it is available on open-access platforms. There are a total of 204 images in the dataset, with 102 images marked as PD and 102 images marked as normal. Figure 2 shows the sample images from the dataset. Then, the voice samples of people with PD are obtained from the UCI-ML library. Voice samples from the UCI-ML library are used because changes in voice patterns can indicate Parkinson's disease.

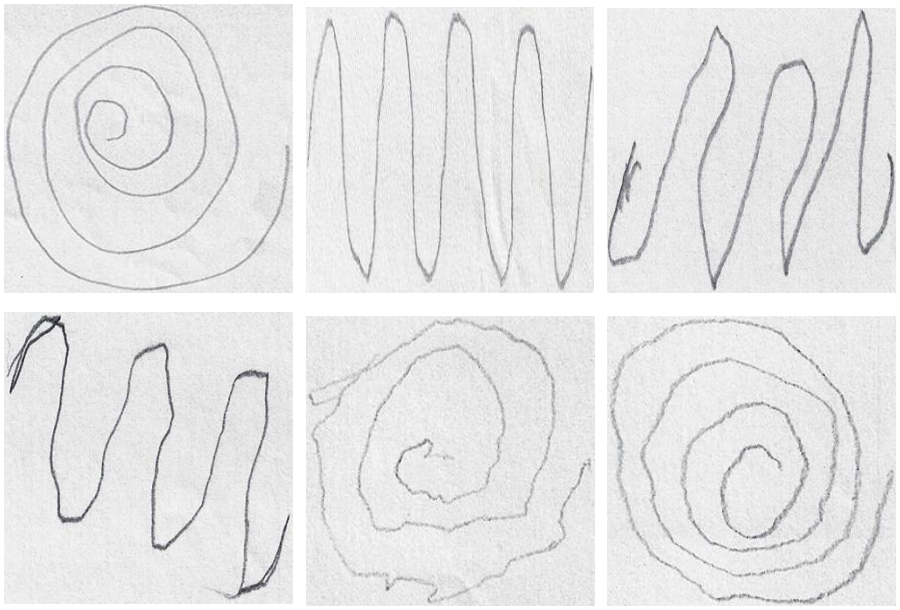


Figure 2: Sample images from the Image-PD dataset.

The voice-PD dataset contains a total of 240 records, 120 of which are normal and 120 of which are PD. The columns of the voice-PD dataset are ID, Recording, Status, Gender, and Jitter_rel Jitter_abs, Jitter_RAP, Jitter_PPQ, Shim_loc HNR05, HNR15, HNR25, HNR35, HNR38, RPDE, DFA, PPE, GNE, MFCC0, MFCC1, MFCC2, MFCC3, MFCC4, MFCC6, MFCC7, MFCC8, MFCC9, MFCC10, MFCC11, MFCC12, Delta0, Delta1, Delta2, Delta3, Delta4, Delta5 and Delta6. Differentiating PD images from normal images involves features like altered vocal pitch, tremor-induced variations, and irregular speech patterns due to motor symptoms like bradykinesia and rigidity. Here, all these columns are different spectral-spatial features of the human voice. Parkinson's disease poses a difficult challenge to treatment due to its complicated, progressive nature, dual presentation of motor and non-motor symptoms, ambiguous etiology, and incomplete understanding of its underlying mechanisms. Current medical practices tackle symptoms rather than the underlying cause of the disease, searching for a cure a drawn-out and challenging procedure.

3.2 VIE-CNN

The VIE-CNN model contains separate training models for voice and images, which are voice-CNN and image-CNN. Standard pre-processing techniques are necessary before studying speech and image collections. Noise reduction, feature extraction, and normalizing are examples of these processes for speech data, whereas scaling, normalization, and data augmentation are frequently used for image data. By performing these procedures, the data is prepared for machine learning analysis, improving its precision and caliber. Convolutional layers are used for image analysis in the VIE-CNN model, and recurrent layers are used for voice analysis. These models have distinct training models for voice and images. Through the use of combined data, this design enables the model to learn specific properties from both modalities, improving its ability to diagnose Parkinson's disease. Here, the kernel sizes in Voice-CNN and Image-CNN are different. However, the number of layers and their operations remain the same. Figure 3 shows the block diagram of Voice-CNN, which contains 1D kernels of size 1x1. Figure 4 shows the block diagram of Image-CNN with a kernel size of 3x3. Convolutional layers for feature extraction, pooling layers for downsampling, and fully connected layers for classification are the main phases in an Image-CNN (Convolutional Neural Network). The operation of each layer is illustrated as follows:

Convolution Neural Layer: The deep VIE-CNN is a multi-layered neural network that has seen recent usage in solving a variety of difficult issues. The synapses of the neurons that make up a convolution layer are coupled to the region of the input data. A convolutional layer enables each neuron to process a particular geographic area and capture local information by connecting its synapses to confined portions of the incoming data through constrained receptive fields. The receptive field of a neuron refers to the maximum extent to which it can process incoming data; this field is expanded by stacking the convolution layers. The portion of the input space that specifically affects a neuron's activity is known as the receptive field (for example, the visual field for a visual neuron). It is the spectrum of stimuli or inputs that the neuron reacts to. To comprehend the way neurons or units process data from their environment, that concept is frequently employed in fields like computer vision and neuroscience. The convolution procedure is expressed as the equation (1) with input data as In_{Map}, where w stands for the convolution kernel weights and its bias term (B_k), respectively, and out_{Map} is the output expression of the convolution technique.

$$out_{Map} = \left(In_{Map} w\right) + B_k \tag{5}$$

Convolution is an operation that is formed by combining one or more of these kernels in various ways. In the suggested model, each layer of convolution is followed by a batch normalization layer, and the activation function is a leaky version of the rectified linear unit (ReLU). The leaky ReLU is a modified version of the rectified linear unit (ReLU) activation function that allows a small gradient for negative inputs, addressing the "dying ReLU" problem where some neurons become inactive.

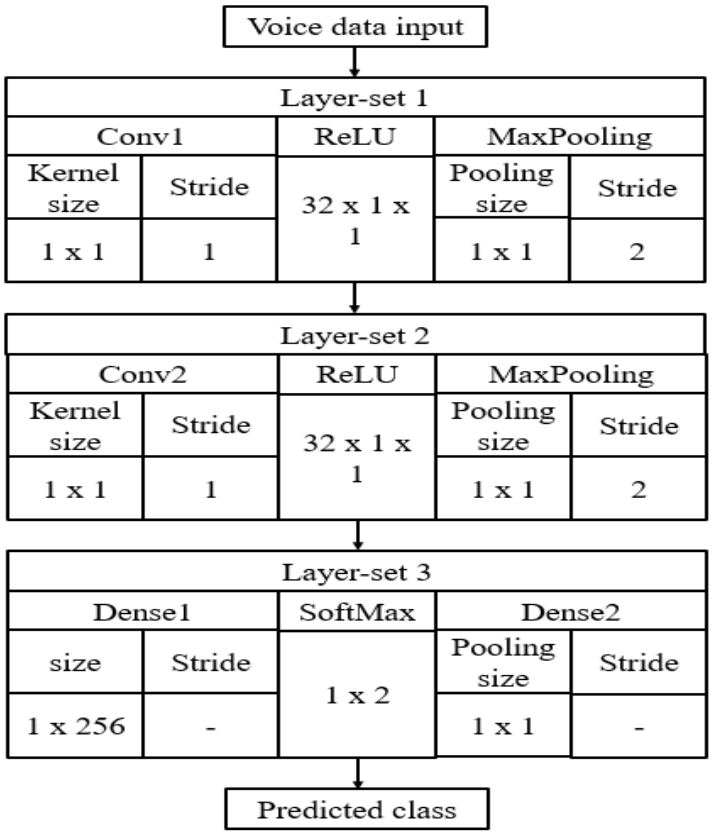


Figure 3: Voice-CNN model.

ReLU activation function: Within a neural network, the activation functions also perform the duties of the transfer functions. The findings of the layer before this one is altered by these layers so that they are mapped onto the information that was provided as the ground truth. There are two different types of activation functions, namely the linear activation function and the nonlinear activation function. Activation in VIE-CNNs is often accomplished via the use of a variety of nonlinear function types. It is possible to activate VIE-CNNs (View-Invariant Embedding Convolutional Neural Networks), which enable the network to recognize and represent a variety of features and patterns across various views or perspectives of the input data. In most cases, these functions are included so that the nonlinearity idea is preserved inside the network. For VIE-CNNs to be able to recognize complex and nonlinear patterns in the data, nonlinearity must be included. The model's ability to learn complex associations is improved by nonlinear activation functions, which also improves the model's overall performance when handling real-world data. ReLU is a linearly rectified function. In the absence of negative input, the output of the ReLU function is always zero; in all other cases, the input is left unaltered (see equation 2). During backpropagation, the model parameters are updated based on input values that are not negative. Because of this, the dying ReLU problem arises, and the leaky ReLU activation function has been implemented in our network to solve this problem. The "dying ReLU problem" in neural networks occurs when ReLU activations always output zero for particular inputs, rendering neurons inactive during training. Gradient updates are interfered with, training is slowed down, and learning of complicated features is constrained. Leaky ReLU and related variations address issues by permitting modest gradients for negative inputs, improving training effectiveness and network expressiveness. In this case, the negative slope does not have a value of zero and instead has a tiny value; as a result, its derivative will have nonzero values for any data that is supplied. The equation that corresponds to the mathematical representation is provided by equation (2), while the equation that corresponds to the function's derivatives is given by equation (3).

$$ReLU(z) = \max\{z, 0\} \tag{3}$$

$$f(z) = \begin{cases} z, z \geq 0, \\ \propto z, z < 0. \end{cases} \tag{4}$$

MaxPooling layer: To broaden the scope of the network's receptive field, the MaxPooling layer is implemented. This process brings down the total cost of computing while also decreasing the size of the feature maps in terms of their spatial dimensions. Only the height and breadth of the supplied data are reduced by this process. There is no change to the total number of feature channels. It is similar to the method of using sliding windows with the maximum element operation selected. The size decrease achieved is proportional to the length of the sliding operation's stride. The pooling operation in the proposed network makes use of a pooling size and takes strides to accomplish its goals. Since the pooling layer is nonparametric, there are no learning parameters at this stage.

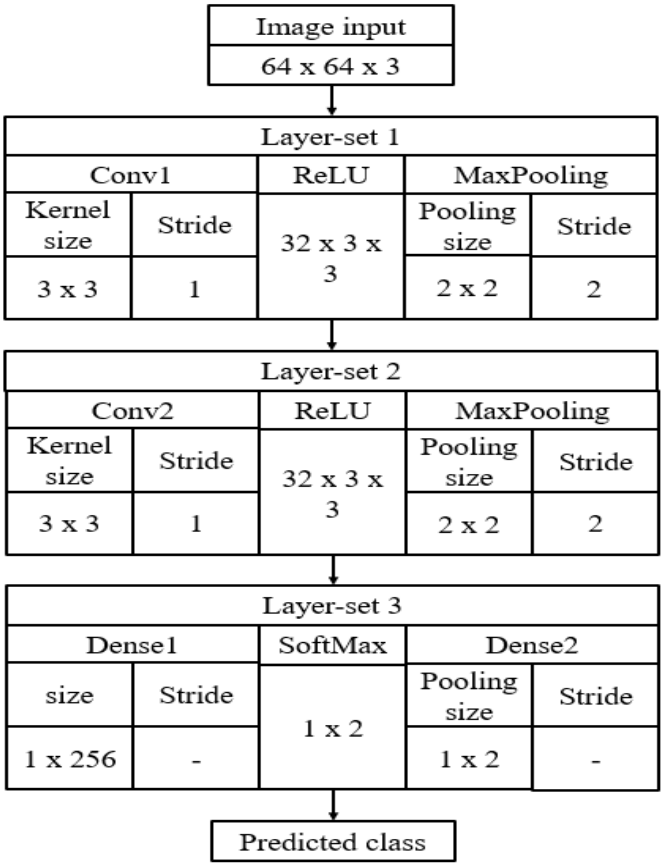


Figure 4: Image-CNN model.

Dense layer: The research has shown that the training of VIE-CNNs is a difficult process that involves a variety of hyperparameters. In most cases, the computational network of a deep neural model has a larger depth, which results in the phenomenon known as the convergence issue. Hyperparameter adjustment and convergence offer difficulties while training VIE-CNNs. The intricacy of the model makes it challenging to choose the appropriate hyperparameters and reach convergence; this frequently necessitates careful initialization, large amounts of data, and powerful computing power. There are a few solutions that have been offered to remedy this problem. The dense layer is used in the suggested model to manage the convergence issue and speed up the training of the network. The dense layer is usually placed immediately after the activation layers. After activation layers, adding a dense layer aids the network in learning complicated representations from the activated features, allowing it to recognize complex patterns in the input.

SoftMax classifier: SoftMax Classifier is also an activation function for activating hyperbolic tangents. It's a form of logistic sigmoid activation function, and it has a significant meaning in terms of the biological neurons. One of the most distinguishing features of a SoftMax tangent function is that its upper derivatives disappear as they approach zero. This is because the SoftMax tangent function keeps its acceptable property of learning discriminative features from a greater variety of varied data samples. The SoftMax activation function has the feature of yielding its normalized score in the range of the output scale. Finally, the SoftMax classifier of the VIE-CNN model is used to classify the PD and normal classes, which compares the test probabilities with trained memory. Probabilities provide a measure of confidence in a machine learning model's classification, influencing the final decision by allowing the adjustment of threshold values for precision and recall trade-offs. The PD (Parkinson's disease) and normal classes in the VIE-CNN model are classified using the SoftMax classifier by turning the raw scores from the last layer into probabilities. It gives each class a probability distribution,

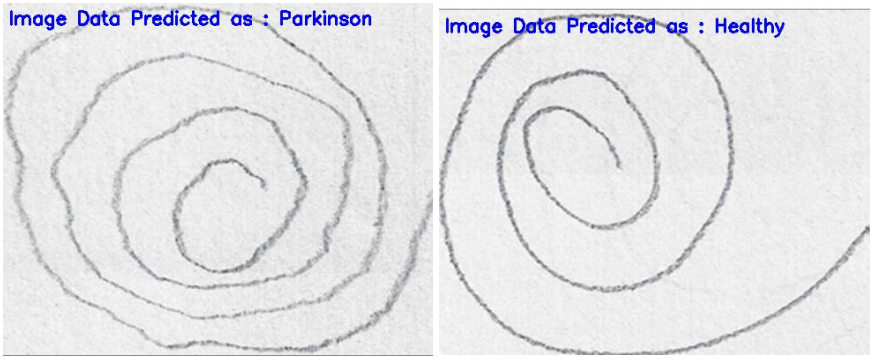
enabling the model to determine which class is most likely given an input.

4 Results and discussion

This section gives a detailed performance analysis of the proposed ParkinsonNet. The performance of the proposed method is measured using several performance metrics, such as accuracy, sensitivity, specificity, F1-measure, and precision. All these metrics are measured for proposed methods as well as existing methods. Then, all the methods use the same dataset for performance estimations.

4.1 Prediction performance evaluation

Figure 5 shows the predicted outcomes using the proposed ParkinsonNet. Figure 5 (a) shows that the outcome is Parkinson's. Here, the two probabilities are generated as 0.9 and 0.1, where the first position contains the maximum value, so the output is detected as Parkinson. Figure 5(b): A healthy outcome is detected. Clinical symptoms, medical history, diagnostic testing (such as imaging and blood tests), and professional medical evaluation are all taken into account to interpret probability and determine with certainty whether a patient has Parkinson's disease or a healthy outcome. Although a final diagnosis is normally made by healthcare experts based on a thorough evaluation of all available data, machine learning algorithms may use these elements to provide probability scores. Here, the two probabilities are generated as 0.2 and 0.8, where the second position contains the maximum value, so the output is detected as healthy.



(a) (b)

Figure 5: Predicted outcomes using ParkinsonNet.

The accuracy and loss values are shown along the y-axis, while the training epoch is shown along the x-axis in Figure 6. We can see that as the number of training epochs rose, the accuracy improved while the loss reduced, and that by the time the training was finished, the accuracy was coming closer to 1 and the loss was getting closer to 0. The graph that can be seen above shows that the blue line in the graph denotes the accuracy of the image, the red line denotes the accuracy of the voice, the green line denotes the loss of the image, and the yellow line denotes the loss of the voice. Table 1 compares the performance of various methods on the voice dataset. Here, the ParkinsonNet resulted in improved performance over various existing methods, such as KNN [19], ANN-SVM [20], and Adaboost [23]. Here, the ParkinsonNet has increased accuracy by 5.74%, sensitivity by 7.61%, specificity by 6.23%, F-measure by 2.26%, precision by 7.24%, and Mathew's correlation coefficient (MCC) by 3.23%, as compared to the KNN [19]. In the third column, the ParkinsonNet has an accuracy of 4.52%, sensitivity of 3.99%, specificity of 6.42%, F-measure of 6.56%, precision of 2.41%, MCC by 1.08% as compared to ANN-SVM [20]. In the last column, ParkinsonNet has an accuracy of 8.29%, sensitivity of 2.34%, specificity of 0.90%, F-measure of 2.75%, precision of 5.71%, and MCC of 0.92% as compared to Adaboost [23].

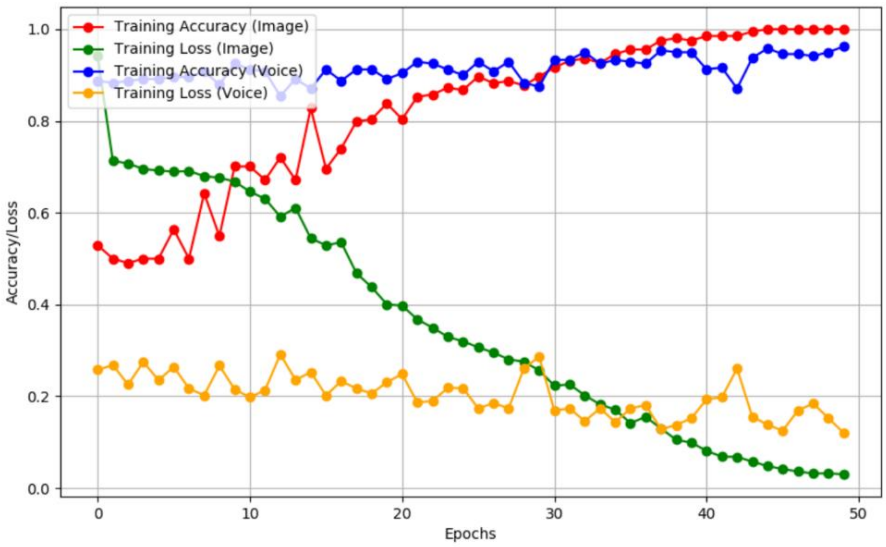


Figure 6: Accuracy and Loss outcomes using ParkinsonNet.

Table 1: Performance comparison of various methods on voice dataset.

Metric	KNN [19]	ANN-SVM [20]	Adaboost [23]	Proposed ParkinsonNet
Accuracy (%)	92.84	93.92	90.65	98.17
Sensitivity (%)	91.75	94.95	96.48	98.74
Specificity (%)	91.72	91.56	96.57	97.44
F-measure (%)	95.69	91.83	95.24	97.86
Precision (%)	91.04	95.34	92.36	97.64
MCC (%)	94.38	96.38	96.54	97.43

Table 2 compares the performance of various methods on the image dataset. Here, the ParkinsonNet resulted in improved performance over various existing methods, such as ANN [13], ISVM [15], and UPDRS [17]. Here, the ParkinsonNet has increased accuracy by 3.47%, sensitivity by 3.89%, specificity by 3.37%, F-measure by 6.56%, precision by 4.35%, and MCC by 0.76% as compared to the ANN [13]. In the third column, the ParkinsonNet has increased accuracy by 3.67%, sensitivity by 5.84%, specificity by 7.45%, F-measure by 4.23%, precision by 5.43%, MCC by 0.76% as compared to ISVM [15]. In the last column, the ParkinsonNet has increased accuracy by 9.09%, sensitivity by 6.88%, specificity by 6.65%, F-measure by 1.33%, precision by 3.04%, MCC by 1.52% as compared to UPDRS [17].

Table 2: Performance comparison of various methods on image dataset.

Metric	ANN [13]	ISVM [15]	UPDRS [17]	Proposed ParkinsonNet
Accuracy (%)	93.89	92.16	96.27	99.65
Sensitivity (%)	95.69	92.06	97.75	99.12
Specificity (%)	91.94	94.00	96.69	98.45
F-measure (%)	94.31	93.35	95.51	98.54
Precision (%)	96.98	92.50	96.25	99.77
MCC (%)	93.89	92.16	91.27	99.67

5 Conclusion

This work implemented the ParkinsonNet using the VIE-CNN model. In the beginning, a combined dataset that includes both the speech and picture samples is taken into consideration. Next, a data processing operation is carried out to normalize the pictures to a size that is consistent throughout the board. This data processing operation also

carries out a data balancing operation inside the voice dataset. Following that, a VIE-CNN model is trained using the pre-processed voice-image data. In this instance, CCE is used to optimize the losses that are sustained during the training. The VIE-CNN model is developed by combining Voice-CNN and Image-CNN, where an ensemble model is used to predict PD from both test samples. After that, the VIE-CNN model uses the test data to make predictions about the normal and abnormal classes. Here, the ParkinsonNet improved accuracy by 3.4% on image data and 2.36% on voice datasets as compared to existing methods. Further, this work can be extended with advanced transfer learning methods for multi-class PD identification.

Acknowledgements

Funding: The authors did not receive any funding.

Conflicts of Interests: Authors do not have any conflicts.

Data Availability Statement: No datasets were generated or analyzed during the current study.

Code availability: Not applicable.

Authors' Contributions: All authors contributed to the design and methodology of this study, the assessment of the outcomes, and the writing of the manuscript.

References

[1] Zhang, H., Song, C., Wang, A., Xu, C., Li, D., & Xu, W (2019). Pdvocal: Towards privacy-preserving Parkinson's disease detection using non-speech body sounds. *In The 25th annual international conference on mobile computing and networking*, Los Cabos, Mexico, pp. 1-16. https://doi.org/10.1145/3300061.3300125.

[2] Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Acharya, U. R (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in biology and medicine*, pp. 103792. https://doi.org/10.1016/j.compbiomed.2020.103792.

[3] Nilashi, M., Ibrahim, O., Samad, S., Ahmadi, H., Shahmoradi, L., & Akbari, E (2019). An analytical method for measuring the Parkinson's disease progression: A case on a Parkinson's telemonitoring dataset. *Measurement*, pp. 545-557. https://doi.org/10.1016/j.measurement.2019.01.014.

[4] Zhang, H (2022). Smart Health Technologies Towards Parkinson's Disease Management, *Doctoral dissertation, State University of New York at Buffalo.*

[5] Dodel, R., Tinelli, M., Deuschl, G., Petersen, G., Oertel, W., & Ahmerkamp-Böhme, J (2021). The economic benefit of timely, adequate, and adherence to Parkinson's disease treatment: the Value of Treatment Project 2. *European Journal of Neurology*, pp. 707-716. https://doi.org/10.1111/ene.14584.

[6] Warnecke, T., Schäfer, K. H., Claus, I., Del Tredici, K., & Jost, W. H (2022). Gastrointestinal involvement in Parkinson's disease: pathophysiology, diagnosis, and management. *npj Parkinson's Disease*, pp. 31. https://doi.org/10.1038/s41531-022-00295-x.

[7] Roberts, A.C., Rafferty, M.R., Wu, S.S., Miao, G., Cubillos, F., Simuni, T., Marras, C., Davis, T., Dahodwala, N., Neault, M. and Ramirez-Zamora, A (2021). Patterns and predictors of referrals to allied health services for individuals with Parkinson's disease: A Parkinson's foundation (PF) QII study. *Parkinsonism & related disorders*, pp.115-122. https://doi.org/10.1016/j.parkreldis.2020.11.024.

[8] Chen, K.K., Jin, Z.H., Gao, L., Qi, L., Zhen, Q.X., Liu, C., Wang, P., Liu, Y.H., Wang, R.D., Liu, Y.J. and Fang, J.P (2021). Efficacy of short-term multidisciplinary intensive rehabilitation in patients with different Parkinson's disease motor subtypes: a prospective pilot study with 3-month follow-up. *Neural Regeneration Research*, pp.1336-1343. https://doi.org/10.4103%2F1673-5374.301029.

[9] Gandhi, P., & Steele, C. M (2022). Effectiveness of interventions for dysphagia in Parkinson's disease: a systematic review. *American Journal of Speech-Language Pathology*, pp. 463-485. https://doi.org/10.1044/2021_ajslp-21-0014.

[10] van Hooren, M. R., Baijens, L. W., Dijkman, R., Kremer, B., Michou, E., Pilz, W., & Vos, R (2022). The effects of surface electrical stimulation plus voice therapy in Parkinson's disease. *The Journal of Laryngology & Otology, pp.* 1-7. https://doi.org/10.1017/s0022215122002031.

[11] ul Haq, A., Li, J. P., Agbley, B. L. Y., Mawuli, C. B., Ali, Z., Nazir, S., & Din, S. U (2022). A survey of deep learning techniques based on Parkinson's disease recognition methods employing clinical data. *Expert Systems with Applications*, pp. 118045. https://doi.org/10.1016/j.eswa.2022.118045.

[12] Pereira, C. R., Pereira, D. R., Rosa, G. H., Albuquerque, V. H., Weber, S. A., Hook, C., & Papa, J. P (2018). Handwritten dynamics assessment through convolutional neural networks: An application to Parkinson's disease identification. *Artificial intelligence in medicine*, pp. 67-77. https://doi.org/10.1016/j.artmed.2018.04.001.

[13] Tanveer, M., Rashid, A. H., Kumar, R., & Balasubramanian, R (2022). Parkinson's disease

diagnosis using neural networks: Survey and comprehensive evaluation. *Information Processing & Management*, pp. 102909. https://doi.org/10.1016/j.ipm.2022.102909.

[14] Zuckerkandl, E., & Pauling, L (1964). Evolutionary Divergence and Convergence in Proteins. *Evolving Genes and Proteins*, pp. 97-166. https://doi.org/10.1016/B978-1-4832-2734-4.50017-6.

[15] Nilashi, M., Ibrahim, O., Ahmadi, H., Shahmoradi, L., & Farahmand, M (2018). A hybrid intelligent system for the prediction of Parkinson's Disease progression using machine learning techniques. *Biocybernetics and Biomedical Engineering*, pp. 1-15. https://doi.org/10.1016/j.bbe.2017.09.002.

[16] Gómez-Vilda, P., Mekyska, J., Ferrández, J.M., Palacios-Alonso, D., Gómez-Rodellar, A., Rodellar-Biarge, V., Galaz, Z., Smekal, Z., Eliasova, I., Kostalova, M. & Rektorova, I (2017). Parkinson disease detection from speech articulation neuromechanics. *Frontiers in neuroinformatics*, pp.56. https://doi.org/10.3389/fninf.2017.00056.

[17] El Maachi, I., Bilodeau, G. A., & Bouachir, W (2020). Deep 1D-Convnet for accurate Parkinson's disease detection and severity prediction from gait. *Expert Systems with Applications*, pp. 113075. https://doi.org/10.1016/j.eswa.2019.113075.

[18] Johri, A., & Tripathi, A (2019). Parkinson's disease detection using deep neural networks. *In 2019 twelfth International Conference on Contemporary Computing (IC3)*, Noida, India, pp. 1-4. https://doi.org/10.1016/j.eswa.2019.113075.

[19] Pahuja, G., & Nagabhushan, T. N (2021). A comparative study of existing machine learning approaches for Parkinson's disease detection. *IETE Journal of Research*, pp. 4-14. https://doi.org/10.1080/03772063.2018.1531730.

[20] Solana-Lavalle, G., Galán-Hernández, J. C., & Rosas-Romero, R (2020). Automatic Parkinson's disease detection at early stages as a pre-diagnosis tool by using classifiers and a small set of vocal features. *Biocybernetics and Biomedical Engineering*, pp. 505-516. https://doi.org/10.1016/j.bbe.2020.01.003.

[21] Zhang, H., Song, C., Rathore, A. S., Huang, M. C., Zhang, Y., & Xu, W (2020). mhealth technologies towards Parkinson's disease detection and monitoring in daily life: A comprehensive review. *IEEE reviews in biomedical engineering*, pp. 71-81. https://doi.org/10.1109/rbme.2020.2991813.

[22] Alzubaidi, M. S., Shah, U., Dhia Zubaydi, H., Dolaat, K., Abd-Alrazaq, A. A., Ahmed, A., & Househ, M (2021). The role of neural network for the detection of Parkinson's disease: a scoping review. *In Healthcare*, p. 740. https://doi.org/10.3390/healthcare9060740.

[23] Ali, L., Zhu, C., Golilarz, N. A., Javeed, A., Zhou, M., & Liu, Y (2019). Reliable Parkinson's disease detection by analyzing handwritten drawings: construction of an unbiased cascaded learning system based on feature selection and adaptive boosting model. *Ieee Access*, pp. 116480-116489. https://doi.org/10.1109/ACCESS.2019.2932037.

[24] Wodzinski, M., Skalski, A., Hemmerling, D., Orozco-Arroyave, J. R., & Nöth, E (2019). Deep learning approach to Parkinson's disease detection using voice recordings and convolutional neural network dedicated to image classification. *In 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, pp. 717-720. https://doi.org/ 10.1109/EMBC.2019.8856972.

[25] Quan, C., Ren, K., & Luo, Z (2021). A deep learning-based method for Parkinson's disease detection using dynamic features of speech. *IEEE Access*, pp. 10239-10252. https://doi.org/10.1109/ACCESS.2021.3051432.

[26] Thirugnanam, T., Galety, M. G., Pradhan, M. R., Agrawal, R., Shobanadevi, A., Almufti, S. M., & Lakshmana Kumar, R. (2023). PIRAP: Medical Cancer Rehabilitation Healthcare Center Data Maintenance Based on IoT-Based Deep Federated Collaborative Learning. *International Journal of Cooperative Information Systems*, 2350005. https://doi.org/10.1142/S0218843023500053.

[27] Ali, M. H., Jaber, M. M., Alfred Daniel, J., Vignesh, C. C., Meenakshisundaram, I., Kumar, B. S., & Punitha, P. (2023). Autonomous vehicles decision-making enhancement using self-determination theory and mixed-precision neural networks. *Multimedia Tools and Applications*, pp. 1-24. https://doi.org/10.1007/s11042-023-14375-4.

DRG-Net: Diabetic Retinopathy Grading Network Using Graph Learning with Extreme Gradient Boosting Classifier

*Venkata Kotam Raju Poranki, B. Srinivasa Rao
[1]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India
E-mail: vkrporanki@gmail.com
*Corresponding author

Keywords: color eye fundus images, Diabetic retinopathy, graph convolutional neural network, synthetic minority over-sampling technique, extreme gradient boosting

Received: August 6, 2023

Diabetic retinopathy (DR) is a leading cause of blindness that occurs in different age groups. So, the early detection of DR can save millions of people from blindness issues. Further, the manual analysis of DR requires much processing time and experienced doctors. Hence, computer-aided diagnosis (CAD)-based artificial intelligence models have been developed for an early DR prediction. However, the state-of-the-art methodologies failed to extract the deep balanced features, which resulted in poor classification performance. Therefore, this work implements the DR grading network (DRG-Net) using graph learning properties. Initially, the synthetic minority over-sampling technique (SMOTE) is applied to the EyePACS and Messidor dataset to balance the instances of each DR class into uniform levels. Then, a deep graph correlation network (DGCN) is applied to extract the class-specific features by identifying the relationship. Finally, an extreme gradient boosting (XGBoost) classifier is employed to perform the DR classification with the pre-trained balanced features obtained using SMOTE-DGCN. The obtained simulation results performed on the EyePACS dataset and the Messidor dataset disclose that the proposed DRG-Net resulted in higher performance than state-of-the-art DR grading classification approaches, with accuracy, sensitivity, and specificity of 99.01%, 99.01%, and 98.43% for the EyePACS dataset, respectively, and 99.6%, 99.08%, and 100% for the Messidor dataset.

Povzetek: Članek opisuje novo metodo DRG-Net, ki uporablja grafovno učenje in ekstremno gradientn spodbujevalno učenje za zgodnje odkrivanje diabetične retinopatije.

1 Introduction

Humans with DR pose a danger to their eyesight is expected to total 103.12 million besides 28.54 million, respectively, in the year 2021. The numbers are expected to rise to 160.50 million and 44.82 million by 2045 [1]. Additionally, in underdeveloped nations without access to basic healthcare facilities or a scarcity of ophthalmologists. The underserved regions of the developed world are likewise affected by this issue. Thus, early detection and routine screening may cut the chance of vision loss to 57.0% while also lowering treatment costs. The two major processes in DR detection are screening and diagnosis [2]. Fine pathognomonic DR signals are generally established for this purpose following dilation of pupils (mydriasis). For accurate diagnosis, early identification, patient education, and treatment planning, clear pathognomonic signals are essential. This improves healthcare workers' capacity to recognize and treat diabetic retinopathy. Then, DR screening is performed using slit lamp bio-microscopy with a + 90.0 D lens and direct indirect ophthalmoscopy [3]. Finding lesions associated with DR and comparing them to the criteria for the grading system allows one to diagnose DR. Figure 1 depicts the normal retinal vision and DR-related anomalies. Diabetes complications, such as diabetic macular edema (DME) and DR, are more common in those who are working age.

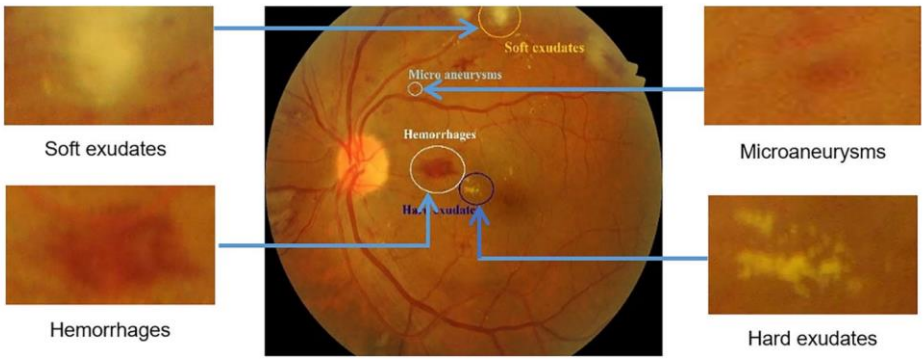


Figure 1: Normal retinal image and abnormalities.

The complicated condition known as DR makes the retinal veins expand and leak blood and fluid [5]. DR causes an issue with visual impairment. Microvascular damage, blood vessel irregularities, inflammation, VEGF, ischemia, compromised blood-fluid barrier, hyperglycemia-induced metabolic alterations, and oxidative stress are some of the variables that contribute to diabetic retinopathy as well as produce retinal vein expansion and leaking. Effective electroretinography transformation, retinal vein type, and retinal blood flow may be used to provide a prompt identification of DR. Clear pathognomonic signals enhance patient care and management by improving diagnostic processes, communication, patient education, severity evaluation, treatment planning, early diagnosis, progression monitoring, and research breakthroughs for diabetic retinopathy. For the early diagnosis of diabetic retinopathy, electroretinography (ERG) is essential because it can spot tiny retinal changes before symptoms manifest. It aids in the distinction of illness phases, forecasts the likelihood of progression, tracks alterations, supports research, and directs individualized therapies. Diabetes patients are more prone to acquire DR if they experience unfavorable symptoms of the condition over an extended length of time [6]. During patients to be examined and during the initial phase of DR therapy to limit the possibility of vision loss, routine retinal screening is essential. DR can be identified by the presence of several signals on a retinal picture. Additionally, these DR symptoms include hard exudates (EX), hemorrhages (HM), soft exudates (EX), and microaneurysms (MA). A diagnostic test used to assess retinal reactions to light, the electroretinogram (ERG) looks for abnormalities that might be signs of diabetic retinopathy (DR). It aids with early discovery, DR severity evaluation, and prompt action. Based on the presence of these symptoms, the DR is divided into five phases: severe DR, no DR, moderate DR, mild DR, and proliferative DR (PDR), which are succinctly described in Table 1 [7]. Non-proliferative DR (NPDR) is used here to refer to all types of DR, including none, mild, moderate, and severe DR. Close examination and monitoring are crucial for patients in the mild stage of DR because they have at least one MA; for patients in the moderate stage who may have HM or MAs in one to three retinal quadrants; for patients in the severe stage who have intraretinal HM and venous beading in two or more quadrants; and for patients in the proliferative stage who need to see a doctor every six to eight months.

Table 1: Various grades of DR are based on signs.

DR severity grading	Class	Signs
No DR	DR-0	Absence of various problems.
Mild DR	DR-1	The presence of MA.
Moderate DR	DR-2	Increase in MA count
Severe DR	DR-3	The intraretinal HM count raises more than twenty. Several quadrants of the vena cava show beading.
Proliferative DR	DR-4	Significant intraretinal microvascular problem with neovascularization in more than one quadrant. Pre-retinal and vascular HM.

The diagnosing stage is currently carried out manually. This technique is expensive, time-consuming, and calls for professionals who have undergone extensive training and have excellent diagnostic skills. Even with all

these tools at hand, a misdiagnosis is still a possibility. The situation is difficult because of this reliance on manual review. Among the often-employed DR grading systems is the Early Treatment Diabetic Retinopathy Study (ETDRS) [10]. The ETDRS isolates highly precise DR characteristics using several layers. This system of grading applies to all seven retinal fundus Fields of View (FOV). Early diagnosis of eye disorders, monitoring health changes, avoiding vision loss, assessing systemic health issues, individualized treatment planning, patient education, and enhanced quality of life are all benefits of routine retinal exams. Even though ETDRS is the gold standard, due to implementation difficulty and technological limitations, alternate systems for grading are also used. One such system is the International Clinical Diabetic Retinopathy (ICDR) [11] scale, which is recognized in both clinical and Computer-Aided Diagnosis (CAD) environments. However, the conventional CAD methods [12] failed to result in maximum performance. So, this work is focused on the development of the following models:

- The SMOTE method is adopted to perform the balancing of instances presented in EyePACS and Messidor datasets, which balances the instances of each DR class into a uniform level.
- The DGCN i.e., graph learning method is developed, and used to extract the class-specific features by identifying the relationship.
- The XGBoost classifier is trained with balanced features and performs the multi-class prediction operation.

The remainder of the text is structured as follows: The overview of DR grading and detection techniques is included in section 2. The planned DRG-Net's analysis is covered in Section 3 along with a description of the sub-modules. The performance comparison with cutting-edge techniques is covered in Section 4 along with a full examination of the simulation results suggested by DRG-Net. The item is concluded in Section 5 with potential future application.

2 Related work

This section gives a detailed analysis of DR grading methods. In [12] authors presented the deep convolutional neural network (DCNN) model for DR classification. Dataset preparation begins with phases for data collecting and data annotation. Furthermore, a median filter was used for data pre-processing. Additionally, DR diagnosis is carried out using DCNN, which categorizes the classes as Normal, Moderate, Heavy, and Severe. This technique is also used in hospitals to provide patients with online services. In [13] authors explored an issue of automatic DR identification and suggested a new deep-learning hybrid to handle. It builds the hybrid model by adding a custom block of CNN layers on top of the pre-trained Inception-ResNet-v2 and using transfer learning on that network.

In [14] authors used single color fundus image and suggested a novel automated deep learning-based technique for severity identification. The suggested method is to build a visual embedding using the DenseNet169 encoder. A convolutional Block Attention Module (CBAM) is also added on top of the encoder to boost its ability to discriminate. To improve network performance and model comprehension, the Convolutional Block Attention Module (CBAM) combines channel and spatial attention processes to boost feature identification in convolutional neural networks. However, this method has high loss values compared to machine learning models. In [15] authors used a fine-tuned ResNet-50, which is developed using two-stage deep learning models. The architecture can effectively classify DR into one of three categories (normal, moderate DR, and severe DR). Further classification of moderate DR was performed using a fine-tuned ResNet-18 [15], while classification of severe DR was performed using a fine-tuned ResNet-50.

In [16] authors proposed the creation and Optimization CNN (OCNN) Model. The pre-trained OCNN model is initially used to extract the features. A classification layer based on gradient boosting is added to enhance features. The examination of the suggested system using a 10-fold cross-validation on two difficult problems shows that it performs better than cutting-edge approaches. In [17] authors proposed DR early detection based on multifractal geometry. Furthermore, automating the diagnostic process and increasing the resulting accuracy by employing a supervised machine learning technique like the Support Vector Machine (SVM) algorithm. In [18] authors proposed a new technique to identify DR using ensemble recurrent neural networks (ERNN). First, a two-stage classifier that uses an assembly approach to merge several machine learning algorithms for categorization. By merging separate predictions and applying a meta-classifier for final judgment, the two-stage classifier assembly strategy, which incorporates numerous machine learning algorithms, improves categorization accuracy. The classifier is also used with DR. So, the issue is that it takes a long time to diagnose this condition, even though an early diagnosis is necessary to prevent total blindness. In [19] authors proposed an

early detection of DR using Principal Component Analysis (PCA) with Firefly optimization. Then, deep learning models were used to perform grading operations. Here, the raw dataset is first normalized using the standard scalar approach, and the most important characteristics in the dataset are then extracted using PCA. Additionally, the Firefly algorithm is used to reduce the number of dimensions. In [20] authors developed the deep belief neural network (DBNN) to predict DR. Here, the Grey Wolf Optimization (GWO) method is used to optimize the PCA features. Additionally, using GWO makes it possible to choose the best training parameters for the DBNN model.

Various machine learning models like support vector machine (SVM) [21], decision tree, and random forest, are used to classify the DR grades from eye fundus images. Eye fundus pictures are entered into a connected graph approach, which represents them as a graph with nodes and edges. With the use of Graph Neural Networks, which analyze the graph and extract characteristics and contextual data, eye ailment diagnoses are now more accurate. Further, the application of different pre-trained convolutional neural network (CNN) [22] architectures was implemented to identify the DR from fundus pictures with rejection resampling (random under-sampling at mini-batch level) technique for effective performance in tackling different imbalanced scenarios of varied dataset sizes. To manage unbalanced datasets, a variety of strategies are utilized, including resampling, algorithm-level, algorithmic, synthetic data creation, transfer learning, and hybrid methods. Experimentation is necessary for the best outcomes. Further, the classification of DR using the Ensemble of Machine Learning (EML) [23] classifier is implemented. However, this method contains higher computational complexity. Further, three-dimensional semantic segmentation of DR lesions with grading operation is implemented using ResNet [24] based transfer learning. The marine predictor algorithm (MPA) is used to select the best features from available features. To address these shortcomings of traditional deep learning models, the explainable and interpretable diabetic retinopathy (ExplainDR) [25] grading approach was developed. Here, neural-symbolic learning is introduced for class-specific feature extraction. Then, a model called the Hinge Attention Network (HA-Net) [26] was created for a thorough study of several DR grades. The HA-Net uses the VGG16 model [27] for feature extraction and grading done by long short-term memory (LSTM). Then, to improve the DR findings, the residual attention network (RAN) [28] was invented. The attention mechanism performs the dilated convolution for feature analysis. In addition, DR diagnosis is carried out using transfer learning based InceptionResNetV2 [29], Xception, and EfficientNetB3 model. Here, InceptionResNetV2 resulted in superior performance. Then, Mask Region-based CNN (MRCNN) [30] was utilized instead of faster region-based CNN (Faster RCNN) and transferred learning model to classify DR grades [31].

3 Proposed system

This section gives a detailed analysis of the proposed DRG-Net. The DRG-Net is a hybrid learning model, that is developed by the properties of graph learning, deep learning, and machine learning models. Figure 2 shows the block diagram of the proposed DRG-Net. However, the EyePACS and Messidor datasets contain the five classes of DR, i.e., DR-0, DR-1, DR-2, DR-3, DR-4. However, the number of images in each grade is uniform, which results in imbalanced features. So, the SMOTE method is applied to balance the number of instances in each class, which maintains an equal number of features. Then, the DR contains the highly correlated features among PDR and NPDR classes. So, DGCN is applied to extract the DR grade-specific features by adopting the graph interconnections of features from SMOTE balanced features. Then, the XGBoost classifier is trained with DGCN balanced features and performs classification of multiple DR grades. Due to its excellent predictive strength and capacity to recognize complicated patterns, the Deep Graph Convolutional Network-trained XGBoost classifier excels at categorizing different diabetic retinopathy grades, although its success is reliant on factors like data variety and practitioner knowledge.

3.1 SMOTE data balancing

The dataset's irregular sizes hurt the performance of deep learning classifiers. The datasets have distinct, imbalanced numbers in them. Therefore, most of the DR class features will be given the highest probabilities (priorities) by the XGBoost classifier, while the minority DR class features will be ignored. Using weighted loss functions, sampling techniques, and adaptable goal functions, XGBoost applies ways to ensure equitable minority class characteristics in the classification of unbalanced datasets. Misclassifications are minimized, and problems are addressed. This may lead to inaccurate categorization and inaccurate forecasting. The block diagram for the SMOTE data balancing approach is shown in Figure 3. It is an oversampling strategy that consistently raises the number of minority-class records. The oversampling procedure in this case generates duplicate samples using the closest possible combinations. The Synthetic Minority Over-Sampling Technique (SMOTE), which ensures variety while keeping patterns and carefully

chooses k nearest neighbors, is used in oversampling to create duplicated samples from minority class instances.

The SMOTE method determines the overall dataset record count as well as the record count for every class. SMOTE is a strategy that enhances model performance by creating synthetic records for minority classes, recognizing them, figuring out the oversampling ratio, and arbitrarily choosing the nearest neighbors. Unbalanced data can impair DRG-Net's effectiveness by introducing biases, lowered sensitivity, false negatives, and overfitting. To create synthetic minority class samples, SMOTE is essential. The average record value will then be calculated. Then, SMOTE uses the notion of randomization to create KNNs for every data. Consider the dataset with X minority class samples. Then, $y_1, y_2, \cdots, and\ y_N$ represent the N showed samples for each minority class. For model performance, SMOTE seeks to provide a balanced dataset; however, normalizing the dataset to equal samples defeats this goal. It is advised to preserve or lessen the class disparity when taking SMOTE. So, there must be more than N-number of KNN samples, i.e., $K > N$. Following this, using the correlation among records, random samples were added to the database. The random correlation factor p_i is constructing the interpolation operation.

$$p_i = X + rand(0,1) \times (y_i - X), \qquad i = 1,2,\cdots,N \tag{1}$$

Here, it is expected that the random integer will fall between [0,1]. The y_i denotes the nearest neighbor for sample i, the X represents data records presented in each minority class. By performing the rounding operation to imbalanced minority class X_{IL}, dataset balance was achieved. Due to biases, restricted minority learning, and other problems, deep learning classifiers have difficulty with unbalanced dataset sizes. Oversampling and under-sampling are rebalancing approaches that enhance the accuracy of classifiers for both classes.

$$X_B = round(X_{IL}) * p_i \tag{2}$$

Here, X_B represents the balanced dataset. Lastly, SMOTE equalizes the whole feature set by inserting these random records into the minority class determined by KNN resemblance.

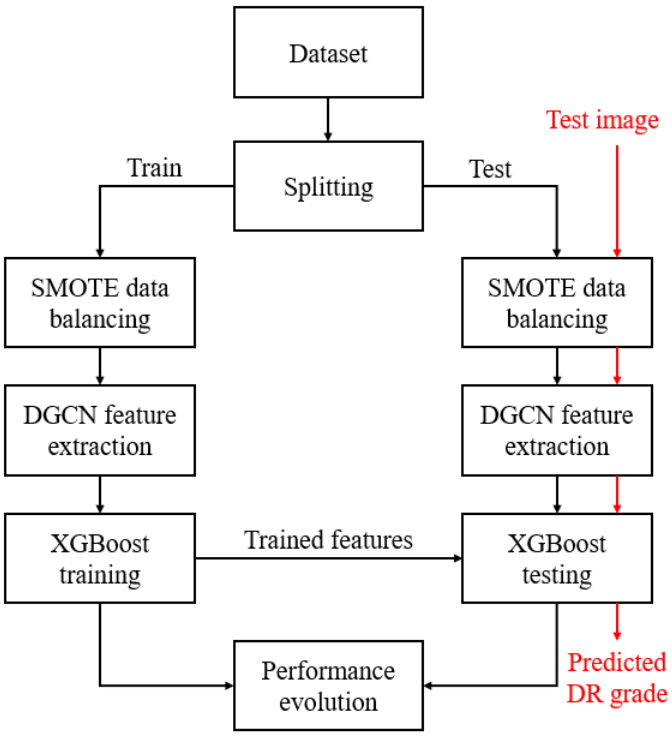


Figure 2: Proposed DRG-Net block diagram.

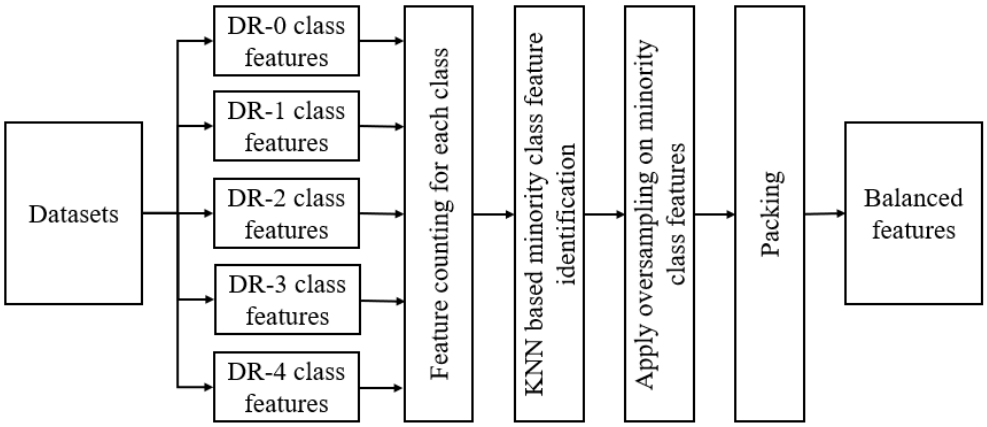


Figure 3: Block diagram of data balancing with SMOTE.

3.2 DGCN-based feature extraction

Graph neural networks (GNNs) have made considerable strides in several practical applications recently, including drug discovery, medical image processing, and recommendation. The graph structure of the GNN technique for recognizing diabetic retinopathy (DR) makes use of both labeled and unlabeled data. By using data relationships, managing unlabeled data, and scaling effectively, it improves accuracy. Despite their success, deep GNNs' performance is still constrained by several fundamental factors, one of which is over-smoothing. In deep Graph Neural Networks (GNNs), over-smoothing impairs generalization, reduces local structure capacity, impairs discrimination, and makes it difficult to capture higher-order connection patterns. Techniques like attentional processes and residual connections help to solve this problem. It shows that the stacked aggregators make the learned node representations almost impossible to discern. The performance deterioration of deep GNNs or their preference for feature correlation issues is examined. Performance problems with Deep Graph Neural Networks (GNNs) are brought on by feature correlation that results in noise, redundancy, and instability. To address these problems, preprocessing, feature selection, normalization, and regularisation approaches are required. To reduce the feature correlation issues, this work proposed DGCN for class-specific feature extraction from DR images. The Diabetic Retinopathy Grading Convolutional Network (DGCN), which is useful for automated diagnosis and grading, improves accuracy, lowers computational overhead, and enables robust feature extraction in diabetic retinopathy pictures. Figure 4 shows the DGCN feature extraction from SMOTE-balanced features.

3.2.1 GNN

The GNN approach moved the difficulty of DR recognition to semi-supervised learning on a graph by combining the properties of unlabelled vertices with those of neighboring labeled vertices. Due to its versatility, capacity for handling a variety of data, and capacity for capturing both local and global information, Graph Neural Networks (GNNs) are an excellent choice for diagnosing diabetic retinopathy. The GNN feature recognition model is presented in Figure 5, which is a graph convolution network. The eye fundus images are used in this instance as input in a connected graph method, where each node denotes the ideal feature, and each edge denotes the disease severity of the same feature.

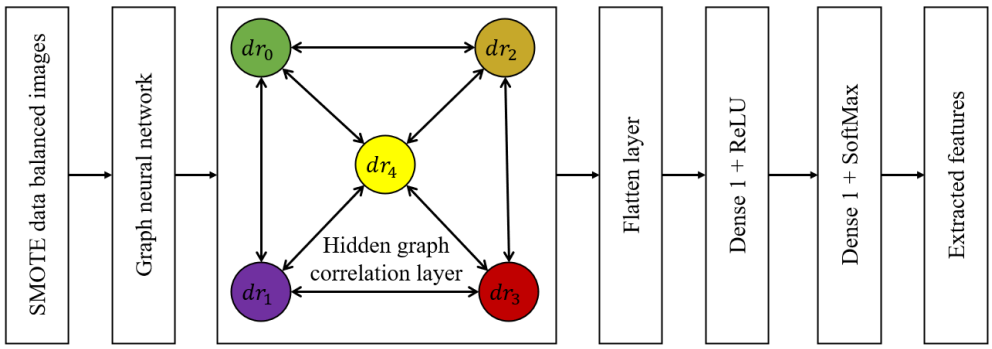


Figure 4: Proposed DGCN-based feature extraction.

The input for a connected graph is denoted by $G = (V, A)$, where V stands for nodes (vertex) and A for an adjacency matrix. Furthermore, the edge weights for all directions are stored in this adjacency matrix. The graph kernel and layer characteristics affect how well GNN performs. The individual layer operation of the GNN is shown in equation (3).

$$H^{l+1} = \sigma\left(D^{-0.5}AD^{-0.5}H^{(l)}W^{(l)}\right) \quad (3)$$

Here, l stands for the l^{th} layer, W for the linked graph's weight, H for its vertices, $\sigma(\cdot)$ represents the Leaky-ReLU activation function, and D for the diagonal matrix of $d_{i,j}$ components. When there is a lot of noise or outliers in the data: Leaky ReLU outperforms ReLU in situations when the data contains a lot of noise or outliers because it may produce a non-zero output for negative input values, preventing the loss of potentially crucial information. Additionally, the adjacency matrix element$a_{i,j}$ are used to create these $d_{i,j}$ elements:

$$d_{i,j}\text{s a} = \sum_{j=1}^{N} a_{i,j} \quad (4)$$

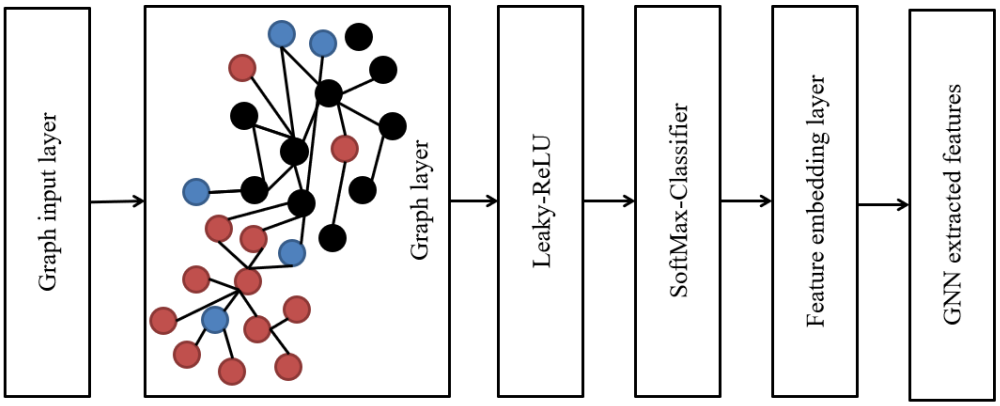


Figure 5: Architecture for GNN feature analysis.

The adjacency matrix and diagonal matrix must be constructed effectively to classify and grade DR using GNN. By recording spatial correlations and aiding feature propagation, the adjacency matrix is essential in Graph Neural Networks (GNNs) for diagnosing and grading diabetic retinopathy, increasing accuracy and interpretability. There are a total of five DR severity grades in the sample. As a result, each vertex in the linked graph layer bears a separate severity grade, and the adjacency matrix is designated as $A \in a_{5\times5}$. During training, an adjacency matrix is created based on the regularity of co-occurrences of various severity classes. In graph-based methods like Graph Neural Networks (GNNs), appearance probabilities are essential for directing connections and affecting graph sparsity. They boost GNN performance, information flow, and the graph's capacity to capture significant linkages. Additionally, each vertex is represented by a single hot encoding, allowing for simplex connections across graphs without any overlap. The adjacency matrix is produced during GNN training by computing the probabilistic co-occurrence probabilities. The appearance probability, $P_{ij} = P(L_i/L_j)$, is developed here by estimating the conditional probability ($P()$) between DR characteristics. Here, L_j stands for the likelihood of the test features, and L_i for the likelihood of the train features. By binarizing P_{ij}, the binary adjacency matrix (A') is created, improving the generalization stability of the model. Additionally, by taking into account the conditional probability-based weight bounds, A' is re-estimated, avoiding the over-smoothing issues brought on by features training. Here is how to represent the final conditional probability (A_f):

$$A'_{ij} = \begin{cases} 0, & P_{ij} < j \\ 1, & P_{ij} \geq j \end{cases} \quad (5)$$

$$A_f = \begin{cases} A'_{ij}, & i \neq j \\ 1 - P_{ij}, & i = j \end{cases} \quad (6)$$

Here, P_{ij} has a range of 0 to 1, depending on the random weight probability (τ). Additionally, at τ = 0.3, the optimal value of P is achieved as 0.25. Based on the updated A_f, the graph hidden layer operation is derived as follows:

$$H^{l+1} = \mathbb{f}\left(H^l, A_f\right) \quad (7)$$

$$\mathbb{f}\left(H^l, A_f\right) = \sigma\left(A_f, H^l, W^l\right) \quad (8)$$

In this case, the multi-grade feature separation is carried out using the feature matrix produced by the SoftMax classification operation, denoted by the symbol $\mathbb{f}(\cdot)$. A weight matrix for the lth neural network layer is called W^l.

3.2.2 Graph correlation layer

The GNN output features ($\mathbb{f}(H^l, A_f)$) were supplied to the graph correlation layer to identify the correlation among various features. Due to its effective use in medical data analysis, particularly for learning inherent correlations among various samples, DGCN is chosen to build learn graph feature representations and topological structure. To link each sample to the mini-batch of O_f is G_b, this research first creates graph correlations for the feature set $G_b = [G_1, \dots, G_k, \dots, G_B]$, where B was the batch size and $1 \leq k \leq B$. By creating an adjacent matrix A from this, DGCN links sample G_k with its K-nearest neighbors (KNN).

$$A_{kj} = \begin{Bmatrix} 0, if\ j \in K_k \\ 1, if\ j \in K_k \end{Bmatrix} \tag{9}$$

The K_k is a group of KNN of I_k, and A_{kj} stands for the correlation between the kth and jth samples. The graph correlations are developed by the mini batch $G(G_b, A)$ from the learned GNN feature G_b and neighboring matrix (A). One input layer, hidden graph correlation layers, and dense layers make up the DGCN used in this work. The hidden graph correlation layer operates in the manner described below.

$$X^{(l)} = \left[G_1^{(l)}, \dots, G_k^{(l)}, \dots, G_B^{(l)}\right].\sigma\left(D^{-1/2}AD^{-1/2}X^{(l-1)}W^{(l)}\right) \tag{10}$$

Here, the diagonal of the neighboring matrix is represented by D, and $X^{(l)}$ indicates the result of the hidden graph correlation layer with l samples.

$$D = diag(d_1, d_2, \dots, d_B) \tag{11}$$

$$d_B = \sum_{j=1}^{B} A_{kj} \tag{12}$$

The ReLU activation function is represented by $\sigma(.)$, the many diagonal components (d) that develop the D matrix. Then, under the restrictions of a graph-center loss, DGCN builds a KNN graph by calculating similarities to choose the top K samples. Similar features may be brought closer to one another, and discriminant correlations can be created by these methods for DGCN. This scenario shows the lack of correlation between head classes and tail classes during the dataset training". So, the hidden graph correlation layer effectively trains the features of head and tail classes by forming multi-level correlations. Finally, the DGCN consists of two dense layers with ReLU and SoftMax activation functions. These dense layers with activation functions formed as the fully connected layer. Here, the SoftMax classifier identifies the disease-specific features. The SoftMax classifier uses a labeled dataset to learn correlations between attributes and grades, giving test cases the grade with the best likelihood to help with multi-class grading problems like detecting diabetic retinopathy.

3.3 XGBoost classifier

As seen in Figure 6, the DGCN collected and balanced data samples are used as input to the XGBoost. In this case, to identify between the different assaults from test data, the XGBoost classifier is used. The training and testing process will employ the XGBoost ensembles of the XGBoost classifier. To deal with classification and regression problems, the XGBoost ensemble learning method, based on DT, is used. A decision tree ensemble learning method called XGBoost is applied to classification and regression problems. It is a potent machine-learning technique that predicts class labels and continuous target variables using gradient boosting and regularization. The well-known machine learning method XGBoost combines gradient boosting, parallelism, and sparsity awareness for resilience and flexibility. It effectively handles classification and regression problems. XGBoost falls under the umbrella of boosting algorithms since it strengthens a collection of poor learners over time. There are several advantages to using XGBoost. First, compared to previous boosting strategies, it learns from the data more quickly since it builds trees concurrently instead of consecutively.

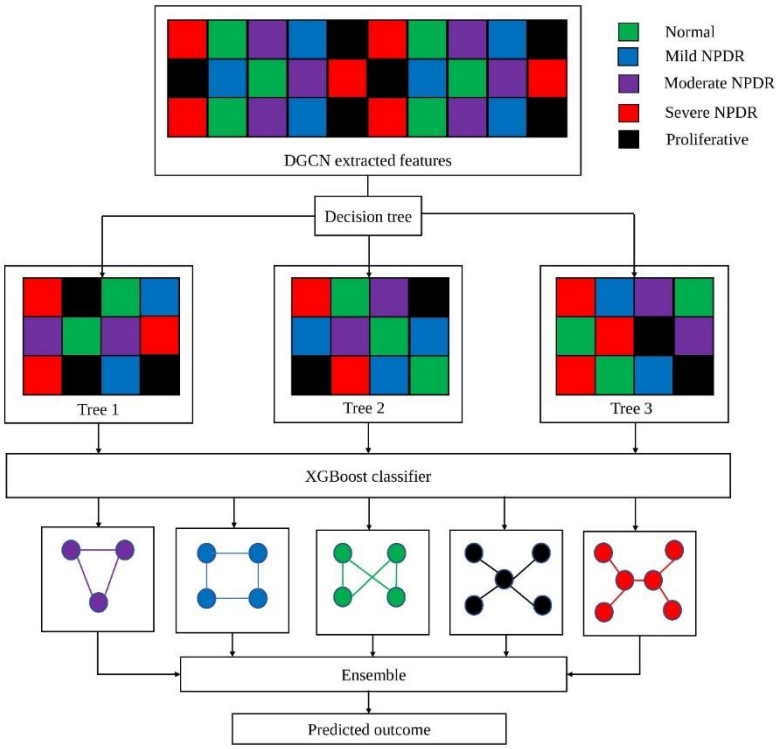


Figure 6: XGBoost classifier block diagram for obtaining prediction outcome.

Second, to reduce overfitting, XGBoost has a regularisation technique built in. Thirdly, it uses an approximation strategy to speed up the model training process. Therefore, XGBoost supports out-of-core processing and effectively manages weighted and sparse data. These elements have contributed to the XGBoost supervised learning algorithm, which is based on DT, becoming widely used. Due to its outstanding performance, efficiency, adaptability, resilience to overfitting, comprehension, community support, flexibility, and availability, XGBoost is a well-liked supervised learning strategy and a top pick among machine learning practitioners. Consider dataset, $D = \{(x_i, y_i)\}$, with $|D| = n, x_i \epsilon R^m$ and m features and n samples to illustrate how XGBoost works. The model output of a boosting technique utilizing T trees is defined as follows:

$$\hat{y}_i = \sum_{t=1}^{T} f_k(x_i),\ f_t \in \mathcal{F} \tag{13}$$

Where $F = \{f(x) = \omega_{II}(x)\}$ is a collection of trees constructed to address a classification problem. A tree is divided into two sections by each f_t: the leaf weight component (ω) and the structural part. With the use of objective functions like mean squared error or log-loss, the decision tree model's performance is assessed both before and after a node split. In XGBoost, the iterative procedure optimizes the tree structure to reduce loss and boost model precision. By reducing the following objective function, f_t may be discovered:

$$\mathcal{O} = \sum_i \ell(\hat{y}_i y_i) + \sum_t \Omega(f_t) \tag{14}$$

In this case, the distance between the prediction ($\hat{y}_i$) and object (y_i) is measured by a training loss function called ℓ, which stands for the cost of model complexity. Conventional methods fail to optimize a boosting algorithm with the aim function denoted by equation (4) in Euclidean space. In the gradient boosting approach, the prediction, and the objective function of the k-th iteration are specified as follows:

$$\hat{y}^{(k)} = \hat{y}^{(k-1)} + f_k(x) \tag{15}$$

$$\mathcal{O}^{(k)} = \sum_{i=1}^{n} \ell\left(y_i, \hat{y}_i^{(k-1)} + f_k(x_i)\right) + \Omega(f_k) \tag{16}$$

The second-order Taylor expansion is used by XGBoost to closely approach Equation (7). The goal function is expressed as follows:

$$\mathcal{O}^{(k)} \simeq \tilde{\mathcal{O}}^{(k)} = \sum_{i=1}^{n} \left[\ell\left(y_i, \hat{y}_i^{(k-1)}\right) + \}_i f_k(x_i) + \frac{||_i (f_k(x_i))^2}{2}\right] + \Omega(f_k) \tag{17}$$

Here k_i represents second-order gradient probability, which depends on the loss function, and $\mathcal{O}^{(k)}$ represents first-order gradient probabilities. Following is a definition of the XGBoost:

$$\Omega(f) = \gamma L + \frac{1}{2}\lambda \|\omega\|^2 \tag{18}$$

The number of leaves on the tree in this case is L. Take $\mathcal{L}_j = \{i: II(x_i) = j\}$ as an example of an instance asset. By adding a coefficient, equation (8) may now be readily expressed as Ω follows:

$$\tilde{\mathcal{O}}^{(k)} = \sum_{j=1}^{L}\left[\left(\sum_{i\in I_j}\}_i\right)\omega_j + \frac{1}{2}\left(\sum_{i\in I_j}||_i + \lambda\right)\omega_j^2\right] + \gamma L \quad (19)$$

The leaf j's solution weight (ω_j^*) for a tree structure may be calculated by:

$$\omega_j^* = -\frac{\sum_{i\in I_j}\}_i}{\sum_{i\in I_j}||_i+\lambda} \quad (20)$$

Equations (19) and (20) put together, we may create.

$$\tilde{\mathcal{O}}(II) = -\frac{1}{2}\sum_{j=1}^{L}\frac{\left(\sum_{i\in I_j}\}_j\right)^2}{\sum_{i\in I_j}||_j+\lambda} + \Upsilon L \quad (21)$$

Tree II(x) may be evaluated using equation (11) to find the best tree architectures. From a single leaf, the structure is gradually increased by adding branching after every repetition. A single leaf node is the starting point for decision tree algorithms like XGBoost, which divides it into two child nodes based on the cost or impurity measure. This procedure is performed several times, increasing the complexity of the tree, and limiting overfitting to produce an ensemble model. To add a specific split to the existing structure, the following method must be called:

$$\mathcal{O}_{split} = \frac{1}{2}\left[\frac{\left(\sum_{i\in I_l}\}_j\right)^2}{\sum_{i\in I_l}||_i+\lambda} + \frac{(\sum_{i\in I_r}\}_i)^2}{\sum_{i\in I_r}||_i+\lambda} + \frac{(\sum_{i\in I}\}_i)^2}{\sum_{i\in I}||_j+\lambda}\right] - \gamma \quad (22)$$

where $= I_l \cup I_r$, I_l, and I_r represent the instance sets of the left and right nodes after the split. The performance of the model as determined by the objective function is taken into account following a node split in the tree. If performance has improved, the appropriate split will be put into effect; if not, it will be stopped. Additionally, while enhancing the target function, XGBoost frequently encounters less overfitting than other boosting algorithms because of this regularisation. The target function, which specifies the desired prediction of the model, serves as the foundation for supervised learning. It directs the learning process by generating the necessary output for the inputs.

4 Results and discussion

This section gives a detailed analysis of simulation results performed using the proposed DRG-Net approach and existing DR grading classification methods which are verified using various parameters through the same dataset.

4.1 Dataset description

EyePACS dataset: There are 1427 images in the dataset. The training and validation sets were created by dividing the dataset's pictures 80:20. This division uses a random-split methodology throughout many training runs. A multi-stage training procedure was used to train several models. Multi-stage training includes simultaneously or sequentially training several models, assessing, and improving performance at each step, maybe modifying parameters, and applying ensemble approaches for improved accuracy. A training set was used for each run's training, and a validation set was used for each run's hyperparameter tuning.

Messidor dataset: The dataset contains 1744 fundus images with varied pixel sizes, which were obtained from three ophthalmologic departments. Due to characteristics including illness incidence, case diversity, clinical significance, patient consent, anonymization, standardization, and availability, the Messidor dataset is preferred for ophthalmological research. Stage 1 scans include those with MA. Stage 2 scans are those discovered in both MA and HM, while stage 3 scans are those discovered in high levels of MA and HM. Table 2 shows the dataset properties for the EyePACS and Messidor datasets. After applying the SMOTE data balancing, the dataset is normalized to equal numbers i.e., 25810 in EyePACS and 1017 in Messidor datasets.

Table 2: Number of instances in each dataset.

Dataset	**EyePACS**	**Messidor**
Class	No. of Instances	No. of Instances
Normal (No-DR)	25810	1017
Mild DR	2443	270
Moderate DR	5292	347
Severe DR	873	75
Proliferative DR	708	35

4.2 Prediction results

Figure 7 shows the predicted grading outcomes using the proposed DRG-Net. Here, the proposed method performs grading operations based on maximum feature-specific probability analysis. The suggested grading technique determines the grade with the highest likelihood based on each feature's specific contribution, allowing for

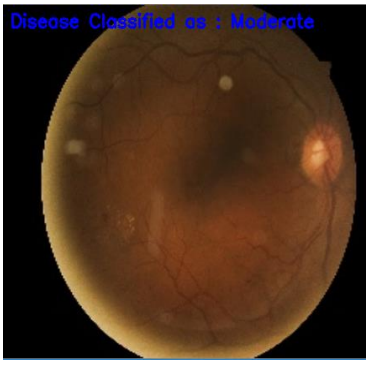
more precise and informed grading. These grades were identified by comparing the test features with the trained dataset using the SoftMax classifier. There are five distinct samples considered, and the yellow color-outlined values indicate the proposed DRG-Net probability estimate by the SoftMax classifier. Here, the DR grade is finalized based on the maximum value. For example, consider Figure 7 (a), which has probability values as **0.91,** 0.02, 0.03, 0.03,0.01. Here, **0.91** is the maximum achieved accuracy, which has the first position, so the output grade is classified as "No-DR".

0.91, 0.02, 0.03, 0.03,0.01

(a)

0.08**, 0.82,** 0.01, 0.06,0.03

(b)



0.04**,** 0.05**, 0.78**, 0.01,0.12

(c)

0.03, 0.06, 0.01, **0.74**,0.16

(e)

0.02, 0.08, 0.07, 0.03,**0.70**

(f)

Figure 7: Obtained results for DR grading prediction.

4.3 Ablation study

Ablation study is the process of performance estimation of DRG-Net in the presence and absence of various modules. The ablation research seeks to assess DRG-Net's constituent parts, such as feature importance, architecture choices, hyperparameters, loss functions, ensemble components, preprocessing techniques, data augmentation, regularisation techniques, and class balancing strategies. Table 3 and Table 4 show the ablation study of the proposed DRG-Net on both EyePACS and Messidor datasets. Here, the proposed method in the presence of all modules (SMOTE, DGCN, XGBoost) resulted in superior performance as compared to the absence of individual modules. Specifically, the absence of SMOTE data balancing and DGCN feature extraction resulted in poor performance. Combining real and fake data, changing the input layer, training, and assessing performance are the steps involved in incorporating balanced SMOTE data features into a dynamic graph convolutional system.

Table 3: Ablation study of proposed DRG-Net on EyePACS Dataset.

Presented modules	**Absence modules**	**Accuracy**	**Sensitivity**	**Specificity**
DGCN, XGBoost	SMOTE	93.01	92.35	96.12
SMOTE,	DGCN	97.54	93.48	97.45

XGBoost				
SMOTE, DGCN	XGBoost	97.89	94.78	98.12
SMOTE, DGCN, XGBoost	-	99.016	99.019	98.4375

Table 4: Ablation study of proposed DRG-Net on Messidor Dataset.

Presented modules	**Absence modules**	**Accuracy**	**Sensitivity**	**Specificity**
DGCN, XGBoost	SMOTE	90.35	93.19	95.61
SMOTE, XGBoost	DGCN	96.60	97.53	98.0
SMOTE, DGCN	XGBoost	97.34	96.35	96.45
SMOTE, DGCN, XGBoost	-	99.60	99.08	100.00

4.4 Performance evaluation

Table 5 shows the performance comparison of the proposed DRG-Net method with DCNN [12], CBAM [14], OCNN [16], and DBNN [20] using the EyePACS dataset. Here, the proposed method resulted in superior performance in terms of accuracy, sensitivity, and specificity, respectively. Table 6 shows the performance comparison of the proposed DRG-Net method with ERNN [18], Firefly-CNN [19], HA-Net [26], and MRCNN [30]. Here, the proposed method resulted in superior accuracy performance in contrast with other approaches using the Messidor dataset.

Table 5: Performance evaluation on EyePACS dataset.

Class	**Accuracy**	**Sensitivity**	**Specificity**
DCNN [12]	93.134	93.616	94.755
CBAM [14]	94.368	95.711	95.300
OCNN [16]	95.116	96.068	95.304
DBNN [20]	96.195	97.244	95.463
DRG-Net	99.016	99.019	98.4375

Table 6. Performance evaluation on Messidor dataset

Class	**Accuracy**	**Sensitivity**	**Specificity**
ERNN [18]	93.352	93.324	93.569
Firefly -CNN [19]	94.735	93.797	94.255
HA-Net [26]	96.511	95.389	94.322
MRCNN [30]	97.588	95.648	95.503
DRG-Net	99.60	99.08	100.00

5 Conclusion

This work developed DRG-Net by making use of the features of graph learning, and deep learning. Datasets, the datasets EyePACS and Messidor are subjected to the SMOTE algorithm, which brings the number of instances belonging to each DR class down to a standard level. So, the number of images in each class is maintained the same. After that, DGCN is used to extract the features that are unique to each class by locating the relationship between the classes. In addition, the prediction operation is carried out by an XGBoost classifier that has been trained with balanced feature sets. The results of the simulations performed on the EyePACS and Messidor datasets showed that the DRG-Net outperformed the more conventional methods in terms of performance. For improved performance, this work can be expanded with feature selection techniques.

Declarations

Funding

No funds or grants were received by any of the authors.

Conflict of interest

There is no conflict of interest among the authors.

Data Availability

All data generated or analysed during this study are included in the manuscript.

Code Availability

Not applicable.

Author's contributions

Venkata Kotam Raju Poranki and B. Srinivasa Rao. contributed to the design and methodology of this study, the assessment of the outcomes and the writing of the manuscript.

References

[1] Li, F., Wang, Y., Xu, T., Dong, L., Yan, L., Jiang, M., ... & Zou, H (2022). Deep learning-based automated detection for diabetic retinopathy and diabetic macular oedema in retinal fundus photographs. *Eye*, pp. 1433-1441. https://doi.org/10.1038/s41433-021-01552-8.

[2] AbdelMaksoud, E., Barakat, S., & Elmogy, M. (2022). A computer-aided diagnosis system for detecting various diabetic retinopathy grades based on a hybrid deep learning technique. *Medical & Biological Engineering & Computing*, pp. 2015-2038. https://doi.org/10.1007/s11517-022-02564-6.

[3] Vaghefi, E., Hill, S., Kersten, H. M., & Squirrell, D. (2020). Multimodal retinal image analysis via deep learning for the diagnosis of intermediate dry age-related macular degeneration: a feasibility study. *Journal of Ophthalmology*, Article ID 7493419. https://doi.org/10.1155/2020/7493419.

[4] Aziz, T., Ilesanmi, A. E., & Charoenlarpnopparut, C. (2021). Efficient and Accurate Hemorrhages Detection in Retinal Fundus Images Using Smart Window Features. *Applied Sciences*, pp. 6391. https://doi.org/10.3390/app11146391.

[5] Saranya, P., Prabakaran, S., Kumar, R., & Das, E (2022). Blood vessel segmentation in retinal fundus images for proliferative diabetic retinopathy screening using deep learning. *The Visual Computer*, pp. 977-992. https://doi.org/10.1007/s00371-021-02062-0.

[6] Zhao, L., Ren, H., Zhang, J., Cao, Y., Wang, Y., Meng, D., ... & Liu, F (2020). Diabetic retinopathy, classified using the lesion-aware deep learning system, predicts diabetic end-stage renal disease in Chinese patients. *Endocr Pract*, pp. 429–443. https://doi.org/10.4158/ep-2019-0512.

[7] Gao, Z., Li, J., Guo, J., Chen, Y., Yi, Z., & Zhong, J. (2019). Diagnosis of diabetic retinopathy using deep neural networks. *IEEE Access*, pp. 3360-3370. https://doi.org/10.1109/access.2018.2888639.

[8] Jabbar, M. K., Yan, J., Xu, H., Ur Rehman, Z., & Jabbar, A. (2022). Transfer learning-based model for diabetic retinopathy diagnosis using retinal images. *Brain Sci*, pp. 535. doi: 10.3390/brainsci12050535. PMID: 35624922; PMCID: PMC9139157. https://doi.org/10.3390/brainsci12050535.

[9] Farag, M. M., Fouad, M., & Abdel-Hamid, A. T (2022). Automatic severity classification of diabetic retinopathy based on DenseNet and convolutional block attention module. *IEEE Access*, pp. 38299-38308. doi: 10.1109/ACCESS.2022.3165193.

[10] Cahoon, S., Shaban, M., Switala, A., Mahmoud, A., & El-Baz, A. (2022). Diabetic retinopathy screening using a two-stage deep convolutional neural network trained on an extremely un-balanced dataset. *SoutheastCon2022*, pp. 250-254. doi: 10.1109/SoutheastCon48659.2022.9764079.

[11] Martinez-Murcia, F. J., Ortiz, A., Ramírez, J., Górriz, J. M., & Cruz, R. (2021). Deep residual transfer learning for automatic diagnosis and grading of diabetic retinopathy. *Neurocomputing (September 2021)*, pp. 452. 424-434. doi: 10.1016/j.neucom.2020.04.148.

[12] Majumder, S., & Kehtarnavaz, N. (2021). Multitasking deep learning model for detection of five stages of diabetic retinopathy. *IEEE Access*, pp. 123220-123230. doi: 10.1109/ACCESS.2021.3109240.

[13] Tavakoli, M., Mehdizadeh, A., Aghayan, A., Shahri, R. P., Ellis, T., & Dehmeshki, J. (2021). Automated microaneurysms detection in retinal images using radon transform and supervised learning: application to mass screening of diabetic retinopathy. *IEEE Access*, pp. 67302-67314. doi: 10.1109/ACCESS.2021.3074458.

[14] F Saeed, F., Hussain, M., & Aboalsamh, H. A. (2021). Automatic diabetic retinopathy diagnosis using adaptive fine-tuned convolutional neural network. *IEEE Access*, pp. 41344-41359. doi: 10.1109/ACCESS.2021.3065273.

[15] Raja Kumar, R., Pandian, R., Prem Jacob, T., Pravin, A., & Indumathi, P. (2021). Detection of diabetic retinopathy using deep convolutional neural networks. *Computational Vision and Bio-Inspired*

Computing, Springer, Singapore, pp. 415-430. https://doi.org/10.1007/978-981-33-6862-0_34.

[16] Gurani, V. K., Ranjan, A., & Chowdhary, C. L. (2019). Diabetic retinopathy detection using neural network. *International Journal of Innovative Technology and Exploring Engineering*, pp. 5. https://doi.org/10.35940/ijitee.j1105.0881019.

[17] Pao, S. I., Lin, H. Z., Chien, K. H., Tai, M. C., Chen, J. T., & Lin, G. M (2020). Detection of diabetic retinopathy using bichannel convolutional neural network. *Journal of Ophthalmology*, 2020. https://doi.org/10.1155/2020/9139713.

[18] Pires, R., Avila, S., Wainer, J., Valle, E., Abramoff, M. D., & Rocha, A (2019). A data-driven approach to referable diabetic retinopathy detection. *Artificial Intelligence in Medicine*, pp. 93-106. https://doi.org/10.1016/j.artmed.2019.03.009.

[19] Gadekallu, T. R., Khare, N., Bhattacharya, S., Singh, S., Maddikunta, P. K. R., Ra, I. H., & Alazab, M (2020). Early detection of diabetic retinopathy using PCA-firefly based deep learning model. *Electronics*, pp. 274. doi:10.3390/electronics9020274.

[20] Farooq, M. S., Arooj, A., Alroobaea, R., Baqasah, A. M., Jabarulla, M. Y., Singh, D., & Sardar, R (2022). Untangling computer-aided diagnostic system for screening diabetic retinopathy based on deep learning techniques. *Sensors, pp*. 1803. https://doi.org/10.3390/s22051803.

[21] Panwar, A., Semwal, G., Goel, S., & Gupta, S (2022). Stratification of the lesions in color fundus images of diabetic retinopathy patients using deep learning models and machine learning classifiers. *Edge Analytics*, Springer, Singapore, pp. 653-666. https://doi.org/10.1007/978-981-19-0019-8_49.

[22] Saini, M., & Susan, S (2022). Diabetic retinopathy screening using deep learning for multi-class imbalanced datasets. *Computers in Biology and Medicine*, pp. 105989. https://doi.org/10.1016/j.compbiomed.2022.105989.

[23] Kalpana Devi, M., & Mary Shanthi Rani, M (2022). Classification of diabetic retinopathy using ensemble of machine learning classifiers with IDRiD dataset. In *Evolutionary Computing and Mobile Sustainable Networks*, Springer, Singapore, pp. 291-303. https://doi.org/10.1007/978-981-16-9605-3_20.

[24] Shaukat, N., Amin, J., Sharif, M., Azam, F., Kadry, S., & Krishnamoorthy, S (2022). Three-Dimensional semantic segmentation of diabetic retinopathy lesions and grading using transfer learning. *Journal of Personalized Medicine*, pp. 1454. https://doi.org/10.3390/jpm12091454.

[25] Jang, S. I., Girard, M. J., & Thiery, A. H (2022). Explainable and interpretable diabetic retinopathy classification based on neural-symbolic learning. *arXiv preprint arXiv:2204.00624*, 2022.

[26] Shaik, N. S., & Cherukuri, T. K. (2022). Hinge attention network: A joint model for diabetic retinopathy severity grading. *Applied Intelligence*, pp. 15105-15121. https://doi.org/10.1007/s10489-021-03043-5.

[27] Da Rocha, D. A., Ferreira, F. M. F., & Peixoto, Z. M. A (2022). Diabetic retinopathy classification using VGG16 neural network. *Research on Biomedical Engineering*, pp. 761-772. https://doi.org/10.1007/s42600-022-00200-8.

[28] Yu, M., & Wang, Y. (2022). Intelligent detection and applied research on diabetic retinopathy based on the residual attention network. *International Journal of Imaging Systems and Technology*, pp. 1789-1800. https://doi.org/10.1002/ima.22734.

[29] Ganesh, M., Dulam, S., & Venkatasubbu, P. (2022). Diabetic Retinopathy Diagnosis with InceptionResNetV2, Xception, and EfficientNetB3. *In Artificial Intelligence and Technologies*, Springer, Singapore, pp. 405-413. https://doi.org/10.1007/978-981-16-6448-9_41.

[30] Erciyas, A., Barışçı, N., Ünver, H. M., & Polat, H. (2023). Improving detection and classification of diabetic retinopathy using CUDA and Mask RCNN. *Signal, Image, and Video Processing*, pp. 1265–1273. https://doi.org/10.1007/978-981-16-6448-9_41.

[31] Aiswarya, R. S., Kumar, V., & Punitha, P (2023). The Effect of Grain Size and Silicon Content on Non-Oriented Grain Steel Anomalous Loss Through Frequency Excitation in The Medical Healthcare by Using Big Data Analysis. *Tamjeed Journal of Healthcare Engineering and Science Technology*, pp. 43-53. https://doi.org/10.59785/tjhest.v1i1.5.

ELM-Based Imbalanced Data Classification-A Review

*[1]Brajendra Singh Rajput, [1]Partha Roy, [1]Sunita Soni, [2]Bhagat Singh Raghuwanshi
[1]Computer Science and Engineering, Bhilai Institute of Technology, Durg, Chhattisgarh, India, Chhattisgarh Swami Vivekanand Technical University (CSVTU)
[2]Information Technology, Madhav Institute of Technology and Science (MITS) Gwalior, Madhya Pradesh, India
E-mail: brajendrarajput89@gmail.com, partha.roy@bitdurg.ac.in, sunita.soni@bitdurg.ac.in, bhagat@mitsgwalior.in
*Corresponding Author

Keywords: SMOTE, Imbalanced Learning, Classification, under bagging, ensemble, Voting approach, variable length, Kernelized ELM, Fraud prediction.

Recieved*:* August 8, 2023

Imbalance issues occur in Machine Learning (ML) when there is high distortion in the class distributions. A great challenging task in ML is the imbalance of data classification. It is because most classification methodologies tend to bias toward the majority class even though high importance is given to the minority class. To enable its stable operation, many techniques are utilized recently that are still in use for classifying imbalanced datasets efficiently. Owing to the assumption with balanced class distribution or equal misclassification, the prevailing learning algorithms are prone to favor the majority class when handling complicated classification issues with skewed class distribution. The most prominently adopted technique to deal with data having imbalance class distribution is Extreme Learning Machine (ELM). Unwanted class boundaries as of data with unbalanced classes may be learned by ELM similar to various other classification algorithms. Grounded on the kernel utilized, elevated weighted ELM, active learning-centered techniques, etc, an augmentation in the ELM framework is done for efficient imbalanced classification. Regarding ELM approaches, the latest studies for imbalance classification are studied here. Finally, regarding G-Mean, Accuracy, and Imbalance Ratio (IR), the research studies' performance was analogized.

Povzetek: Prispevek obravnava neuravnotežene učne primere. Podaja pregled zlasti novejših metod za klasifikacijo neuravnoteženih podatkov z ELM in poudarja izzive pristranskosti modelov do večinskih razredov ter opisuje tehnike, kot so SMOTE in jedrski ELM.

1 Introduction

Data stream classification has garnered wide interest in the modern era owing to the massive expansion in data availability on the Internet and in other fields. Unstructured data streams that consistently arrive on time are difficult to classify because they lack class labels along with accumulating over time [1]. Many data stream algorithms do not perform satisfactorily or fail altogether in mixed data streams comprising categorical along with numerical values or in limited labeled samples. Multi-classes have more categories than two, while binary classes contain only two categories in the dataset [2].

The dataset is affected by imbalance problems like text classification, web fault prediction, Credit Card Fraud Detection (CCFD), high error rate classification models, etc. [3]. The time-dependent alteration in data streams is one major issue in data stream classification schemes [4]. Another typical classification issue is Class Imbalance (CI) which develops when one of the classes, known as the minority class, has a smaller amount than another class, known as the majority class. To deal with unstructured class streams, many strategies were engendered. Unsupervised learning and supervised learning are the two diverse learning strategies that are employed to address CI issues in ML paradigms [5]. Resampling, cost-sensitive learning, one-class learning, feature selection, and other techniques are also used. However, multiclass classification was not accurately dealt with by conventional methodologies [6]. Hence, the imbalance issue engendered in the data classification process was solved by ELM. Figure 1 shows a diagrammatic representation of imbalance classification.

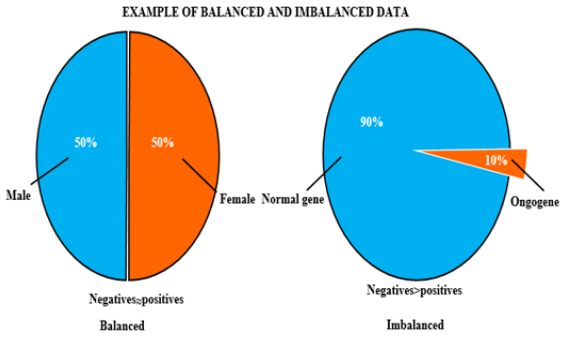


Figure 1: Diagrammatic representation of imbalance classification.

A Single Hidden Layer Feed-Forward Neural Network (SHLFN) is the foundation of an ELM. Utilizing kernels, random neurons with or without specified shapes, and optimization constraints are all combined in this method [7]. To determine the SHLFN's output weights, it randomly chooses the hidden units [8]. Owing to the lack of iterative tuning with training utilizing the generalization operation, ELM is quicker than SHLFN. Additionally, the fundamental benefit of employing ELM is that it models a standard approach for both binary class and multi-class situations [9] in addition to offering universal approximation capability and classification ability. Medical data classification, medical diagnostics, image quality evaluation, and many more are the several application disciplines of ELM. Moore-Penrose generalized inverse is utilized by ELM to update its weight apart from standard single hidden neural networks. The classification accuracy is maximized by several prevailing extreme learning algorithms by classifying the majority class accurately while incorrectly classifying the minority class [10]. For classifying imbalanced datasets, a variety of sampling-based strategies, including under-sampling and oversampling, cost-sensitive learning techniques, along with ensemble learning, have recently been employed

2 Related works

Grounded on the ELM approach, diverse prevailing research methodologies related to the imbalanced classification problem were explained as follows,

2.1 ELM algorithm-based approaches for imbalanced class problem

Bhagat *et al.* [11] presented a novel Synthetic Minority Oversampling Technique centered Class-Specific ELM (SMOTE-CSELM) for elevating imbalanced classification complexities with the skewed class distribution. The minority class samples were elevated by SMOTE that created synthetic samples pertained to the minority class to determine the classifier's decision region. Regarding computation complexity, the experiential outcomes illustrated the presented approach's higher efficacy. However, the large size imbalanced classification complexity could not be tackled by the presented model.

Bhagat Singh *et al.* [12] developed Under Bagging ensemble-centered variants of a Kernelized ELM (UBKELM) where the random under-sampling's strength along with bagging was integrated. By a majority of class samples' random under-sampling, balanced training subsets of large numbers were integrated. As UBKELM had more stability along with promising generalization performance, it was utilized as a component classifier to execute ensemble operations. However, the approach's accuracy was impacted by the dependence of several training subsets on the CIs degrees.

Sanyam Shukla *et al.* [13] explicated a CS-ELM's variant named Online Sequential Class-Specific ELM (OSC-SELM). Here, it utilized a class-specific regularization approach that makes assigning weights to the training samples unnecessary. The class-specific regularization coefficients were computed by utilizing the class proportion along with the regularization coefficient's values. Thus, better generalization was attained by the presented scheme than various other prevailing schemes. However, the presented method's stability was affected by the overfitting complexities.

Hualong *et al.* [14] projected a proficient Active Online-Weighted ELM (AOW-ELM) classification scheme to tackle larger time consumption issues. To tackle the CI problem, a cost-sensitive learning scheme was adopted to choose the Weighted ELM (WELM). An active learning model was built by utilizing an AL-ELM algorithm, subsequently. For an effective weight update rule, an efficient online learning mode of WELM was designed. Finally, a flexible along with effective early stopping criterion was deduced centered on the margin exhaustion criterion. The missed clusters were avoided by the utilized clustering techniques. By this, the suggested method was recognized to be more efficient than numerous prevailing models. However, the output was affected by the random initialization of weights variation betwixt the input and HL.

Wendong et al. [15] suggested a scheme for classification issues with imbalanced data distributions by utilizing the Class-specific Cost Regulation ELM (CCR-ELM) algorithm. Kernel-centered CCR-ELM was formed by introducing kernel functions in CCR-ELM. The

number of class sample's effects was pondered in CCR-ELM along with the effects of data's dispersion degree. An elevation in classification performance with the superior status diagnosis was explicated in the experiential outcomes. However, direct utilization in multiclass imbalanced classification issues was not handled by the presented model.

Bhagat *et al.* [16] provided a variant of ELM named Class-Specific ELM (CS-ELM) for handling binary CI issues. For better-imbalanced classification, class-specific regularization parameters were employed here, that are computed utilizing class distribution. As there was no weight assigning the process to the training instances, the presented model's performance was elevated. However, the ELM's stability cannot be addressed by the presented scheme under several operational uncertainties.

Yong *et al.* [17] developed a Differential Evolution (DE)-centered ensemble learning stratagem to resolve the imbalanced data classification problem. Numerous WELMs were chosen initially with distinct activation functions as base learners to embody the university. Additionally, each base learner's weight was optimized by utilizing the DE algorithm. A candidate weight vector's population was presented subsequently. Then, choosing the individual having best fitness value as the base learner's weight in the ensemble occurs. Regarding Geometric mean (G-mean), a superior classification performance than other analogized models was provided by the suggested approach. Nevertheless, the technique's performance was affected by the outlier's attendance.

Hui *et al.* [18] explained an Evolutionary ELM with a sparse cost matrix for imbalanced learning. Here, engendering the case-weighting ELM on a sparse cost matrix occurred in a diagonal form. The misclassification issue was optimized by the multi-objective optimization regarding penalty factors utilizing an evolutionary algorithm merged with an error-bound model. As the presented model was aided by the link betwixt the generalization ability and case-weighting factors, it was utilized as adaptive cost-sensitive learning. However, identifying the penalty factor's perfect set along with thumb rules was critical as it was too specific to fit generic situations.

Chengbo *et al.* [19] introduced an Improved Weighted-ELM (IW-ELM) algorithm for imbalanced classification. The appropriate weights were assigned for the imbalanced data classification by presenting a voting-centered weighting scheme. The weighted ELM classifier's training was involved in the presented approach. Following that, proper classifiers for voting were determined by eliminating unusable classifiers. And finally, grounded on majority voting, the classification outcomes were determined. When analogized with other ELM-centered algorithms, high accuracy was exhibited by the simulation outcomes. But, the training time was elevated by the difficulty in the misclassification cost matrix's generation

Tianlei *et al.* [20] employed a Deep-Weighted ELM (DWELM) algorithm for imbalanced data classification. The representation capacity was augmented by the elevated stacked multilayer deep representation network that was trained with ELM (EH-DrELM). The sample weights for imbalanced multiclass data were optimized by a rapid AdaBoost algorithm. The data imbalance issue in sequential learning was alleviated by meta-cognitive online sequential ELM. However, pushing the data as of distinct classes in diverse directions was not sufficiently made by the linear ELM.

Honghao *et al.* [21] established Weighted ELM (WELM) to optimize imbalanced classification. Grounded on the dandelion's behavior, a new swarm intelligence algorithm named Dandelion Algorithm (DA) was employed for imbalance classification. A diverse number of seeds are engendered by every dandelion in DA. The next iteration's dandelion was opted as of the seeds formed by diverse dandelions. The dandelion's position was altered via sowing along with selection operations. High detection performance was exhibited by the experiential outcomes than other conventional schemes. It was also utilized in CCFD. However, the system's efficacy was affected by the degradation in convergence rate.

2.2 Ensemble-based ELM imbalanced classification problem.

In table1, CI problem classification based on ensemble-based ELM is explained.

Table 1: Analysis of Ensemble-based ELM imbalanced classification problem

Researcher Name	Methodology	Description	Result	Drawback
Arnis *et al.*[22]	Entropy-centered Classifier (EC) algorithm.	Grounded on the original class proportions in the training dataset, weights were encompassed in the entropy computation initially. For each class, the weights or class importance was computed that remains unaltered during the learning process.	Regarding classification accuracy along with sensitivity, a more promising solution was attained in a complex environment.	If the initial dataset was unbalanced, a high false negative rate was shown.
Xiaoyan *et al.*[23]	Ensemble of ML- K-Nearest Neighbour (EML-KNN)	Metadata extraction was the initial step that encompasses meta-target identification along with the meta-feature collection. In the model construction stage, a meta-model was then constructed for each candidate's algorithms. Finally, a new classification problem was recommended in the recommendation step.	Augmented recommendation performance was resulted by recommending diverse algorithms for various classification issues automatically.	The improper distance method affected the quality.
Wentao *et al.*[24]	Sparse Weighting centered on Online Sequential ELM (SW-OSELM)	By oversampling utilizing the SMOTE, a balanced training set was attained initially. Grounded on training errors, a balanced process was executed after that. Centered on the final training set, the initial mode was set.	Better classification accuracy with diminished accuracy loss was exhibited by the experiential outcomes.	Unnecessary training time results in the presence of redundant virtual samples.
Junhai *et al.*[25]	Map Reduce and	Utilizing the Map Reduce	Classifying the imbalance	The system's

	Ensemble-centered ELM technique.	algorithm, the positive class instance's center was computed initially. The sampling was executed followed by the classifier component's training utilizing the ELM algorithm. By utilizing the voting approach, the integration was executed finally.	d large datasets elevated the positive class instance's learning region.	performance was affected by the big data partitioning into small pieces automatically along with their deployment in a parallel computing node.			weights along with biases are generated randomly. Further, the original data's main features are obtained by ELM as an autoencoder (ELM-AE).		
Shifei *et al.*[26]	UnSupervised-ELM(US-ELM)	By framing a new cost function, the embedded matrix was obtained initially. The clustering in the embedded matrix was executed by the k-means algorithm. Thus, input	faster responding along with greater generalization	It was not suitable for high-dimensional datasets as it takes a longer training time.					

Fulong *et al.* [27] presented an ensemble-centered adaptive over-sampling scheme to overcome the CI problem. Elevated microaneurysm detection was attained by utilizing Boosting and Bagging along with Random subspace. Thus, the positivity of adaptive over-sampling and ensemble was integrated by the presented ensemble-centered over-sampling schemes. The induction biases introduced as of the imbalanced data were diminished by the amalgamation of ensemble and adaptive over-sampling. Generalization along with ELM classification performance was elevated as a result of this. However, the classification of imbalanced data was made more critical owing to the presence of false alarms.

Zhenyu *et al.* [28] developed the Easy-SMT ensemble approach to tackle the imbalance of learning's impact. With a SMOTE-centered oversampling policy to supplement minority defective classes along with Easy Ensemble to alter a CI learning problem into an ensemble-grounded balanced learning subproblem, Easy-SMT was nothing more than an integrated ensemble-based approach. Regarding good classification capability, the presented method outperformed other conventional methods. Because of the minority class sample's small-sized features, a large number of minority class samples were not recognized clearly.

Guillem *et al.* [29] projected a Probability Threshold (PT) - Bagging approach to solving CI problems raised in the network. Cost-sensitive learning, rebalancing mechanisms, along with threshold moving were the '3' stages in the presented PT-Bagging approach. Various misclassification costs were assigned to different classes

initially. To balance the training data, the data was resampled after that. By utilizing the corresponding threshold, a model as of the dataset was altered into a class label. As the presented work discarded various training data, more efficient computation than the traditional schemes was attained by it. However, the error rate was elevated by the miscalibrated posterior probability and the prior shift. Shown in Figure 2.

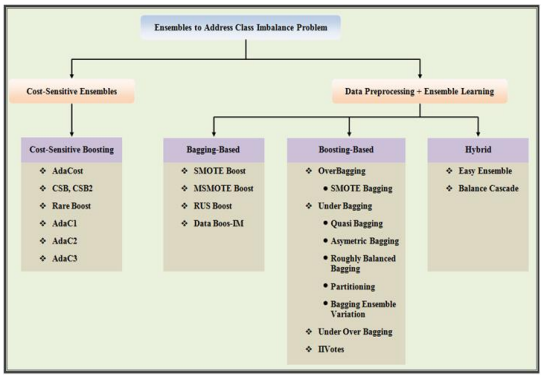


Figure 2: Different ELM-based approaches for imbalance class problems.

Hossam *et al.* [30] ejected a hybrid approach based on ensemble approaches. Feature selection, data balancing, along with classification were the '3' stages in this model. In the prediction process, the influencing factors were identified initially. Then, a least powerful oversampling technique named SMOTE was utilized to handle the imbalanced data distribution. Regarding accuracy, Area under Curve (AUC), along with Geometric-Mean (G-Mean), better prediction outcomes were obtained. But, the outlier's presence made the outputs unstable.

Yanjiao *et al.* [31] suggested a Parallel one-class ELM (P-ELM) centered Bayesian methodology for solving an imbalanced classification issue. Grounded on the class attribution of samples, the dataset was partitioned into k components here, which was further classified by subjecting it to the corresponding k Kernel-grounded one-class ELM classifiers. The suggested model's superiority was validated by the experiential analysis which was utilized in both binaries along with multiclass classification. However, the samples lost their intrinsic property owing to the sample area's overlapping.

2.3 Semi-Supervised ELM-based imbalanced classification problem.

Table 2 details the semi-supervised ELM-based techniques, their purpose, results, and limitations.

Table 2: Analysis of imbalance classification based on semi-supervised ELM

Authors	Method used	Purpose	Results	Limitation
Zhiqiong *et al.*[32]	Distributed and weighted ELM (DW-ELM) algorithm.	Primarily, each class's data center was engendered randomly. Grounded on the multivariate Gaussian distribution, the data of each class was engendered, where the formerly produced center point was pondered as the mean whereas the variance was the number of classes' reciprocal. Finally, the experimental data was made by combining all classes' generated data.	The approach's efficacy was validated by relatively stable training time.	The HLs dimensionality was augmented by elevating the number of hidden nodes resulting in the algorithm's increased computation time.

Junjie *et al.*[33]	Self-Training Regularized Weighted Online Sequential ELM (ST-RWOS-ELM)	Initially, the positive and negative sample ratio along with the data chunk was unaltered. Also, by applying the weighting procedure, the errors in the samples were totally rebalanced.	Short calibration time along with low computational complexity and better accuracy of about 90%.	Imbalance classification was executed only for a highly labeled dataset.			The ELM's performance in semi-supervised tasks was elevated by a weighted assembled regularization approach via the sigmoid activation function.	tion tasks were elevated by the combination of Laplacian and Hessian terms.	
Nan *et al.*[34]	UnSupervised ELM with Kernel (US-KELM)	The regularised cost function was diminished by utilizing a wavelet kernel function rather than output HL.	To attain better performance, it was capable of utilizing unlabeled data.	The training time was elevated by utilizing '2' loops in the kernel matrix computation.	**Jin *et al.*[36]**	Distributed Semi Supervised-ELM (DSS-ELM) with Zero Gradient Sum (ZGS) optimization strategy.	Each node in the communication network has an identical basic function and random parameters. The globally optimal coefficient vector was computed by ZGS strategy utilizing an iterative process	Guaranteed privacy-conserving algorithm with enhanced convergence	Disconnection might occur in communication nodes that cause data loss
Yongxiang *et al.*[35]	Laplacian-Hessian regularization SS-ELM (LHRSS-ELM)	Primarily, Laplacian and Hessian functions were incorporated with semi-supervised learning methods.	The extrapolating power, accuracy, along with robustness in multiclass classifica	When the function was heavily oscillating, it was inaccurate.	**Hongyu *et al.*[37]**	Cost-Sensitive-Dissimilar-ELM(CS-D-	Grounded on the output's uncertainties acquired as of the	This scheme was insensitive to the choice of both the	The algorithm was made invalid and unavailable by

	ELM)	basic classifier, testing samples were partitioned into numerous groups initially. After that, sample groups with small uncertainties were pondered. The basic classifier that predicted the samples with their labels was summed to the original training set. Via the chosen training algorithm, the classifier was restrained on the enlarged training set finally.	classifier training algorithms along with the specific representation forms of uncertainty.	adding up low fuzziness to the training samples.

Carlos *et al.* [38] developed a Semi-Supervised Online Elastic ELM (SSOE-ELM) scheme to overcome the CI issues. During the training phase, each dataset was bifurcated into labeled and unlabelled instances. Grounded on the chunk utilized, training was executed in an online manner. Utilizing classification accuracy's mean and standard deviation as performance metrics, the experiment has recurred 30 times for each division. Especially in a situation like a relatively low amount of labeled data, SSOE-ELM was pondered as a feasible semi-supervised online learning algorithm. Nevertheless, the algorithm's accuracy was altered by the minute variation in the regularization parameter, as it was very sensitive.

Jie Yanga *et al.* [39] projected a Regularized Correntropy criterion-centered SSELM (RC-SSELM) technique. The correntropy was employed in the cost function's formulation. Non-Gaussian noises and outliers were dealt with by adopting the Maximum Correntropy Criterion (MCC) in the RC-SSELM's optimization strategy. A second-order statistical resemblance measure in the kernel space was the Correntropy. Thus, it provided better performance for a high percentage of outliers along with founding robust. However, poor performance was exhibited while readily propagating the misclassified outcomes in the labeled training data to the adjacent samples.

Parsa *et al.* [40] established online ensembles from multi-class imbalanced data for learning. Grounded on individual recall rates, the minority instances were over-sampled initially by the Improved SMOTE Online Ensembles (ISOE) while dynamically re-sampling whole classes. To re-balance the sample sizes of various classes, Improved Online Ensembles (IOE) were utilized by altering the parameter rate grounded on performance along with each class's number of instances. The unknown labels were processed by combining self-training into the online learning process. Thus, a better outcome was provided by the presented approach than the conventional schemes regarding balancing the predictive accuracies. Utilizing only static data sets was a major drawback here; the system's durability gets affected by the streams that do not contain concept drifts.

Zheng *et al.* [41] introduced a Semi-Supervised Broad Learning System (S2-BLS) for solving CI issues. Here, the mapped features were obtained by utilizing the ELM-grounded Auto Encoder (ELM-AE). Consequently, the discriminative projecting weights betwixt ground truths and the changed features encompassing the mapping features along with the enhancement nodes were computed. Subsequently, training the discriminant linear mode with altered samples along with ground truths occurs. Betwixt the labeled along with unlabelled samples, the information was explored efficiently. But, the training time was elevated by the useless information comprised in the retrieved mapped features along with enhancement nodes.

Adnan *et al.* [42] presented a Multi Kernel Semi-Supervised ELM (MKSSELM) method to overcome the

unbalanced data conflicts. Utilizing the multi-kernel approach, the kernel parameters were fine-tuned initially. The kernel regulation was controlled by commanding a norm constraint on the kernel combination weights and optimizing the ELM structural parameters along with the kernel combination weights. The system's generalization performance was augmented by the Kernelized ELM. However, the local optima in the dataset affected the presented approach's performance.

Chuangquan Chena *et al.* [43] enumerated an Extreme Semi-Supervised Learning (ESSL) scheme to overcome imbalance classification. Initially, both binary, as well as multiclass classification issues, were handled by ELM in a unified model. Then, the HL was encoded by a tiny Approximate Empirical Kernel Map (AEKM) that diminished the computational cost along with memory usage. Subsequently, via the weighting function's elaborative design, the balance constraint or prior knowledge in the unlabeled data was eliminated. The imbalanced class problem was effectively along with efficiently solved by the ESSL without any fine-tuning parameters. However, the scalability was affected by the ESSL's dependence on regularization parameters.

2.4 Kernel- Fuzzy-based ELM imbalanced class problem.

Table 3 illustrated the diverse kinds of existing kernel-Fuzzy-based ELM approaches for imbalanced class problems.

Table 3: Evaluation of imbalance class problem based on kernel and fuzzy based ELM.

Researcher	Techniques	Procedure	Performance measure	Disadvantage
Weipeng *et al.*[44]	Fuzziness-centered Online Sequential ELM (FOS-ELM).	Utilizing the fuzzy classifier, meaningful data was chosen as of the given data initially. Here, the sequential learning utilized only the selected samples with high-output fuzziness.	Higher generalization performance with high testing accuracy.	The loss of vital information was caused by the utilization of predefined conditions.
Lu Li *et al*. [45]	AdaBoost Weighted Composite Kernel ELM (AdaBoost-WCKELM)	Grounded on the AdaBoost algorithm, the training sample's weights were adjusted here. The spatial information with original spectral bands was embedded by extracting the local mean and standard subsequently. Finally, the imbalance issue was solved by applying the composite kernel method in the ELM	Better accuracies were obtained even though the training samples were imbalanced or very limited.	Insufficient discriminative information for multiclass was provided by utilizing unlabelled data

		algorithm.				LM)	class, a subset of training data was utilized as the centroid of those kernel functions chosen by the ensemble method. The balance data was chosen by adapting the random under-sampling method.	when analogized with traditional methods.	computational complexity with elevated dataset size was resulted by utilizing training samples for the kernel matrix.
Ping *et al*.[46]	(MKELM)-centered random forest binary classifier (MKELM-RF)	Multiple kernel functions comprised of different sub-feature spaces were utilized for space mapping. To tackle the CI issue, synthetic data points for the minority samples were generated by utilizing SMOTE. Finally, the features were chosen by executing bootstrap sampling to train MKELM-centered learners.	accuracy 98.1%, average positive predictive value 93.9%, and sensitivity 94.4%	Poor minority class sample detection.	**Bhagat Singh Raghuwanshia *et al.* [48]**	Generalized Class Specific Kernelized ELM(GCS-KELM)	The class proportion along with the regularization parameter's value was determined by the Class-specific regularization parameters. The input data was converted into feature space by utilizing the Gaussian	As it does not need the weight allotment to each training sample, the computational cost was diminished along with an accuracy of 92%.	The overfitting problem was raised in the approach.
Bhagat *et al*.[47]	Reduced Kernelized WELM (RKWE	To maintain equal instances as of each	A significantly lower running time	A significant augmentation in the					

		Kernel function.					base classifiers outcomes were combined.		
Bhagat *et al.*[49]	Reduced Kernelized-WELM(RK-WELM)	The CI issue was handled by utilizing an ensemble-centered solution initially. The base classifier used here was the RKWELM which generated balanced subsets on a training dataset. Finally, utilizing majority voting along with soft voting, those	Owing to the diminished number of kernels utilization, rapid classification along with 80% accuracy was obtained	Elevated data complexity resulted from choosing centroid kernels.					

Shuya *et al.* [50] presented a Weighted Online Sequential ELM with Kernels (WOS-ELMK) for CI learning. To tackle the CI issue, the initialization phase, decremental learning phase, along with sequential learning phase was utilized in both binary class as well as multiclass datasets. Further, the non-optimal hidden node issue linked with random feature mapping was avoided by utilizing kernel mapping in WOS-ELMK. Thus, by getting not being adapted to the new data, the instability was avoided in the presented approach thereby elevating the presented approach's performance. However, the approach was made unsuitable for large datasets with the fixed memory scheme utilization.

Bhagat *et al.* [51] presented Balance Cascade-centered KELM (BCKELM) to tackle the CI problem. The training subsets were balanced by adapting a random under-sampling model. Further, the base learner was engendered by the ensemble method sequentially. A new balance training subset was engendered in every iteration. Therefore, better stability along with good generalization performance was provided by the presented approach. However, the computation time was elevated by the balanced training subsets dependence on the IR.

Hongyun *et al.* [52] developed the K-means SMOTE method along with the ANN-centered KELM algorithm (K-mean SMOTE+KELM) for solving CI problems. Utilizing the KELM approach, the data was balanced by integrating the K-means SMOTE technique with the ANN. Data weighting operation was executed by utilizing Deep Weighted ELM (DWELM) through an enhanced multiclass AdaBoost imbalanced learning algorithm (AdaBoost-ID) and an enhanced deep representation network centered ELM (EH-DrELM). The imbalance issue was sorted along with features' learning capacity was elevated by the K-mean SMOTE+KELM algorithm. However, the accuracy was affected by the value of the allocated weights.

Yang et al. [53] introduced Enhanced Kernel-centered Multilayer Fuzzy Weighted ELM (EML-KFWELM) to solve the CI problem. The minority class's classification performance was elevated by embedding a weighted strategy into ML-KELM by ML-KFWELM. Then, the classification errors caused by outliers and noise samples

were eliminated by the fuzzy membership. Moreover, the ML-KFWELM's parameters were optimized by the EGWO strategy. The classification performance, accuracy along with stability was elevated by the optimization approach than other models. However, a slow response was made in the system owing to the existence of noise.

Zhennao *et al.* [54] addressed a hybrid ML model grounded on enhanced Grey Wolf Optimization (GWO) along with KELM. The GWO algorithm's global along with local search capabilities was elevated by a new hierarchical mechanism initially. The key parameters for KELM were tuned by utilizing IGWO. Therefore, a better convergence rate along with performance elevation was indicated by the IGWO-KELM's experiential outcomes. However, the computational burden was elevated by utilizing a discrete optimization strategy.

Pattaramon *et al.* [55] projected Radial Basis Function Network (RBFN) to tackle the class overlapping problem in imbalanced data. In instances where a common space was shared by more than one class, class overlaps occur generally. Additionally, shifting the decision boundary towards the negative class led to positive instances of misclassification near the class boundary. It cannot be utilized in a real-world application even though elevated classification accuracy was presented by this approach.

El Barakaz Fatima *et al.* [56] established three algorithms namely RONS (Reduce Overlapping with No-sampling), ROS (Reduce Overlapping with SMOTE), and ROA (Reduce Overlapping with ADASYN) to solve the CI problem. The classification of class-imbalanced data was optimized by diminishing the overlap degree along with looping the learning to separate datasets. However, the high dimensionality affected the presented system's performance.

Everlandio *et al.* [57] exaggerated evolutionary inversion of class distribution in overlapping areas for multi-class imbalanced learning (EVINCI) to tackle the multi-class imbalance problem. A set of samples evolved as of an imbalanced dataset by utilizing a Multiobjective Evolutionary Algorithm (MOEA). In the overlapping regions, while choosing samples that generated precise models, the concentration of majority classes' fewer representative instances was opted. Regarding predictive accuracy, the approach's superiority was evaluated by the experiential outcomes. But the performance was affected by the larger IR.

2.5 Various other ELM-based approaches in imbalance classification.

Mingjing *et al.* [58] introduced a KELM parameter centered on a swarm intelligence algorithm called GWO. By exploring this for optimal parameter prediction, the KELM classifier's generalization capability was maximized by the swarm intelligence algorithms. Owing to the presented approach's superiority, it was utilized in the bankruptcy prediction. However, for datasets containing a large number of samples, it was not suitable.

Dong *et al.* [59] engendered an efficient bankruptcy prediction model centered on the KELM method. The optimal parameters were searched by a two-step grid search strategy that unites coarse search with fine search. The prediction was executed by the obtained optimal parameter finally. When analogized with the prevailing models, a better performance was exhibited by the presented model. However, a large variation in the classification accuracy was engendered by randomly assigned weights in distinct trials.

Yan Wei *et al.* [60] projected effective hybrid Gaussian Barebones (GB) improved Harris Hawk's Optimizer (HHO) centered KELM (GBHHO-KELM) method for predicting imbalance classification issues in students' intentions on self-employment. The global along with local search capabilities was balanced by introducing the GB mechanism into the HHO algorithm. Thus, better parameter combinations along with higher prediction sensitivity with more stable performance were attained by the presented scheme. However, the accuracy was affected by accomplishing the smaller fitness together with variance by the presented model.

Yongshan *et al.* [61] established a new Instance Cloned ELM (IC-ELM) in order to compute the class label of a testing instance. The testing instance's k nearest instances were selected by introducing the instance cloning technique. In the extended data set, each training instance's weight was computed. The underlying class label for the testing machine was predicted by formulating the learning model with an extended training dataset. Thus, the overfitting issue was more effectively handled by the presented model than the prevailing models. However, it cannot be utilized in practical applications like disease diagnosis, etc.

Xiaowei *et al.* [62] developed a Genetic Algorithm-centered ELM approach for the CI problem. The GA's population diversity was elevated by the Extinction and Immigration (EI) strategy. Here, the feature selection process utilized the Error Minimized-ELM (EM-ELM).

As per the ELM's generalization theory, an effective ranking method was formed for the ELM's selection process. Finally, the selected ELMs computed the final ensemble. Grounded on higher prediction accuracy along with better stability, the efficacy was demonstrated by the experiential outcomes. However, the system was made more time-consuming by the filter and embedded ensemble methods than the other schemes.

Ali Asghar *et al.* [63] introduced the Opposition Based Learning Grey Wolf Optimization (OBLGWO) algorithm for solving the CI problem. This algorithm included a greedy selection operator and LF mechanisms, random leaders, along with the strategy of random spiral motions. Regarding the faster convergence rates, the outcomes revealed the presented approach's better solution quality. The model selection issue was tackled effectively. However, the scalability was affected by the elevation in the average error and standard deviation that was augmented by the population dimension.

Xu Xiaolong *et al.* [64] presented a Density-based DSMOTE. Initially, the minority class's samples were partitioned into core samples, borderline samples, and noise samples by the optimized DBSCAN. To synthesize more effective samples, the minority class's noise samples were eliminated. Thus, the core samples along with borderline samples were oversampled by utilizing distinct strategies. Regarding precision, recall, and F-value, a better outcome was demonstrated by the experiential outcomes. However, the majority sample's information was not properly utilized and the presented model was susceptible to noise data.

Author name	Approach	Steps involved	Advantage	Drawbacks
Yang Lu *et al.* [65]	Individual Bayes Imbalance Impact Index (IBI^3)	The imbalance problem of a single minority class sample was measured by IBI^3. The degradation degree in the imbalanced data set was determined by BI^3.	Obtained higher correlated data with high F-scores.	The imbalance issue was not solved effectively since each type of classifier had distinct sensitivity.
Zhining *et al.* [66]	Self-paced Ensemble (SPE) approach.	The concern in classifying the classifier's samples was tackled by a classification hardness strategy. Here, better outcomes were attained by the resampling strategy which was guided by this hardness approach.	Robust performance even under highly overlapping classes.	The performance was affected by the computational cost's dependency on the dataset.
Mugdha *et al.* [67]	Convolutional Neural Network-based ELM	Initially, filters were used to learn the intuitive features like edges and basic shapes. More abstract features like texture were	By diminishing the overfitting problem, the translation variance was elevated by the CNN's utilization.	Some issues like elevated error rate and instability in classification were generated by the imbalance.

		acquired followed by features classification.			**Adamu *et al.* [70]**	Generative Adversarial Network(GAN)	A fine-grained generation was assured by utilizing multiple fake classes along with minority class instance's classification. Additionally, minority class instances aimed at rebalancing were also engendered.	Superiority regarding classification performance along with generated sample quality.	The performance was affected by the formation of poor minority class samples.
Linbin *et al.*[68]	CI loss	The imbalance degree (ID) in the process of gradient descent was alleviated by a new loss function that could be employed among arbitrary imbalanced datasets.	Elevated recognition rate.	The efficiency was affected by the presented method's dependency on the number of samples.					
Borowska *et al.* [69]	A Rough Granular computing (RGA)Approach	The granules with parameters were introduced initially. Utilizing the RGA approach, the parameters were tuned. RGA involved pre-processing followed by selective oversampling.	Small disjuncts, class overlapping, along with noises were carefully removed.	The complexity was elevated by the tendency to over-generalization and variance.	**Xiaofen *et al.* [71]**	Weighted ELM centered on Hybrid Artificial Bee Colony (WELM-HABC)	The ABC algorithm optimized the input bias initially. The HABC optimized the weights allocated to the training samples. The perturbed parameter vector's diversities of differential evolution	Superior classification performance with efficient AUC.	Failed to work in multi-class datasets.

		combine with the ABC's best solution information in the HABC effectively.		
Fatima *et al.* [72]	Multilayer Perceptron (MLP) and ELM.	The Credit card transaction's vector features were comprised in the input layer initially. The weighted inputs as of the input layer were received by the first HL that forward the earlier layer's data to the subsequent one. The classification outcome was contained in the output layer finally.	New fraudulent transactions were predicted rapidly with 95.46% accuracy.	The performance was affected by the false alarm's presence.

2.6 Performance analysis

The ELM algorithm's performance is validated here by means of a CI problem. Various models like SMOTE-CSELM [11], UBKELM [12], Ensemble-based WELM [17], and Demand Response Algorithm OSCSELM [13] were utilized by the prevailing schemes for the CI issue in crime scenes. Regarding Geometric Mean (G-Mean), Accuracy, and IR, the performance analysis are exhibited in Figures 3, 4, and 5.

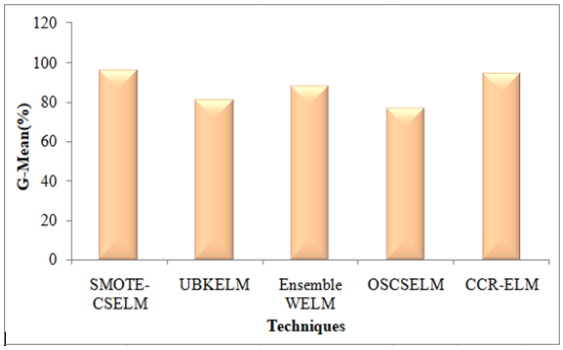


Figure 3: Performance evaluation based on G-Mean.

The model's G-mean is depicted in figure 3. 96.02% G-Mean was attained by the SMOTE-CSELM method whereas the UBKELM obtained only 81%. Similarly, 88.13%, 76.87%, and 94.50% G-mean were shown by the Ensemble-based WELM, OSCSELM, and the CCR-ELM method. Here, the SMOTE-CSELM approach is identified as more efficient than the other methodologies.

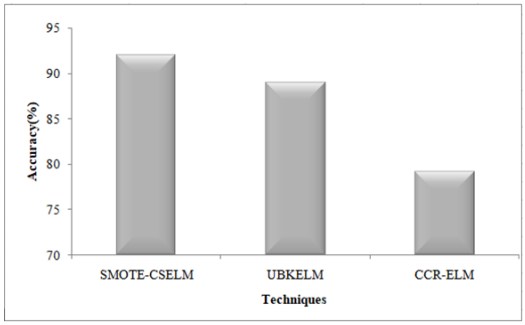


Figure 4: Performance assessment based on accuracy.

Grounded on accuracy rate, the performance of diverse models utilized in imbalance classification was revealed in figure 4. The number of appropriately predicted data points out of all data points is defined as Accuracy. Here, SMOTE-CSELM's accuracy is higher than the prevailing schemes. The SMOTE-CSELM's accuracy is 92%, whereas lower accuracy was attained by the CCR-ELM than the prevailing schemes; the UBKELM method's accuracy is 89%, which is much better than the CCR-ELM having 79.22%. It is concluded

that better outcomes were attained by the SMOTE-CSELM than other approaches.

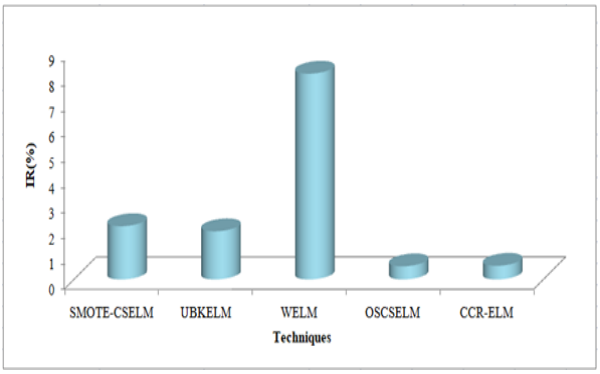


Figure 5: Performance evaluation based on Imbalance Ratio.

Regarding IR, other prevailing model performance evaluations were depicted in figure 5. The ratio of the number of examples belonging to the minority class to the majority class is named IR. The imbalance problem is higher if the IR value is smaller. It is evident that a lower IR value was exhibited by the OSCSELM method (0.52%) inflicting large imbalance issues in the dataset. Also, 2.1%, 8.11%, and 1.90% were the IR of SMOTE-CSELM, UBKELM, and ensemble-based WELM, respectively.

3 Conclusion

A significant issue in ML is classification with imbalanced class distributions. Regarding this topic, more related studies along with findings have been presented. A research review on several extreme ML techniques for classifying imbalances is presented here, along with a brief discussion on their drawbacks. Various prevailing models of extreme ML approaches for imbalance classification like Ensemble-based ELM for CI problems, Semi-supervised ELM algorithms for CI problems, and Kernel-based ELM for CI problems was explained by this literature work. Grounded on G-Mean, Accuracy, and IR performance indicators, the techniques' performance is examined. Finally, superior outcomes were exhibited by SMOTE-CSELM-based approaches than other methods that were concluded from this review. This review article recommends that future studies focus on making the system more reliable for imbalance classification via optimization and hybridized ML enhancements.

Declarations

Funding: No funds, grants were received by any of the authors.

Conflict of interest: There is no conflict of interest among the authors.

Data availability: All data generated or analysed during this study are included in the manuscript.

Code availability: Not applicable.

Author's contributions: All Author is contributed to the design and methodology of this study, the assessment of the outcomes and the writing of the manuscript.

References

[1] Haseeb Ali, Mohd Najib Mohd Salleh, Rohmat Saedudin, Kashif Hussain and Muhammad Faheem Mushtaq. Imbalance class problems in data mining a review. Indonesian Journal of Electrical Engineering and Computer Science, 14(3):1560-1571, 2019.

[2] Paula Lauren, Guangzhi Qu, Feng Zhang and Amaury Lendasse. Discriminant document embeddings with an extreme learning machine for classifying clinical narratives. *Neurocomputing*, 277:129-138, 2017.

[3] Li Li, Ruizhi Sun, Saihua Cai, Kaiyi Zhao and Qianqian Zhang. A review of improved extreme learning machine methods for data stream classification. *Multimedia Tools and Applications*, 78(24):33375-33400, 2019.

[4] Jinhong Huang, Zhu Liang Yu and Zhenghui Gu. A clustering method based on extreme learning machine. *Neurocomputing*, 277:108-119, 2017.

[5] Jian Wang, Siyuan Lu, Shui-Hua Wang and Yu-Dong Zhang. A review on extreme learning machine. *Multimedia Tools and Applications*, 80(11-13):1-50, 2021.

[6] Peter Adeniyi Alaba, Segun Isaiah Popoola, Lanre Olatomiwa, Mathew Boladele Akanle, Olayinka Ohunakin, Emmanuel Adetiba, Opeoluwa David Alex, Aderemi A. A Atayero and Wan Mohd Ashri Wan Daud. Towards a more efficient and cost-sensitive extreme learning machine a state-of-the-art review of recent trend. *Neurocomputing*, 350:70-90, 2019.

[7] Huimin Pei, Kuaini Wang, Qiang Lin and Ping Zhong. Robust semi-supervised extreme learning machine. *Knowledge-Based Systems*, 159:203-220, 2018.

[8] Xiong Luo, Changwei Jiang, Weiping Wang, Yang Xu, Jenq-Haur Wang and Wenbing Zhao. User behavior prediction in social networks using weighted extreme learning machine with distribution optimization. *Future Generation Computer Systems*, 93:1023-1035, 2018.

[9] Kemal Akyol. Comparing of deep neural networks and extreme learning machines based on growing and pruning approach. *Expert Systems with Applications*, 140:1-14, 2019.

[10] Parashjyoti Borah and Deepak Gupta. Unconstrained convex minimization based implicit Lagrangian twin extreme learning machine for classification (ULTELMC). *Applied Intelligence*, 50(4):1327-1344, 2020.

[11] Bhagat Singh Raghuwanshi and Sanyam Shukla. SMOTE based class-specific extreme learning machine for imbalanced learning. *Knowledge-Based Systems*, 187:1-46, 2019.

[12] Bhagat Singh Raghuwanshi and Sanyam Shukla. Class imbalance learning using underbagging based kernelized extreme learning machine. *Neurocomputing*, 329:172-187, 2018.

[13] Sanyam Shukla and Bhagat Singh Raghuwanshi. Online sequential class-specific extreme learning machine for binary imbalanced learning. *Neural Networks*, 119:235-248, 2019.

[14] Hualong Yu, Xibei Yang, Shang Zheng and Changy. Active learning from imbalanced data a solution of online weighted extreme learning machine. *IEEE Transactions on Neural Networks and Learning Systems*, 30(4):1088-1103, 2018.

[15] Wendong Xiao, Jie Zhang, Yanjiao Li, Sen Zhang and Weidong Yang. Class-specific cost regulation extreme learning machine for imbalanced classification. *Neurocomputing*, 261:70-82, 2016.

[16] Bhagat singh Raghuwanshi and Sanyam Shukla. Class-specific extreme learning machine for handling binary class imbalance problem. *Neural Networks*, 105:206-217, 2018.

[17] Yong Zhang, Bo Liu, Jing Cai and Suhua Zhang. Ensemble weighted extreme learning machine for imbalanced data classification based on differential evolution. *Neural Computing and Applications*, 28(12):259-267, 2016.

[18] Hui Li, Xi Yang, Yang Li, Li-Ying Hao and Tian-Lun Zhang. Evolutionary extreme learning machine with sparse cost matrix for imbalanced learning. *ISA Transactions*, 100:198-209, 2019.

[19] Chengbo Lu, Haifeng Ke, Gaoyan Zhang, Ying Mei and Huihui Xu. An improved weighted extreme learning machine for imbalanced data classification. *Memetic Computing*, 11(1):27-34, 2017.

[20] Tianlei Wang, Jiuwen Cao, Xiaoping Lai and Badong Chen. Deep weighted extreme learning machine. *Cognitive Computation*, 10(6):890-907, 2018.

[21] Honghao Zhu, Guanjun Liu, Mengchu Zhou, Yu Xie, Abdullah Abusorrah and Qi Kang. Optimizing weighted extreme learning machines for imbalanced classification and application to credit card fraud detection. *Neurocomputing*, 407:50-62, 2020.

[22] Arnis Kirshners, Sergei Parshutin and Henrihs Gorskis. Entropy-based classifier enhancement to handle imbalanced class problem. *Procedia Computer Science*, 104:586-591, 2017.

[23] Xiaoyan Zhu, Chenzhen Ying, Jiayin Wang, Jiaxuan Li, Xin Lai and Guangtao Wang. Ensemble of ML-KNN for classification algorithm recommendation. *Knowledge-Based Systems*, 221:1-10, 2021.

[24] Wentao Mao, Jinwan Wang and Zhanao Xue. An ELM-based model with sparse-weighting strategy for sequential data imbalance problem. *International Journal of Machine Learning and Cybernetics*, 8(4):1333-1345, 2016.

[25] Junhai Zhai, Sufang Zhang and Chenxi Wang. The classification of imbalanced large data sets based on MapReduce and ensemble of ELM classifiers. International Journal of Machine Learning and Cybernetics, 8(6):1009-1017, 2015.

[26] Shifei Ding, Nan Zhang, Jian Zhang, Xinzheng Xu and Zhongzhi Shi. Unsupervised extreme learning machine with representational features. *International Journal of Machine Learning and Cybernetics*, 8(2):1-9, 2015.

[27] Fulong Ren, Peng Cao, Wei Li, Dazhe Zhao and Osmar Zaiane. Ensemble based adaptive over-sampling method for imbalanced data learning in computer aided detection of microaneurysm. *Computerized Medical Imaging and Graphics*, 55:54-67, 2016.

[28] Zhenyu Wu, Wenfang Lin and Yang Ji. An integrated ensemble learning model for imbalanced fault diagnostics and prognostics. *IEEE Access*, 6:8394-8402, 2018.

[29] Guillem Collell, Drazen Prelec and Kaustubh R Patil. A simple plug-in bagging ensemble based on threshold-moving for classifying binary and multiclass imbalanced data. *Neurocomputing*, 275:330-340, 2017.

[30] Hossam Faris, Ruba Abukhurma, Waref Almanaseer, Mohammed Saadeh, Antonio M Mora, Pedro A Castillo and Ibrahim Aljarah. Improving financial bankruptcy prediction in a highly imbalanced class distribution using oversampling and ensemble learning a case from the Spanish market. *Progress in Artificial Intelligence*, 9(1):31-53, 2019.

[31] Yanjiao Li, Sen Zhang, Yixin Yin, Wendong Xiao and Jie Zhang. Parallel one-class extreme learning machine for imbalance learning based on Bayesian

approach. *Journal of Ambient Intelligence and Humanized Computing*, 2018, https://doi.org/10.1007/s12652-018-0994-x.

[32] Zhiqiong Wang, Junchang Xin, Hongxu Yang, Shuo Tian, Ge Yu, Chenren Xu and Yudong Yao. Distributed and weighted extreme learning machine for imbalanced big data learning. *Tsinghua Science and Technology*, 22(2):160-173, 2017.

[33] Junjie Wang, Zhenghui Gu, Zhuliang Yu and Yuanqing Li. An online semi-supervised P300 speller based on extreme learning machine. *Neurocomputing*, 269:148-151, 2016.

[34] Nan Zhang and Shifei Ding. Unsupervised and semi-supervised extreme learning machine with wavelet kernel for high dimensional data. *Memetic Computing*, 9(2):129-139, 2016.

[35] Yongxiang Lei, Lihui Cen, Xiaofang Chen and Yongfang Xie. A hybrid regularization semi-supervised extreme learning machine method and its application. *IEEE Access*, 7:30102-30111, 2019.

[36] Jin Xie, Sanyang Liu and Hao Dai. Manifold regularization based distributed semi-supervised learning algorithm using extreme learning machine over time-varying network. *Neurocomputing*, 355:24-34, 2019.

[37] Hongyu Zhu and Xizhao Wang. A cost-sensitive semi-supervised learning model based on uncertainty. *Neurocomputing*, 251:106-114, 2017.

[38] Carlos A. S da Silva and Renato A Krohling. Semi-supervised online elastic extreme learning machine for data classification. *International Joint Conference on Neural Networks (IJCNN)*, 08-13 July 2018, Rio de Janeiro, Brazil, 2018.

[39] Jie Yang, Jiuwen Cao, Tianlei Wang, Anke Xue and Badong Chen. Regularized correntropy criterion based semi-supervised ELM. *Neural Networks*, 122:117-129, 2019.

[40] Parsa Vafaie, Herna Viktor and Wojtek Michalowski. Multi-class imbalanced semi-supervised learning from streams through online ensembles. *International Conference on Data Mining Workshops (ICDMW)*, 17-20 November 2020, Sorrento, Italy, 2020.

[41] Zheng Liu, Shiluo Huang, Wei Jin and Ying Mu. Broad learning system for semi-supervised learning. *Neurocomputing*, 444:38-47, 2021.

[42] Adnan OM Abuassba, Zhang Dezheng and Zahid Mahmood. Semi-supervised multi-kernel extreme learning machine. *Procedia Computer Science*, 129:305-311, 2018.

[43] Chuangquan Chen, Yanfen Gan and Chi-Man Vong. Extreme semi-supervised learning for multiclass classification. *Neurocomputing*, 376:103-118, 2019.

[44] Weipeng Cao, Jinzhu Gao, Zhong Ming, Shubin Cai and Zhiguang Shan. Fuzziness based online sequential extreme learning machine for classification problems. *Soft Computing*, 22(13-14):3487-3494, 2018.

[45] Lu Li, Chengyi Wang, Wei Li and Jingbo Chen. Hyperspectral image classification by adaboost weighted composite kernel extreme learning machines. *Neurocomputing*, 275:1725-1733, 2017.

[46] Ping Yang, Dan Wang, Wen-Bing Zhao, Li-Hua Fu, Jin-Lian Du and Hang Su. Ensemble of kernel extreme learning machine based random forest classifiers for automatic heartbeat classification. *Biomedical Signal Processing and Control*, 63:1-13, 2021.

[47] Bhagat Singh Raghuwanshi and Sanyam Shukla. UnderBagging based reduced Kernelized weighted extreme learning machine for class imbalance learning. *Engineering Applications of Artificial Intelligence*, 74:252-270, 2018.

[48] Bhagat Singh Raghuwanshi and Sanyam Shukla. Generalized class-specific kernelized extreme learning machine for multiclass imbalanced learning. *Expert Systems with Applications*, 121:244-255, 2018.

[49] Bhagat Singh Raghuwanshi and Sanyam Shukla. Classifying imbalanced data using ensemble of reduced kernelized weighted extreme learning machine. International *Journal of Machine Learning and Cybernetics*, 10(8):3071-3097, 2019.

[50] Shuya Ding, Bilal Mirza, Zhiping Lin, Jiuwen Cao, Xiaoping Lai, Tam V Nguyen and Jose Sepulveda. Kernel based online learning for imbalance multiclass classification. Neurocomputing, 277):139-148, 2017.

[51] Bhagat Singh Raghuwanshi and Sanyam Shukla. Classifying imbalanced data using BalanceCascade-based kernelized extreme learning machine. *Pattern Analysis and Applications*, 23(3):1157-1182, 2019.

[52] Hongyun Qin, Houpan Zhou and Jiuwen Cao. Imbalanced learning algorithm based intelligent abnormal electricity consumption detection. *Neurocomputing*, 402:112-123, 2020.

[53] Yang Wang, An-Na Wang, Qing Ai and Hai-Jing Sun. Enhanced kernel-based multilayer fuzzy weighted extreme learning machines. *IEEE Access*, 8:166246-166260, 2020.

[54] Zhennao Cai, Jianhua Gu, Jie Luo, Qian Zhang, Huiling Chen, Zhifang Pan, Yuping Li and Chengye Li. Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf

optimization strategy. *Expert Systems With Applications*, 138:1-29, 2019.

[55] Pattaramon Vuttipittayamongkol, Eyad Elyan and Andrei Petrovski. On the class overlap problem in imbalanced data classification. *Knowledge-Based Systems*, 212:1-17, 2021.

[56] El Barakaz Fatima, Boutkhoum Omar, El Moutaouakkil Abdelmajid, Furqan Rustam, Arif Mehmood and Gyu Sang Choi. Minimizing the overlapping degree to improve class-imbalanced learning under sparse feature selection application to fraud detection. *IEEE Access*, 9:28101-28110, 2021.

[57] Everlandio R. Q Fernandes and Andre C. P. L. F de Carvalho. Evolutionary inversion of class distribution in overlapping areas for multi-class imbalanced learning. *Information Sciences*, 494:141-154, 2019.

[58] Mingjing Wang, Huiling Chen, Huaizhong Li, Zhennao Cai, Xuehua Zhao, Changfei Tong, Jun Li and Xin Xu. Grey wolf optimization evolving kernel extreme learning machine application to bankruptcy prediction. Engineering *Applications of Artificial Intelligence*, 63:54-68, 2017.

[59] Dong Zhao, Chunyu Huang, Yan Wei, Fanhua Yu, Mingjing Wang and Huiling Chen. An effective computational model for bankruptcy prediction using kernel extreme learning machine approach. *Computational Economics*, 49(2):1-17, 2016.

[60] Yan Wei, Huijing Lv, Mengxiang Chen, Mingjing Wang, Ali Asghar Heidari, Huiling Chen and Chengye Li. Predicting entrepreneurial intention of students an extreme learning machine with gaussian barebone harris hawk's optimizer. *IEEE Access*, 2017, https://doi.org/10.1109/ACCESS.2020.2982796.

[61] Yongshan Zhang, Jia Wu, Chuan Zhou and Zhihua Cai. Instance cloned extreme learning machine. *Pattern Recognition*, 68:52-65, 2017.

[62] Xiaowei Xue, Min Yao and Zhaohui Wu. A novel ensemble-based wrapper method for feature selection using extreme learning machine and genetic algorithm. *Knowledge and Information Systems*, 57(2):389-412, 2017.

[63] Ali Asghar Heidari, Rahim Ali Abbaspour and Huiling Chen. Efficient boosted grey wolf optimizers for global search and kernel extreme learning machine training. *Applied Soft Computing*, 81:1-57, 2019.

[64] Xu Xiaolong, Chen Wen and Sun Yanfei. Over-sampling algorithm for imbalanced data classification. *Journal of Systems Engineering and Electronics*, 30(6):1182-1191, 2019.

[65] Yang Lu, Yiu-Ming Cheung and Yuan Yan Tang. Bayes imbalance impact index a measure of class imbalanced data set for classification problem. *IEEE Transactions on Neural Networks and Learning* Systems, 31(9):3525-3539, 2019.

[66] Zhining Liu, Wei Cao, Zhifeng Gao, Jiang Bian, Hechang Chen, Yi Chang and Tie-Yan Liu. Self-paced ensemble for highly imbalanced massive data classification. *IEEE 36th International Conference on Data Engineering (ICDE)*, 20-24 April 2020, Dallas, TX, USA, 2020.

[67] Mugdha Jain, William Andreopoulos and Mark Stamp. Convolutional neural networks and extreme learning machines for malware classification. *Journal of Computer Virology and Hacking Techniques*, 16(17):1-16, 2020.

[68] Linbin Zhang, Caiguang Zhang, Sinong Quan, Huaxin Xiao, Gangyao Kuang and Li Liu. A class imbalance loss for imbalanced object recognition. IEEE *Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:2778-2792, 2020.

[69] Borowska K and Stepaniuk J. A rough–granular approach to the imbalanced data classification problem. *Applied Soft Computing*, 83(4):1-40, 2019.

[70] Adamu Ali-Gombe and Eyad Elyan. MFC-GAN Class-imbalanced dataset classification using multiple fake class generative adversarial network. *Neurocomputing*, 361:212-221, 2019.

[71] Xiaofen Tang and Li Chen. Artificial bee colony optimization-based weighted extreme learning machine for imbalanced data learning. *Cluster Computing*, 22(3):6937-6952, 2018.

[72] Fatima Zohra El hlouli, Jamal Riffi, Mohamed Adnane Mahraz, Ali El Yahyaouy and Hamid Tairi. Credit card fraud detection based on multilayer perceptron and extreme learning machine architectures. *International Conference on Intelligent Systems and Computer Vision (ISCV)*, 09-11 June 2020, Fez, Morocco, 2020.

Machine Bias: A Survey of Issues

Ana Farič *, Ivan Bratko
Faculty of Computer and Information Science, University of Ljubljana, Ljubljana, Slovenia
E-mail: af27987@student.uni-lj.si, bratko@fri.uni-lj.si
*Corresponding author

Keywords: machine learning, artificial intelligence, bias, fairness, discrimination, COMPAS

Received: April 1, 2024

Some recent applications of Artificial Intelligence, particularly machine learning, have been strongly criticised in general media and professional literature. Applications in domains of justice, employment and banking are often mentioned in this respect. The main critic is that these applications are biased with respect to so called protected attributes, such as race, gender and age. The most notorious example is the system COMPAS which is still in use in the American justice system despite severe criticism. The aim of our paper is to analyse the trends of discussion about bias in machine learning algorithms using the COMPAS as an example. The main problem we observed is that even in the field of AI, there is no generally agreed upon definition of bias which would enable operational use in preventing bias. Our conclusions are that (1) improved general education concerning AI is needed to enable better understanding of AI methods in everyday applications, and (2) better technical methods must be developed for reliably implementing generally accepted societal values such as equality and fairness in AI systems.

Povzetek: Analizirali smo trende v diskusijah o pristranskosti odločitev strojnega učenja, kjer smo za primer vzeli sistem COMPAS.

1 Introduction

With the widespread use of machine learning, there have been cases in the last 5 to 10 years where applications received significantly negative feedback for being biased, primarily from the general media and also within professional literature. Typical applications come from domains such as judiciary, employment, and banking. Critics warn that "machine learning algorithms and systems are unfair and biased" with respect to so-called protected attributes, such as race, gender, and age of an individual. They argue that artificial intelligence recommendations depend on these attributes rather than on the objective evaluation of facts [15].

Some noteworthy article headlines describing discriminatory practices allegedly promoted by machine learning algorithms include: "There's software used across the country to predict future criminals. And it's biased against blacks [2]," "New Zealand passport robot tells applicant of Asian descent to open eyes [24]," "A beauty contest was judged by AI and the robots didn't like dark skin [19]," and "Amazon scraps secret AI recruiting tool that showed bias against women [10]." Such examples contribute to escalating concerns (and sometimes panic) about the potentially harmful impacts of artificial intelligence (AI) on our lives [20]. Experts from various fields address the issue of machine learning bias, attempting to define what bias means, where it originates, and, most importantly, what should be done about it.

In the evolving field of ethics in AI (e.g., UNESCO 2021 [26]), the topic of machine learning bias prominently appears. Policymakers often mention it in relation to regulatory principles aimed at ensuring the ethical use of AI (e.g., European AI Act, 2023, 2024 [3]). However, in these discussions, it is often not clear what exactly machine learning bias and AI bias mean. Therefore, regulatory measures in this direction are not clearly defined, except in a very abstract form. The term bias in relation to machine learning means different things to different authors. Even in the AI literature, there is no complete consensus and no universally accepted technical definitions of bias that could be operationally used to prevent bias [15]. For various meaningful definitions of bias, it has even been mathematically proven that, except in special cases, they cannot be satisfied simultaneously [17].

In this paper, we review various definitions of bias and different opinions on how to address the problem most effectively in practice. The conclusions converge towards the idea that addressing bias appropriately requires considering societal values and operationalizing them through interdisciplinary collaboration with a democratically accepted social agreement in the form of appropriate legislation. Better general education on AI and its methods would contribute to a better general understanding of bias in AI in practice. The lack of uniformity in dealing with bias in AI will be in this paper demonstrated using the example of the COMPAS system [2, 11, 12, 16].

2 COMPAS

The COMPAS system is considered in a series of publications (Figure 1) as arguably the most controversial case illustrating the bias of AI. COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is a decision-making system used by many American courts where judges assess the risk of recidivism, specifically estimating the likelihood that an offender will reoffend within two years if released. COMPAS was developed by an American company, then known as Northpointe (Equivant today). COMPAS takes into account 137 attributes for each offender, obtained either from the individual or their criminal record. This data is analyzed by a specific algorithm, which, as a trade secret of the company, is not publicly known. Based on this analysis, the algorithm provides a score ranging from 1 to 10, where a higher score indicates a higher risk of recidivism.

Figure 1 illustrates 10 highly cited articles on this system and the mutual citation between articles. An arrow from paper A to paper B indicates that paper B is cited in paper A.

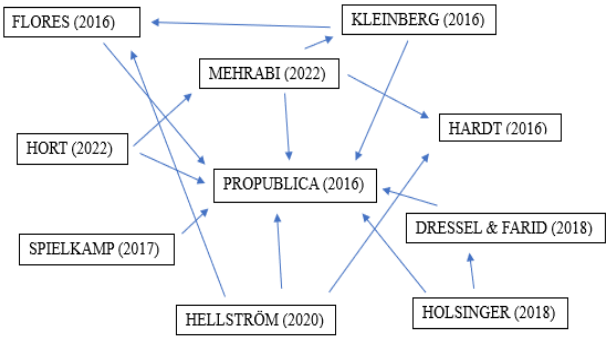


Figure 1: Interconnectedness of publications on the COMPAS system.

At the center of the graph is an article from the ProPublica newsroom [2], which, according to [1], sparked interest in studying bias in AI. In [2], a group of investigative journalists described their analysis of the COMPAS system and experiments with real data on over 7,000 defendants from Broward County, Florida in the years 2012 and 2013. They found that only 61% of people assessed as likely to reoffend actually did so. In further analysis, they focused mainly on the racial aspect, concluding that the program is biased against African American defendants. They monitored how many of them were re-convicted in the next two years and compared predictions with actual outcomes. 44.9% of African Americans marked as high risk did not reoffend. In contrast, 47.7% of whites marked as low risk reoffended within two years. These two metrics for system mispredictions are commonly referred to as (1) FPR (false positive rate), which is the proportion of the negative class incorrectly predicted as positive, and (2) FNR (false negative rate), which is the proportion of the positive class incorrectly predicted as negative. Complete results regarding the FPR and FNR rates in the COMPAS system predictions were:

	White	Black
FPR	23.5%	44.9%
FNR	47.7%	28.0%

These results were interpreted as evidently biased against African American defendants, leading to an assessment of the use of the COMPAS system as inappropriate and discriminatory [2]. This conclusion seems quite justified.

An additional issue highlighted in [2] is the fact that the decision criterion used by COMPAS is not transparent, as the algorithm is protected as a trade secret. COMPAS itself does not provide an explanation for its predictions. This article is highly cited, and consequently, COMPAS became the most well-known example of bias in machine learning, both within the professional circles of machine learning and among the public without expertise in AI. Despite this, COMPAS is still in use.

In response to the ProPublica article, a group of experts from the American justice system published a rejoinder with the telling title "False positives, false negatives and false analysis: a rejoinder to Machine Bias …" [12] They pointed out several controversial decisions in ProPublica's analysis, conducted their own experimental research, and concluded that ProPublica's assertions were incorrect. While this criticism seems justified, it would be more convincing if they clearly showed where the crucial mistake in ProPublica occurred. Instead, they presented their own experimental results, claiming that offenders are treated fairly regardless of race. They obtained these results by considering risk assessments of offenders on a scale from 1 to 10, as assessed by COMPAS. From these assessments, they calculated the AUC (Area under ROC curve), a standard performance measure in machine learning systems. AUC is interesting because it is equal to the probability that the predictive system correctly distinguishes between positive and negative cases. This means that, if we take two random cases (two defendants), one positive (did reoffended) and the other negative (did not reoffend), the system will correctly determine which is positive and which is negative with the probability AUC. [12] reported that the AUC value for whites was 0.69, and 0.70 for African Americans. The difference is not statistically significant. From this, they concluded that COMPAS is not racially discriminatory, and ProPublica's results indicating discrimination cannot be correct. However, this indirect argument allows for doubt since AUC, FPR, and FNR are not uniquely related to each other.

Dressel and Farid [11] reported on a relevant experiment where they were interested in the accuracy of predictions about the risk of recidivism achieved by randomly selected people without domain knowledge. They also compared the accuracy of COMPAS with that of a simple

linear classifier. They conducted a human prediction experiment (performed with crowd sourcing) on a subset of Broward County data (about 1000 out of a total of around 7000 defendants) from the studies [2] and [12]. Since using all 137 attributes for prediction by people would be impractical, they only used 7 selected attributes from the original set. The predictive accuracy of non-experts in these experiments was surprisingly almost the same as that of the COMPAS system. Interestingly, human predictions in this experiment were similarly biased as COMPAS, measured by FPR and FNR for whites and African Americans. These results hardly changed when additional information about the race was given to the human evaluator. They also found that a simple linear classifier achieved a similar predictive accuracy using only two attributes, and more sophisticated classifiers did not improve predictive accuracy (or fairness).

Holsinger et al. [16] criticized the study by Dressel and Farid [11]. The criticism is based on the following arguments. Participants were recruited via Amazon's Mechanical Turk, and they received payment for participation. The participants were only shown the values of seven selected attributes: age, gender, offense type, offense severity, adult convictions, juvenile felony charges, and juvenile misdemeanor charges. All these attributes are known as important risk factors for recidivism. According to Holsinger et al. [16], this reduction of the original set of 137 attributes made the prediction task easier than the original task with 137 attributes. This view is indeed justified as it is known that appropriate selection of useful attributes in machine learning may be difficult. After providing an individual rating, participants received feedback regarding the correctness of the answer and their average accuracy. Participants who achieved high accuracy were rewarded with slightly higher payment. All this significantly differs from the real context where expert decision-makers face a plethora of (often irrelevant and biased) information, which makes their task more difficult [16]. A counter argument could be used here that these decision-makers could easily inform themselves about the most important recidivism factors by a simple web search. But the authors in [16] conclude that, in the context of all circumstances, Dressel and Farid did not reveal anything new and just "rediscovered what has been well-established in a large body of risk assessment literature: Compared to unstructured human judgement, structured human judgment and actuarial approaches are more accurate. Structuring decisions limits consideration and unnecessary emphasis on factors that are unrelated to risk of recidivism (i.e. bias)."

Regardless of these results, Rudin [22] used machine learning to synthesize a very simple and completely understandable predictive model from the mentioned Florida data. This model comprises three simple if-then rules (shown below) and uses three attributes only (gender, age and number of past crimes). Unlike the COMPAS model, these rules are trivially understandable:

```
IF age between 18-20 and sex is male
  THEN predict arrest (within 2 years)
  ELSE IF age between 21-23 and 2-3 prior offenses
    THEN predict arrest
    ELSE IF more than three priors
      THEN predict arrest
      ELSE predict no arrest
```

This predictor is equally accurate for recidivism prediction as COMPAS, and it has similar FPR and FNR on data from Broward County, Florida.

From the described results, the recidivism predictive problem seems challenging despite extensive available information about the defendant, and better accuracy apparently cannot be achieved. At the same time, almost everything seems to be achieved with just three most useful attributes, and the additional 130+ attributes do not bring anything substantially new. From this, the authors in [2] and [11] jump to the conclusion that the use of machine learning in the justice system is not promising in general. This is, of course, a hasty and overly simplistic conclusion, as Spielkamp [23] points out. In many other applications, machine learning has surpassed the predictive accuracy of experts, as for example confirmed already by many early experiments with machine learning in medical diagnosis [6].

All these different opinions regarding bias and usefulness of the COMPAS system highlight the lack of universally accepted operational definitions of bias and fairness in machine learning. This situation is nicely illustrated by the highly cited article by Mehrabi et al. [20], which discusses more than a dozen relevant definitions but does not provide a synthesis that would limit this conceptual complexity and offer a practically useful approach. Additionally, this article causes further uncertainty by quickly dismissing the COMPAS system and categorizing it as evidently biased, flawed, and useless, without addressing counter arguments in [12].

3 Issues in definitions of bias and fairness

In the general media, machine learning is often accused of bias based more on intuition, without precisely defining mathematically verifiable criteria by which bias can be detected. Statements such as "the system has shown bias towards people of color in the judiciary," [2] or "the system is biased against women in employment evaluations," [5] use general phrases like "bias of algorithms," "algorithmic bias", "machine learning bias" or "artificial intelligence bias." Sometimes these statements are accompanied by a simple explanation like "machine learning systems are developed almost exclusively by white men, so ..."

Today, it is clear that the matter is not so trivial. Overly simplistic explanations are becoming rarer. It is also becoming clearer that the phrase "algorithm bias" is not suitable and gives the wrong impression that algorithms can have malicious intentions and do not operate according to mathematical and statistical principles [20]. The goal of these methods is always to discover the laws that apply in the real world from real-world data. However, a problem arises if biased practices already exist in the real world. Data collected in such a world reflects this bias, and the learning algorithm detects and reproduces this bias. If the results obtained from biased data in the real world are then used again in the real world, we will reproduce the existing bias [14].

However, it is still not precisely defined what bias actually is. Often, it is an impression of bias, where biases for or against an individual or a group manifest in a way that is perceived as unjust [21].

Let us look at the problems with defining bias. In the field of machine learning, various explanations are found, all meaningful in their own ways. The term bias in machine learning refers to several phenomena [15]:

1. Inductive Bias: This is the principle by which an algorithm chooses one of the typically large number of possible hypotheses, all in some way justified by the training data. It is a set of assumptions made by a learning algorithm to generalize a finite set of observations (training data) into a general model of the domain [17]. This type of bias is a necessary component of machine learning, without it machine learning is not possible. An example of such bias is Occam's razor, which says: if we have two explanations that explain the data equally well, we should choose the simpler one [13], [15], [21]. Although the term bias has a negative connotation, inductive bias is a positive and even an inevitable component of machine learning, as explained by [15], and is a basic concept in AI textbooks.

2. Bias in Training Data: This bias reflects actual biases in established decision-making in a given application area (e.g., bias in the judgments of experts in actual judicial practice in the environment from which the training data are drawn) [5], [10]. [1] emphasizes that biases in training data can be attributed to cognitive biases of human thinking. It is a natural phenomenon where the human brain filters infinite types of information in a way that retains what is relevant to us. Because algorithms are trained on data representing human behavior, they reflect these cognitive biases. This bias is referred to as negative legacy [5], or as historical bias [15].

3. Bias from Improper Data Collection or Sampling: For example, if there are significantly fewer examples available for a particular group of people than for other groups, according to mathematically grounded statistical and probability principles, some groups, typically minorities, appear to be discriminated against simply because probability estimation methods correctly assess probabilities differently when there is little data available. This bias is referred to as underestimation [5]. [25] emphasizes that we must also question where the training data come from. If algorithms traditionally relied on reliable labels determined by experts, today algorithms may learn from data originating from the broader society, where labels and patterns are often biased.

The above sources of bias are relatively widely accepted. However, the problem remains of how to precisely define criteria that objectively indicate whether a system is biased or to quantitatively assess that bias. There are numerous measures that seem relevant, but they may turn out to be contradictory, and for now, there is no simple, universally accepted measure.

The situation is well illustrated by the comprehensive review of different definitions of fairness by Mehrabi et al. [20]. In “fairness through awareness”, an algorithm is considered fair if it gives similar predictions to similar individuals; in “treatment equality,” ; the ratio of FNR and FPR in both groups (based on a protected attribute) is the same; in “fairness in relational domains,” ; in addition to attributes, social, organizational, and other connections between individuals are considered; in “fairness through unawareness,” ; an algorithm is considered fair as long as any protected attributes are not explicitly used in the decision-making process.

Berk et al. [4] make a similar point. They examine different ways that fairness can be formally defined, how these different kinds of fairness can be incompatible, how risk assessment accuracy can be affected, and various algorithmic remedies that have been proposed. They conclude that, except in most trivial cases, it is impossible to maximize accuracy and fairness at the same time and impossible to simultaneously satisfy all kinds of fairness. Kleinberg et al. [18] explore this problem more thoroughly. They define three natural, seemingly obvious conditions that a system must meet to be unbiased (fair). However, it turns out that these three conditions cannot be satisfied simultaneously, except in (trivial) special cases that are uninteresting for practical purposes. So, these three basic requirements together are unattainable. These three requirements are:

(1) Calibration of Probability Estimates: if the algorithm identifies a set of people who are supposed to belong to the positive class with a given probability P, then approximately proportion P of that set must actually belong to the positive class. The same condition must apply to all groups of individuals who differ in the "protected attribute," such as race or gender. In other words, estimates must mean what they are supposed to mean and must be independent of the group (based on protected attributes) to which the individual belongs.

(2) Balance of the Positive Class: the average risk score of the individuals in the positive class must be the same for all groups. In the case of COMPAS, for example, white and black convicts belonging to the positive class should have comparable risk scores.

(3) Balance of the Negative Class: analogous to the average in positive class.

Kleinberg et al. [18] mathematically prove that these three requirements, although essentially aiming for the same goal of reducing bias, are incompatible with each other, except in special cases.
When bias occurs, the question is how to eliminate it. There are various ideas for this, of which the two most obvious are (a) "protected attributes" and (b) reverse discrimination. Typical protected attributes are race and gender.

The principle of protected attributes means that we forbid the learning algorithm from using these attributes when deciding on the classification of an instance. This idea usually does not work well, as the learning algorithm effectively reconstructs their values from other, unprotected attributes that correlate with the protected ones. For example, from data on education or residence location, the algorithm may probabilistically infer a person's race [14].

The principle of reverse discrimination is to deliberately give certain advantages to underprivileged groups in treatment to counteract the effects of discrimination. This measure is obviously well-intentioned, but it actually introduces additional injustice, which is sometimes questionable [1]. Such injustice (reverse discrimination) may be justified, but not from the perspective of fairness, but from the perspective of "higher" values, such as rectifying historical injustices and achieving long-term justice through temporary injustice. Therefore, it is a strategic implementation of socially accepted values that are not easily achievable in practice due to historical reasons and persistence. The difficult question remains to what extent reverse discrimination makes sense. This should be determined by democratically accepted social consensus, formalized with appropriate laws for each case. In practice, addressing bias is approached within three phases of machine learning: (1) pre-processing phase, where we augment the minority sample, (2) mid-processing phase, where we introduce constraints to compensate for an uneven sample, and (3) post-processing phase, where we adjust thresholds for minorities [5][20][21].

When developing methods and tools, we must be aware of potential pitfalls. Alelyani [1] and Holsinger et al. [16] emphasize that certain solutions can lead to new injustices. Chakraborty et al. [7] state that the common side effect of mutating training data is the loss of significant connections between variables or degradation of learner's performance (as measured by accuracy and F1 score). They identify prior decisions that generated the training data as the root causes of bias. They propose a Fair-SMOTE tactic that makes it possible to mitigate bias while maintaining (or even improving) performance at the same time. The key point is to mutate the data in a way that extrapolates all the variables by the same amount. This way, we don't lose connections between variables. Specifically, it finds data imbalance and improperly labeled data points (by situation testing) and then use oversampling to balance the data and remove improperly labeled data points. As an outcome, it generated fair results. They reject worries by [4] that the cost of fairness is a reduction in learner performance. They conclude that it is always better to reflect on the domain and use those insights to guide improvements than blindly applying some optimization methods.

4 Conclusions

Bias has become a popular and controversial topic in some significant machine learning applications. The discussion is dominated by confusion, stemming from the fact that people have different intuitive understandings of the concepts of fairness and bias. Fairness is experienced in various ways, and there is no perfect consensus on the details. Similarly, there is no agreement on what a clear, mathematically formulated criterion should be to unequivocally quantify the bias of a specific system. There is a multitude of disagreements, controversies, and open issues where consensus is lacking. There is no consensus on the origin of bias, nor on which tool or method is most suitable for addressing bias.

The fairness of machine learning should mean producing decisions that society would be satisfied with. However, we are not unified in this regard. The case of COMPAS illustrates how crucial this unity is. COMPAS was tested by multiple experts, and their opinions are entirely contradictory. Some argue that COMPAS is biased, while others say it is not. Spielkamp [23] believes that everyone is correct because they understand fairness in various ways. The study by Kleinberg et al. [18] is particularly illuminating, where the authors mathematically prove the mutual exclusion of certain definitions of fairness that seem equivalent and necessary at first glance. We must clearly define important societal values, consider historical data and the real state of the world, educate ourselves about at least the basic workings of algorithms, and then articulate our expectations accordingly. Only after this can we decide which concept of fairness is suitable in a particular case. Corbet-Davies et al. [8] suggest that, in certain contexts, without a proper understanding of the domain and desired outcomes, acting according to popular formal conceptions of fairness can even have opposite effects than desired. The authors propose that, in algorithm development, instead of adhering to axiomatic notions of fairness, we should focus on their consequences, which strongly depend on the context.

In general literature, there doesn't seem to be anyone anticipating how challenging it will be in practice to

address the issue of bias. Expectations regarding values will need to be precisely formulated with appropriate laws. For example, should reverse discrimination be implemented in a specific application due to historical injustices, and to what extent? This formulation will have to be more technical than usual in regulations and laws, as it will be the basis for the concrete implementation in artificial intelligence algorithms. It is clear that the problem of bias is not solely of a technological nature, and therefore effective approaches to solving the problem must also include broader solutions. We must strive for interdisciplinary research, where AI engineers collaborate with disciplines dealing with ethics, legislation and decision-making [27].

For a proper general understanding and action in this field, there is a need for quality general education. The lack of it is evident in the way information is reported, in people's responses, and even in the confusion of experts. Various algorithms are becoming an inevitable part of our lives. It is unacceptable that we not only know too little about them but also have incorrect perceptions. On the other hand, we have governments and democratic institutions that do not understand the workings of artificial intelligence systems, yet they are the ones who commission and then implement such systems into their decision-making processes. Institutions often lack the knowledge and resources to know how to ask for appropriate algorithmic tools. It is imperative to educate people so that they can articulate what the algorithms should actually measure, what the output should be, and what criteria need to be met for the algorithm to be fair [9].

Acknowledgement

This work was supported by ARIS (Agency of Research and Innovation of Slovenia) and the Ministry of Digital Transition of Slovenia, as part of the research project V2-2272.

References

[1] Alelyani, S. (2021). Detection and Evaluation of Machine Learning Bias. *Applied Sciences, 11*(14). https://doi.org/10.3390/app11146271.

[2] Angwin, A., Larson, J., Mattu, J. & Kirchner, L. (2016). Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks. *ProPublica*.

[3] *Artificial Intelligence Act*, European Parliament, 14 June 2023; current updated unofficial version January 2024.

[4] Berk, R., Heidari, H., Jabbari, S., Kearns, M & Roth, A. (2021) Fairness in Criminal Justice Risk Assessments: The State of the Art. *Sociological Methods & Research, 50*(1), 3-44. https://doi.org/10.1177/0049124118782533.

[5] Blanzeisky, W. & Cunningham, P. (2021). Algorithmic Factors Influencing Bias in Machine Learning. In: Kamp, M., et al. Machine Learning and Principles and Practice of Knowledge Discovery in Databases. ECML PKDD 2021. Communications in Computer and Information Science, 1524, Springer, Cham. https://doi.org/10.1007/978-3-030-93736-2_41.

[6] Cestnik, B., Kononenko, I. & Bratko, I. (1987). ASSISTANT 86: A knowledge-elicitation tool for sophisticated users: In: Bratko, I., Lavrač, N. (eds) *Progress in Machine Learning: Proc. of European Working Session on Learning EWSL 87*. Sigma Press, 1987, 31-45.

[7] Chakraborty, J., Majumder, J. & Menzies, T. (2021). Bias in Machine Learning Software: Why? How? What to do? arXiv: 2105.12195. https://doi.org/10.48550/arXiv.2105.12195.

[8] Corbett-Davies, S., Gaebler, J. D., Nilforoshan, H., Shroff, R. & Goel, S. (2018). The Measure and Mismeasure of Fairness. arXiv: 1808.00023. https://doi.org/10.48550/arXiv.1808.00023.

[9] Courtland, R. (2018). Bias detectives: the researchers striving to make algorithms fair. *Nature, 558*, 357-360. https://doi.org/10.1038/d41586-018-05469-3.

[10] Dastin, J. (11.10.2018). *Amazon scraps secret AI recruiting tool that showed bias against women.* Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

[11] Dressel, J. & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*(1). https://doi.org/10.1126/sciadv.aao5580.

[12] Flores, A. W., Bechtel, K. & Lowenkamp, C. T. (2016). False Positives, False Negatives, and False Analyses: A Rejoinder to "Machine Bias: There's Software Used Across the Country to Predict Future Criminals and It's Biased Against Blacks." *Federal Probation Journal, 80*(2).

[13] Gordon, D. F. & Desjardins, M. (1995). Evaluation and Selection of Biases in Machine Learning. *Machine Learning*, 20, 5-22. https://doi.org/10.1023/A:1022630017346.

[14] Hardt, M., Price, E. & Srebro, N. (2016). Equality of Opportunity in Supervised Learning. arXiv: 1610.02413. https://doi.org/10.48550/arXiv.1610.02413.

[15] Hellström, T., Dignum, V. & Bensch, S. (2020). Bias in Machine Learning – What is it Good for? arXiv: 2004.00686. https://doi.org/10.48550/arXiv.2004.00686.

[16] Holsinger, A. M., Lowenkamp, C. T., Latessa, E. J., Serin, R., Cohen, T. H., Robinson, C. R., Flores, A. W. & vanBenschoten, S. W. (2018). A Rejoinder to Dressel and Farid: New Study Finds Computer Algorithm Is More Accurate Than Humans at Predicting Arrest and as Good as a Group of 20 Lay Experts. *Federal Probation Journal*, 82(2), 51-56. https://doi.org/10.2139/ssrn.3271682.

[17] Hüllermeier, E., Fober, T. & Mernberger, M. (2013). Inductive Bias. *Encyclopedia of Systems Biology*. https://doi.org/10.1007/978-1-4419-9863-7_927.

[18] Kleinberg, J., Mullainathsn, S. & Raghavan, M. (2016). Inherent Trade-Offs in the Fair Determination of Risk Scores. arXiv: 1609.05807. https://doi.org/10.48550/arXiv.1609.05807.

[19] Levin, S. (8.9.2016). *A beauty contest was judged by AI and the robots didn't like dark skin*. The Guardian. https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people.

[20] Mehrabi, A., Morstatter, F., Saxena, N., Lerman, K. & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM computing surveys (CSUR), 54*(6), 1-35. https://doi.org/10.48550/arXiv.1908.09635.

[21] Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M. E. … Staab, S. (2020). Bias in data-driven artificial intelligence systems – An introductory survey. *WIREs Data Mining and Knowledge Discovery, 10*(3). https://doi.org/10.1002/widm.1356.

[22] Rudin, C. (2019). Stop explaining black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x.

[23] Spielkamp, M. (2017). Inspecting Algorithms for Bias. *MIT Technology Review*, July 2017.

[24] Staff, R. (7.12.2016). *New Zealand passport robot tells applicant of Asian descent to open eyes*. Reuters. https://www.reuters.com/article/us-newzealand-passport-error-idUSKBN13W0RL.

[25] Sun, O., Nasroui, O. & Shafto, P. (2020). Evolution and impact of bias in human and machine learning algorithms interaction. *PLoS ONE*, 15(18). https://doi.org/10.1371/journal.pone.0235502.

[26] UNESCO Recommendation on the Ethics of Artificial Intelligence, 2021. https://unesdoc.unesco.org/ark:/48223/pf0000381137.

[27] Yu, H., Shen, Z., Miao, C., Lesser, V. R. & Yang, Q. (2018). Building Ethics into Artificial Intelligence. arXiv: 1812.02953. https://doi.org/10.48550/arXiv.1812.02953.

Utilizing an Ensemble Machine Learning Framework for Handling Concept Drift in Spatiotemporal Data Streams Classification

Ature Angbera[*1,2], Huah Yong Chan[*1]
[1]Universiti Sains Malaysia, School of Computer Sciences, Pulau Pinang 11800, Malaysia
[2]Joseph Sarwuan Tarka University, Department of Computer Science, Makurdi, Nigeria
E-mail: angberaature@student.usm.my, hychan@usm.my

Keywords: spatiotemporal data stream, concept drift, ensemble learning, machine learning, spatiotemporal data classification

Received: May 18, 2023

The number of systems and devices broadcasting spatiotemporal data has recently significantly increased. Streaming data analytics provides the foundation of various spatiotemporal data services and functions. The non-stationary characteristics of these platforms and the constantly altering trends of the spatiotemporal data streams present concept drift issues for spatiotemporal data analytics. As a result, when concept drift occurs, it harms the model. The model's performance will eventually decline. The learning algorithms need the proper adaptive techniques to deal with concept drift on the spatiotemporal data streams with accurate predictions. This paper proposes an average weighted performance ensemble model (AWPEM). The AWPEM framework is for drift adaptation for spatiotemporal data stream classification. The framework is evaluated using real-world spatiotemporal data and compared to other state-of-the-art methods. The results show that the proposed framework outperforms other methods in terms of classification accuracy and robustness to concept drift. Further research will focus on enhancing the adaptability of the proposed framework to diverse and dynamic spatiotemporal data environments, exploring mechanisms for automated parameter tuning, investigating computational efficiency and scalability to large-scale spatiotemporal datasets.

Povzetek: Za večjo robustnost učnih sistemov je razvit ansambel WPEM, povprečno uteženi performančni ansambel za prilagoditev konceptualnim spremembam pri klasifikaciji spatiotemporalnih podatkovnih tokov.

1 Introduction

In recent years, it has become evident that the volume of data generated by technologies such as social media, sensor data, and other sources is rapidly increasing. Particularly, spatiotemporal data is streamed and vastly outpaces analysis tools' memory and processing power. There are currently 2.7 zeta bytes of data in the digital realm, which is growing daily [1]. The amount of data produced by systems like email and network monitoring [2], forecasting air quality grade [3], assessing credit risk [4], and analysing mobile users' behaviour changes [5] are unfit for disc storage because they are so huge. In light of this, streaming algorithms are made to process data as it comes in, online, and without storing a sizable amount of data in the main memory. As a result, real-time analytics on non-stationary data have recently caught researchers' attention. Spatiotemporal data streams are data collections that flow continuously and alter as they enter a system. According to [6], data streams can be enormous, timely ordered, changing quickly, and potentially endless in duration. Due to the periodic data changes in the streaming spatiotemporal data, the typical mining method is faced with the problem of concept drift [7]. The mining algorithm needs to be upgraded. Concept drift is a problematic issue in online learning since it significantly affects the consistency of streaming data classification [8]. If it goes undiscovered, concept drift can negatively impact the accuracy of predictions. We can handle distributional changes and maintain great accuracy of the prediction over time by employing concept drift detection models.

Assume that X and Y are the random variables representing the streaming observations and the labels that go with them. According to [9], concept drift is analogous to a change in the joint probability P(Y, X) at different time steps $t, z \in \{1, \dots, T\}$, that is:

$$P_t(Y, X) \neq P_z(Y, X)$$
$$\Leftrightarrow P_t(Y|X)P_t(X) \neq P_z(Y|X)P_z(X)$$

At time step t, $P_t(Y, X)$ is referred to as the active concept. We also distinguish between drift in actual and drift in virtual concepts. A shift in $P(X)$, that is: $P_t(X) \neq P_z(X)$, is referred to as virtual concept drift (X). Virtual concept drift is, therefore, unrelated to the target distribution and has no impact on the decision boundary [9]. Real concept drift, also known as concept shift, on the other hand, refers to a change in the conditional target distribution, that is: $P_t(Y|X) \neq P_z(Y|X)$. Real concept drift moves the decision border, which could affect predictions in the future [9]. In order to prevent severe

declines in prediction performance, it is imperative to spot changes in $P(Y|X)$ in a timely manner.

Sadly, concept drift does not exhibit a consistent trend in actual practice. Instead, we might see significant variations in concept drift's length and intensity. To achieve this, we distinguish various categories of concept drift as sudden or abrupt, gradual or reoccurring [1]. Figure 1 depicts the types or categories of concept drift in streaming data.

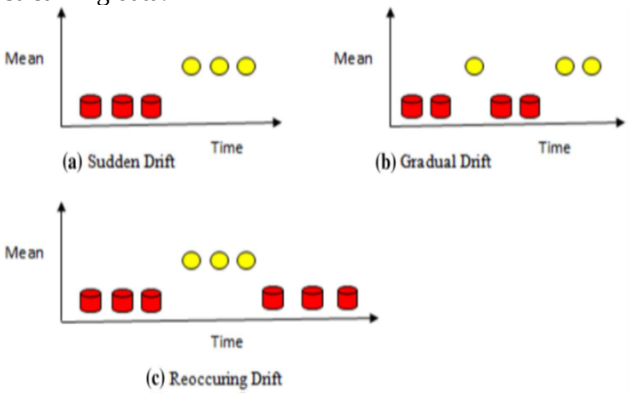


Figure 1: Types of concept drift [1]

This study investigates a precise and trustworthy concept drift approach in spatiotemporal data streams. The work suggests an average weighted performance ensemble model (AWPEM) for efficient concept drift detection and enhanced classification in spatiotemporal crime. The suggested model is an ensemble learning framework built on base learners and integrates two cutting-edge drift adaptation strategies. The two drift adaptation approaches employed are adaptive random forest (ARF) [10] and streaming random patches (SRP) [11]. As drift detectors, three popular drift detection approaches are used: adaptive windowing (ADWIN) [12], early drift detection method (EDDM) [13], and drift detection method (DDM) [14]. In order to build a strong spatiotemporal crime classification ensemble model with improved drift adaptability capability, the fundamental learners are combined and weighted based on their real-time success. The following are the study's ultimate goals:

i. Concept drift adaption techniques are examined.
ii. It suggests the AWPEM, a novel drift adaption technique, to solve the performance issues with the current concept drift techniques in spatiotemporal data streaming.
iii. Using two datasets, it assesses the proposed AWPEM framework for spatiotemporal crime predictions with concept drift detection and adaptation.

2 Related work

This part reviews previous efforts on spatiotemporal dataset (crime as a use case) analysis and provides an overview of cutting-edge techniques for concept drift detection and adaptation.

2.1 Spatiotemporal crime analysis

The interest in spatiotemporal modelling is a rising subject of open research [15]. The dynamic interaction between space and time allows for discovering relevant patterns through spatial-temporal data mining [16]. In [17] present a predictive strategy based on spatial analysis and auto-regressive models to effectively anticipate crime patterns in each place and automatically identify high-risk crime locations in urban areas. The algorithm's output is a spatiotemporal crime forecasting model, which includes several crime-prone areas and related crime predictors. Each of these predictor's functions as a predictive model to estimate the likelihood that crimes will occur to the extent it is allocated. In [18], the XGBoost machine learning algorithm was used to predict crime using actual crime and environmental data. The predictions were then interpreted using the SHAP method. With the aid of past crime statistics and environmental variables, XGBoost forecasts future crime. The prediction was then analyzed using SHAP, a machine learning interpreter, to show each variable's contributions from a global and local perspective. A cutting-edge deep learning technique called "Geographic-Semantic Ensemble Neural Network (GSEN)," which stacks a geographic prediction neural network with a semantic prediction neural network, was proposed [19]. The GSEN model combines the "Predictive Recurrent Neural Network (PredRNN), Graph Convolutional Predictive Recurrent Neural Network (GC-PredRNN)," and Ensemble Layer structures for spatiotemporal crime classifications in order to capture spatiotemporal dynamics from a variety of angles. The Fuzzy K-Nearest Neighbor algorithm and geospatial operations were utilised to create a crime prediction model in [20], and the study recommended integrating a Safe Route Travel app with a Crime Mapping system. Based on prior crime data, the model forecasts the type of crime a location is most likely to experience. In [21] suggested using deep inception-residual networks (DIRNet) to develop precise forecasts of theft-related crime based on data from non-emergency service requests (311 events). The outcomes demonstrated that the DIRNet outperforms alternative prediction models, averaging an F1 of 71% on average.

For spatiotemporal crime predictions, the strategies mentioned above produce excellent results. However, these are static machine-learning models created for offline learning. They are unable to adapt to real-time changes in spatiotemporal data. Due to this drawback, they are useless when used in real-world crime prediction systems.

2.2 Methods of concept drift

Concept drift problems are typically encountered in spatiotemporal streaming data analysis when data distribution changes over time due to the non-stationary environment. Concept drift problems often cause spatiotemporal crime prediction models to perform poorly, which has profound security implications. According to the rate at which the data distribution

changes, concept drifts are categorized into abrupt and gradual drifts [12]. An efficient crime prediction model should swiftly adjust to the discovered drifts to address concept drift and retain high prediction accuracy [5], [22].

i. Detection of Concept Drift

A well-known prototype result-based methodology called the drift detection method (DDM) establishes dual parameters, a "*warning level*" and a "*drift level*," to track changes in the prototype's *"error rate and standard deviation"* for drift detection [18]. The concept drift presence is noticed by a substantial rise in the prototype's overall error rate and DDM standard deviation. Although a learner will be modified if its results dramatically worsen, the drift detection method is easy to work with and can prevent pointless model modifications. DDM can detect abrupt drifts efficiently but frequently responds slowly to gradual drifts; this is because memory overflows are brought on by the necessity to retain a sizable number of data samples to reach the drift level of a long, gradual drift [23], [24].

The Early Drift Detection Method (EDDM) keeps track of the separation between two successive errors to identify gradual drift. When there is little space between two subsequent errors, gradual drift occurs [25]. The separation between two misclassification errors should grow as predictions get more accurate. The process for resizing a window is the same as with DDM. The main problem with EDDM is that at least 30 mistakes must be included in the calculation, making it difficult to utilize with unbalanced datasets [23]. The EDDM is an improvement of the DDM; hence the DDM could not detect gradual drift.

The Adaptive sliding window (ADWIN) algorithm separates the data stream into two sequential sub-windows within a variable-size sliding window and compares the means [26]. Drift is discovered when the difference between the means exceeds a threshold value determined using the Hoeffding bound [27]. All previous data samples from before the discovered drift point are discarded once the drift point is identified [23]. Due to the sliding window's capacity to be stretched to a large windowing size to notice long shifts, ADWIN successfully identifies progressive drifts. The mean value, however, is often not a good indicator of change.

ii. Concept Drift Adaptation:

After drift detection, a suitable drift adaptation strategy should be employed to handle the found drifts and maintain strong learning performance. Incremental learning and ensemble procedures are the two most common types of current drift adaptation strategies. The learning model is incrementally modified by studying each sample one at a time in chronological order. The Hoeffding tree (HT), which employs the Hoeffding constraint, is a particular kind of "decision tree (DT)" whereby data streams may incrementally adapt [23]. Instead of using a decision tree to choose the optimal schism, the hoeffding tree uses the "Hoeffding bound" to determine how many samples are needed to choose the split node. Now that its node has been updated, the HT can adapt to new samples. However, the HT lacks any processes for handling certain types of drift. A modernized version of the HT, the Hoeffding Anytime Tree (HATT), often referred to as the "Extremely Fast Decision Tree (EFDT)" [28], instead than waiting to find the best split in the HT, splits nodes when it reaches the confidence threshold. The EFDT can respond better to concept drifts than the hoeffding tree thanks to this division mechanism, albeit its efficiency could be enhanced [29].

Ensemble learning approaches have been suggested as a way to improve concept drift adaption and create reliable learners for data stream analytics. Block-based and online ensembles are two additional categories for ensemble approaches [30], [31]. Block-based ensembles partition the data streams into fixed-size blocks, after which each block is trained with a base learner. The base learners will be assessed and modified each time a new block is released. Although they typically take longer to react to abrupt drifts, block-based ensembles react to drifts gradually and properly. Deciding on the right block dimension to allow for a drift reaction speed and the learning trade-off between the base learners' results is another issue with block-based ensemble systems [31]. Three popular block-based ensembles are "Accuracy Updated Ensemble (AUE), Accuracy Weighted Ensemble (AWE), and Streaming Ensemble Algorithm (SEA)" [5], [32]. Online ensembles include different incremental learning models, such as HTs, to enhance learning performance. The adaptive random forest (ARF) approach, developed by Gomes et al. [10], uses HTs as base learners and ADWIN as each tree's drift detector. The drift detection mechanism replaces the underperforming base trees with new trees that better fit the new concept. Since the random forest is also an effective machine learning algorithm, ARF frequently outperforms many other approaches. ARF also includes a powerful resampling method and the flexibility to accommodate various drifts. For streaming data analytics, Gomes et al. [11] also put forth the Streaming Random Patches (SRP) innovative adaptive ensemble approach. To create predictions, SRP combines the online bagging method and random subspace. SRP and ARF use the same technology, but SRP uses a global subspace randomization strategy, while ARF uses a local subspace randomization approach. The more adaptable global subspace randomization strategy increases the diversity of base learners. SRP frequently has better prediction accuracy than ARF, although its execution time is often longer [29].

Although numerous concept drift adaptation techniques are now in use, their prediction accuracy and drift reaction time are performance limited. Due to their weak capacity to react to drift and low model complexity, incremental learning approaches frequently perform poorly. In contrast, block-based ensembles face significant difficulties in determining block size and drift reaction speed. Online ensembles like ARF and SRP

consistently outperform block-based ensembles and incremental learning; however, their randomization strategies also cause unstable learning models by adding more unpredictability to building their models. In order to increase drift adaptation performance, this research proposes an ensemble model that is stable and reliable.

3 Proposed model framework

Figure 2 depicts a high-level view of the proposed system for spatiotemporal crime prediction based on data stream analytics. The primary steps are listed below. First, generated streams of spatiotemporal data are pre-processed. Second, the critical learners for the first classification of crimes and drift adaptation are constructed using six drift adaptation techniques of concept drift: "ARF+DDM, ARF+EDDM, ARF+ADWIN, SRP+EDDM, SRP+DDM, and SRP+ADWIN." The ensemble approach is then created by merging the six base learners' probabilities of their prediction according to the specified AWPEM architecture. The ensemble model is employed, which can predict crime and adjust to concept drifts.

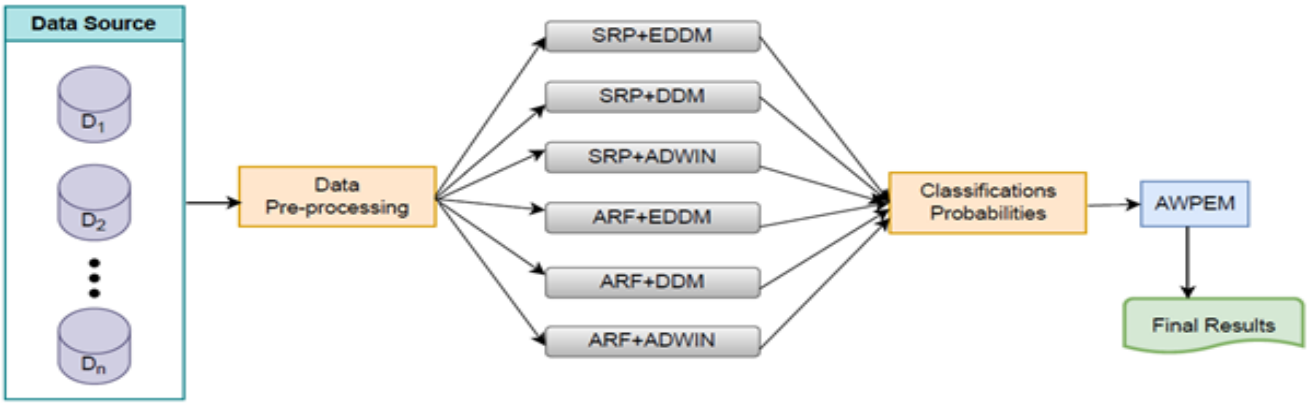


Figure 2: Proposed model framework

Three fundamental approaches for detecting drift (ADWIN, DDM, and EDDM) and two cutting-edge techniques for adjusting drift (ARF and SRP) are combined to create a robust ensemble model. Innovative methods for drift adaption, "ARF and SRP," have demonstrated superior performance to other drift adaptation techniques in experimental tests observed from literature [33], [34]. This motivated us to adapt and use them in our proposed model as our concept drift adaption techniques. Additionally, as was already mentioned, the concept drift detectors (ADWIN, DDM, and EDDM) employed in this work have the advantage of detecting both sudden and gradual drift. As a result, the suggested ensemble model can identify both types of drift effectively thanks to both drift detection techniques.

The average weighted performance ensemble model (AWPEM), a unique ensemble technique, is proposed in this study as a means of integrating the base learners for spatiotemporal data stream analytics. Many other ensemble techniques combine fixed weights, whereas AWPEM gives learners adjustable weights based on how they perform at the moment. Assuming that the goal attribute contains c various labels, $y \in 1, \ldots, c$ for each x data input and that the data stream $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$. It is possible to write the target class predicted by AWPEM as seen in Equation 1.

$$\hat{y} = \underset{i \in \{1,\ldots,c\}}{argmax} \frac{\sum_{j=1}^{k} w_j P_j(y = i | L_j, x)}{k} \quad (1)$$

$P_j(y = i|L_j, x)$ is the likelihood that the value of the class I will appear in the data sample x utilizing, j_{th} base learner, L_j. Each weight of the L_j is w_j and the suggested model base learners' number is k (i.e., $k = 6$).

After each data sample has been processed, the instantaneous error rate (see equation 2) is determined by dividing the sum of incorrectly classified labels by the overall number of samples analyzed.

$$ErrorRate = \frac{Total\ number\ of\ misclassified\ samples}{Total\ number\ of\ processed\ samples} \quad (2)$$

Each base learner's weight w_j, is obtained by taking the inverse of the real-time error rate. The w_j is denoted as given in Equation 3.

$$w_j = \frac{1}{ErrorRate + \epsilon} \quad (3)$$

Where ϵ (epsilon) represents a little constant that serves as a safeguard against the denominator falling to zero.

The idea of real-time error rates is employed in our proposed AWPEM model to create the base learners' weights for further dependable adaptation to data changes as opposed to the "mean square error rates" of the blocks of data utilized in the accuracy updated ensemble model. The ensemble model can account for the outcome of the base learners individually on a particular job due to the actual-time error rate on all compiled data.

The AWPEM provides several benefits since it can enhance the weights of base learners who perform at the top of their class while also considering other base learners. The suggested model's inverse-based weighting operation has improved AUE's weighting operation.

Additionally, the flexible weights generated with "real-time error rates" can be utilized to change the relevance of the base learners concerning their instantaneous results, in contrast to the fixed weights used in many existing ensemble techniques. These shifting weights guarantee that the present base learners who excel at the highest levels will be awarded higher weights.

4 Experiments and results

The Scikit-Multiflow [35] framework was expanded to implement the suggested framework using Python 3.9 on a computer with a core i5 processor.

4.1 Data pre-processing

We used two datasets for our model training and testing. These are crime datasets; the first dataset (Dataset1) has 319073 rows × 16 columns. The prediction class is the OFFENSE, whether it was violent or nonviolent. The second dataset (Dataset2) has 28303 rows × 20 columns. The prediction class here is whether there was an attack or no attack. In the hold-out evaluation, the initial model training took up 10% of the data, while the testing portion took up the remaining 90%. Before using the learning model for model training and updating, the learning model is tested via prequential validation, sometimes referred to as test-and-train validation. The five criteria, "accuracy, precision, recall, f1-score, and execution time," were utilized to assess the effectiveness of the suggested approach.

4.2 Results and discussion of the experiment

Prediction accuracies of the base learners (six) in the suggested model were lower than that of the proposed AWPEM on both datasets used in the experiments.

Figure 3 depicts the accuracies of the crime predictions of the base learning models of ARF+ADWIN and ARF+DDM on both datasets used in this study. Figures 3(a) and 3(b) show that the base learners' model accuracies were 86.33% and 98.43% on both datasets, respectively. Figure 3(c) and 3(d) prediction accuracies were 86.93% and 98.70% on both datasets, respectively.

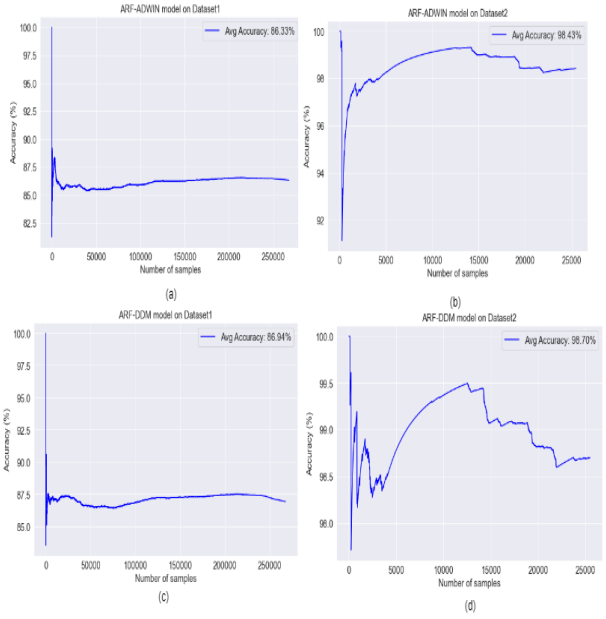


Figure 3: Accuracies of ARF+ADWIN and ARF+DDM, both datasets.

Figure 4 depicts the accuracies of the crime predictions of the base learning models of ARF+EDDM and SRP+ADWIN on both datasets used in this study. Figures 4(a) and 4(b) show that the base learners' model accuracies were 87.38% and 98.75% on both datasets, respectively. Figure 4(c) and 4(d) prediction accuracies were 80.56% and 98.56% on both datasets, respectively.

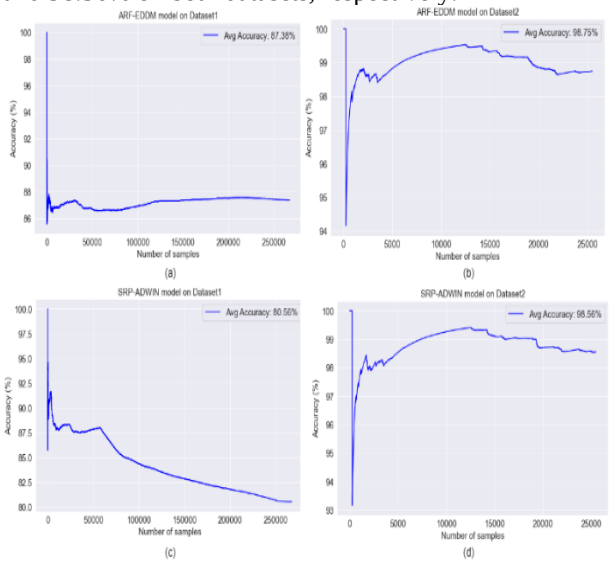


Figure 4: Accuracies of ARF+EDDM and SRP+ADWIN, both datasets

Figure 5 depicts the accuracies of the crime predictions of the base learning models of ARF+EDDM and SRP+ADWIN on both datasets used in this study. Figures 5(a) and 5(b) show that the base learners' model accuracies were 95.01% and 98.43% on both datasets, respectively. Figure 5(c) and 5(d) prediction accuracies were 95.00% and 98.64% on both datasets, respectively.

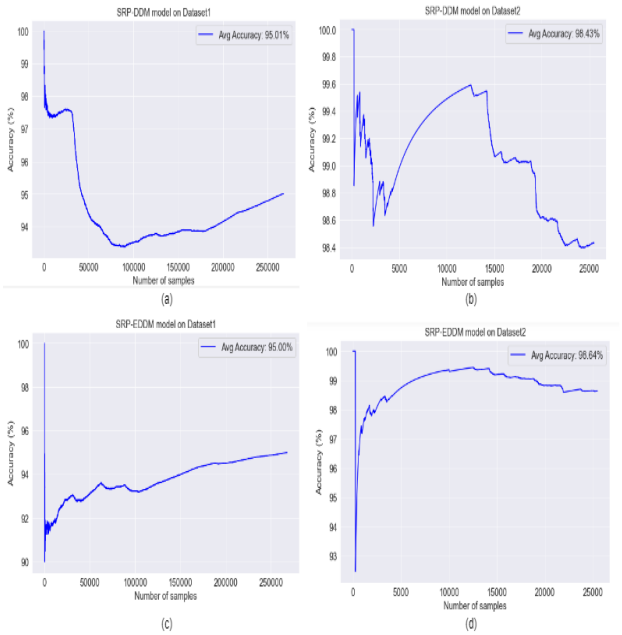


Figure 5: Accuracies of SRP+DDM and SRP+EDDM both datasets

Table 1: Comparison of the effectiveness of drift adaptation techniques on Dataset1

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Average time (ms)
ARF+ADWIN	86.33	76.58	86.80	81.37	6.07
ARF+DDM	86.93	77.17	88.08	82.26	11.87
ARF+EDDM	87.38	78.13	87.90	82.73	12.4
SRP+ADWIN	80.56	70.92	73.66	72.26	22.28
SRP+DDM	95.01	91.33	94.47	92.87	20.88
SRP+EDDM	95.00	91.84	93.77	92.80	22.18
HT	84.49	71.09	92.49	80.39	1.17
LB	92.48	86.29	92.87	89.46	29.07
Proposed AWPEM	98.45	97.87	97.63	97.75	35.08

Table 2: Comparison of the effectiveness of drift adaptation techniques on Dataset2

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	Average time (ms)
ARF+ADWIN	98.43	96.15	94.63	95.38	0.20
ARF+DDM	98.70	97.36	95.04	96.18	0.25
ARF+EDDM	98.75	96.71	95.98	96.34	0.23
SRP+ADWIN	98.56	96.52	95.02	95.76	1.02
SRP+DDM	98.43	96.34	94.45	95.39	1.33
SRP+EDDM	98.64	96.62	95.43	96.02	1.12
EFDT	91.02	69.35	85.54	76.60	0.41
HT	91.61	73.03	81.11	76.86	0.09
LB	97.79	92.76	94.54	93.64	0.02
Proposed AWPEM	99.24	98.54	97.01	97.77	4.20

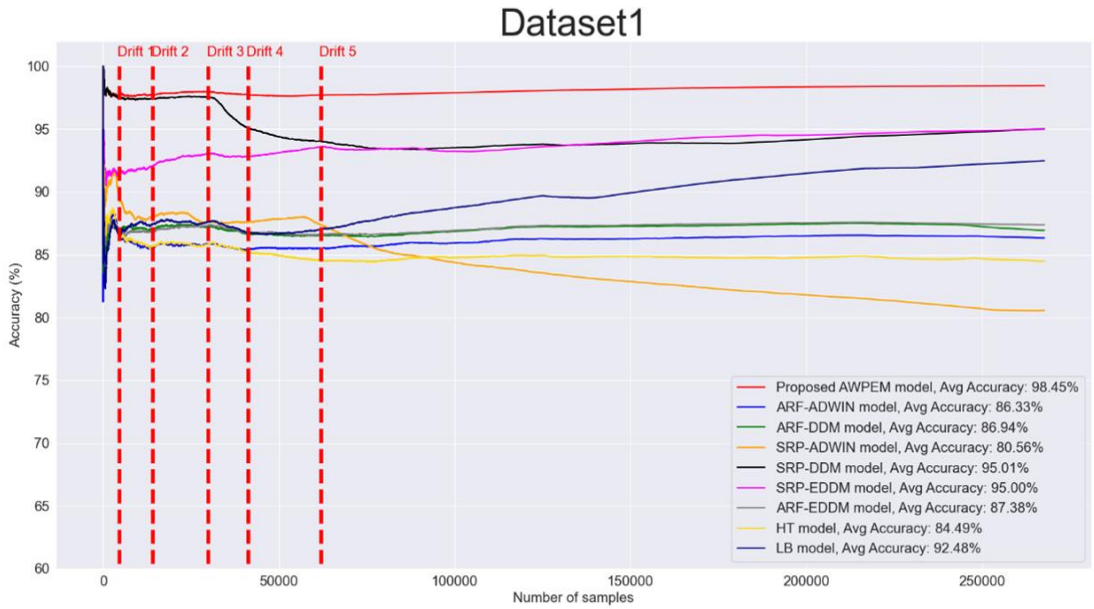


Figure 6: Accuracy comparison of drift adaptation methods on dataset1

The suggested AWPEM framework's performance is compared in Tables I and II to other cutting-edge drift adaptive methods presented in this paper, such as ARF, SRP, HT, EFDT, & LB. The suggested AWPEM technique performs better than all other models, as demonstrated in Tables I and II. On dataset1, from Figure 6, five drifts were seen early on; these drifts were noticed as a result of an increase in crime occurrences. Despite having different levels of adaptability, all the methods used could adjust to the drifts swiftly. However, our proposed AWPEM method swiftly adapted to the drifts and maintained a higher accuracy of 98.45%.

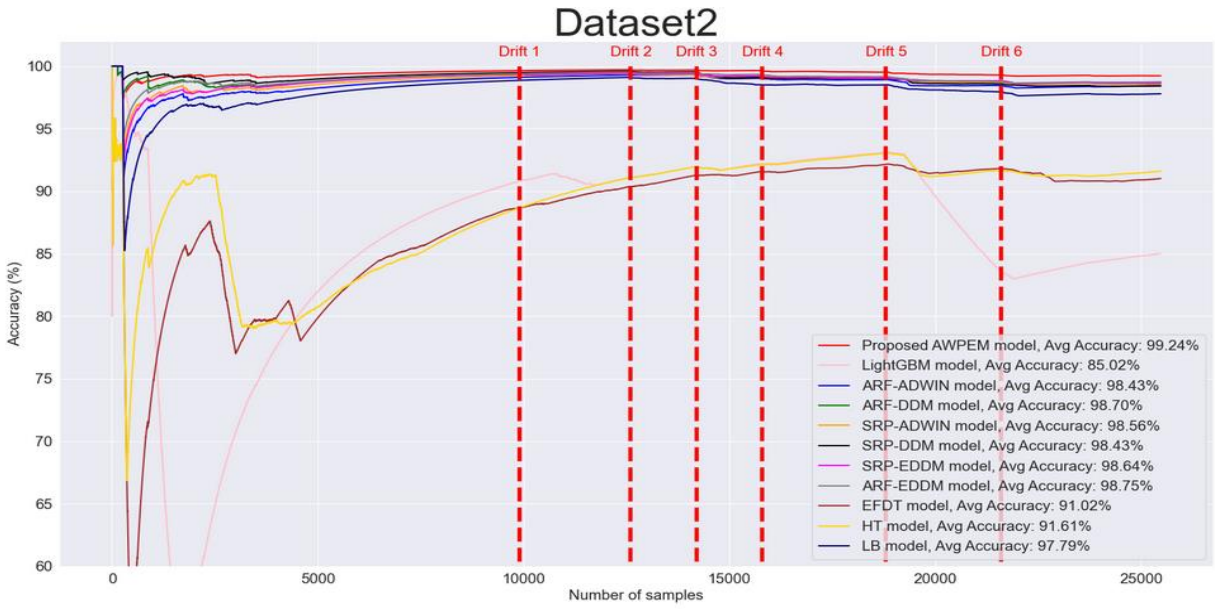


Figure 7: Comparison of drift adaption techniques' accuracy on dataset 2

There were six concept drifts in the tests for dataset2, as depicted in Figure 7. Both sudden (drifts 1, 3, and 5) and gradual drifts were seen in this series (drifts 2, 4, 6); these changes resulted from increased crime committed at various points. Once more, the suggested AWPEM approach quickly corrected for drifts and preserved a greater accuracy of 99.24 %. The justification for selecting the six base learners as base learners is further strengthened by their superior performance compared to other cutting-edge drift adaption techniques.

Figure 8 illustrates how the suggested AWPEM strategy takes longer to execute on datasets 1 and 2 than the other tested methods, but the mean time execution

remains reasonable for each occurrence. M1, M2, M3, M4, M5, and M6 in figure 8 stand for, respectively, "ARF+ADWIN, ARF+DDM, ARF+EDDM, SRP+ADWIN, SRP+DDM, and SRP+EDDM".

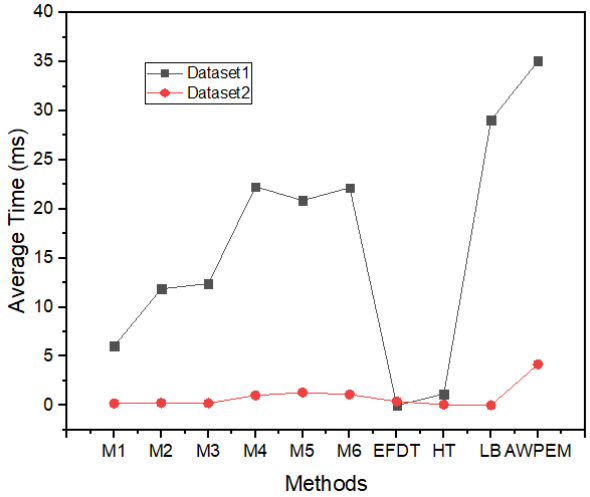


Figure 8: Average time comparison of the models on the two datasets.

Overall, while the average time taken by AWPEM may be higher compared to some other models, its outstanding performance across multiple metrics demonstrates its efficacy in addressing concept drift. Additionally, the difference in processing time between the two datasets highlights the importance of considering dataset characteristics and computational resources when selecting an appropriate drift adaptation technique.
Certainly, addressing the computational efficiency of the AWPEM model can be a promising avenue for future research. While AWPEM demonstrates exceptional performance in handling concept drift, its relatively longer average processing time, especially in Dataset1, suggests room for improvement in terms of computational efficiency. One potential direction for future studies could involve optimizing the algorithmic implementation of AWPEM to reduce its computational overhead without compromising its predictive accuracy. This optimization may involve refining the model architecture, streamlining the feature selection process, or implementing more efficient data processing techniques. Furthermore, exploring parallelization and distributed computing strategies could help accelerate the processing speed of AWPEM, making it more scalable and suitable for handling large-scale spatiotemporal datasets in real-time applications.

In terms of precision, recall, and f1-score, Figures 9 and 10 compare the suggested method to state-of-the-art techniques. The suggested model's precision shows that it can correctly recognize and predict positive data samples even when drift occurs. On the other hand, current methods consider the classification model's response to the input data at hand, resulting in a wrong prediction when concept drifts occur.

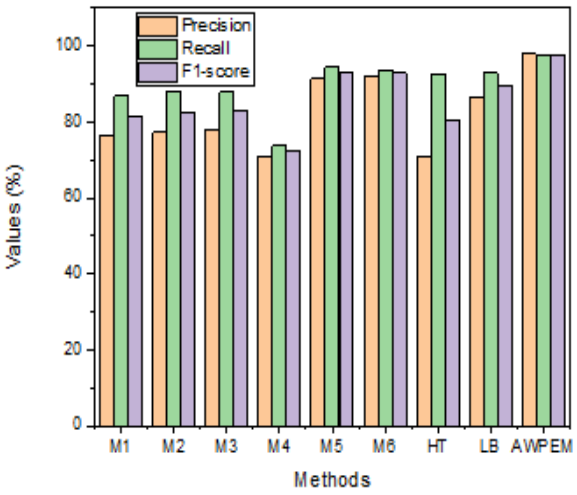


Figure 9: Performance comparison with precision, recall, and f1-score of the methods on dataset1

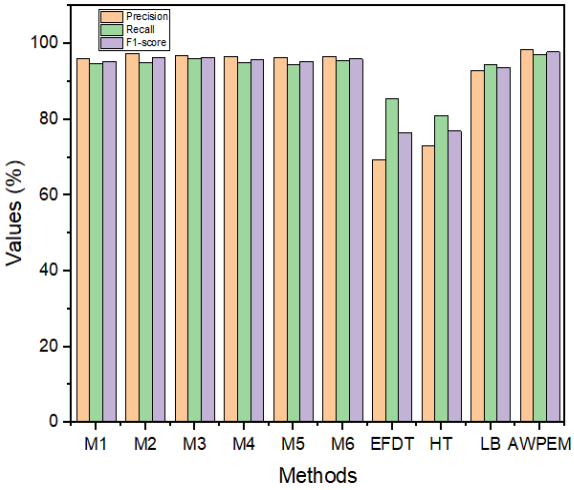


Figure 10: Performance comparison with precision, recall, and f1-score of the methods on dataset2

Recall demonstrates the capacity of the suggested model to generate fewer false negatives. For dataset1 and dataset2, the proposed model outperformed earlier techniques in the recall by 97.63% and 97.01%, respectively. When false negatives and false positives are critical, and the class distribution in the real-world dataset is uneven, the F1-score is calculated and compared. For datasets 1 and 2, the proposed model beat existing techniques by 97.75% and 97.77% f1-score.

5 Conclusion

This research offers the AWPEM framework for drift adaptive for spatiotemporal data streams classification, built on a collection of cutting-edge drift adaptation techniques. The proposed framework predicted crime in response to adaptation to the changes in data (concept drift) by attaining 98.45% and 99.24% accuracies on the two datasets. These accuracies' performance is significantly higher than other cutting-edge algorithms' accuracies based on the experimental performance of the two spatiotemporal data streams. To expand it, more drift

adaption strategies that offer better performance, diversity, and speed can be added to the proposed framework in subsequent studies. Also, it can be further optimized to reduce its high computational efficiency.

Data availability statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

[1] S. Priya and R. A. Uthra, "Comprehensive analysis for class imbalance data with concept drift using ensemble based classification," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 4943–4956, 2020, http://dx.doi.org/10.1007/s12652-020-01934-y.

[2] K. Pradeep Mohan Kumar, M. Saravanan, M. Thenmozhi, and K. Vijayakumar, "Intrusion detection system based on GA-fuzzy classifier for detecting malicious attacks," Concurrency and Computation: Practice and Experience, vol. 33, no. 3, pp. 5–10, 2019, http://dx.doi.org/10.1002/cpe.5242.

[3] W. Liu, H. Zhang, and Q. Liu, "An Air Quality Grade Forecasting Approach Based on Ensemble Learning," Proceedings - 2019 International Conference on Artificial Intelligence and Advanced Manufacturing, AIAM 2019, pp. 87–91, 2019, http://dx.doi.org/10.1109/aiam48774.2019.00024.

[4] Z. Li, Y. Tian, K. Li, F. Zhou, and W. Yang, "Reject inference in credit scoring using Semi-supervised Support Vector Machines," Expert Systems with Applications, vol. 74, pp. 105–114, 2017,. http://dx.doi.org/10.1016/j.eswa.2017.01.011

[5] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under Concept Drift: A Review," IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 12, pp. 2346–2363, 2019, http://dx.doi.org/10.1109/tkde.2018.2876857.

[6] J. Gama, "Knowledge Discovery from Data Streams (1st ed.)," in Chapman and Hall/CRC, New York, 2010, p. 255, http://dx.doi.org/10.1201/ebk1439826119-c1.

[7] C. Raab, M. Heusinger, and F. M. Schleif, "Reactive Soft Prototype Computing for Concept Drift Streams," Neurocomputing, vol. 416, pp. 340–351, 2020, http://dx.doi.org/10.1016/j.neucom.2019.11.111.

[8] L. Yang, Y.-M. Cheung, and Y. Yan Tang, "Adaptive chunk-based dynamic weighted majority for imbalanced data streams with concept drift," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 8, pp. 2764–2778, 2020, http://dx.doi.org/10.1109/tnnls.2019.2951814.

[9] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A Survey on Concept Drift Adaptation," *ACM Computing Surveys*, vol. 1, no. 1, p. 35, 2013, http://dx.doi.org/10.1145/2523813

[10] H. M. Gomes *et al.*, "Adaptive random forests for evolving data stream classification," *Machine Learning*, vol. 106, no. 9–10, pp. 1469–1495, 2017, http://dx.doi.org/10.1007/s10994-017-5642-8.

[11] H. M. Gomes, J. Read, and A. Bifet, "Streaming random patches for evolving data stream classification," *Proceedings - IEEE International Conference on Data Mining, ICDM*, vol. 2019-Novem, no. Icdm, pp. 240–249, 2019, http://dx.doi.org/10.1109/icdm.2019.00034.

[12] A. Bifet and R. Gavaldà, "Learning from time-changing data with adaptive windowing," in *Proceedings of the 7th SIAM International Conference on Data Mining*, 2007, pp. 443–448, http://dx.doi.org/10.1137/1.9781611972771.42.

[13] M. Baena-García, J. del Campo-Ávila, R. Fidalgo, A. Bifet, R. Gavaldà, and R. Morales-Bueno, "Early Drift Detection Method," *4th ECML PKDD International Workshop on Knowledge Discovery from Data Streams*, vol. 6, pp. 77–86, 2006, [Online]. Available: http://www.lsi.upc.edu/~abifet/EDDM.pdf.

[14] J. Gama, P. Medas, G. Castillo, and P. Rodrigues, "Learning with drift detection," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3171, pp. 286–295, 2004, http://dx.doi.org/10.1007/978-3-540-28645-5_29.

[15] R. de Medrano and J. L. Aznarte, "A spatio-temporal attention-based spot-forecasting framework for urban traffic prediction," *Applied Soft Computing Journal*, vol. 96, 2020, http://dx.doi.org/10.1016/j.asoc.2020.106615.

[16] A. Hamdi, K. Shaban, A. Erradi, A. Mohamed, S. K. Rumi, and F. D. Salim, *Spatiotemporal data mining: a survey on challenges and open problems*, no. 0123456789. Springer Netherlands, 2021, http://dx.doi.org/10.1007/s10462-021-09994-y.

[17] C. Catlett, E. Cesario, D. Talia, and A. Vinci, "Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments," *Pervasive and Mobile Computing*, vol. 53, pp. 62–74, 2019, http://dx.doi.org/10.1016/j.pmcj.2019.01.003.

[18] X. Zhang, L. Liu, M. Lan, G. Song, L. Xiao, and J. Chen, "Interpretable machine learning models for crime prediction," *Computers, Environment and Urban Systems*, vol. 94, no. January, p. 101789, 2022, http://dx.doi.org/10.1016/j.compenvurbsys.2022.101789.

[19] G. Jin, H. Sha, Y. Feng, Q. Cheng, and J. Huang, "GSEN: An ensemble deep learning benchmark model for urban hotspots spatiotemporal prediction," *Neurocomputing*, vol. 455, pp. 353–

367, 2021, http://dx.doi.org/10.1016/j.neucom.2021.05.008.

[20] S. Sharma, A. Uniyal, P. Srinivasan, and S. Chaudhari, "Fuzzy Based Geo-Spatial Crime Category Prediction for Crime Mapping and Safe Route Travel," in *2022 IEEE Region 10 Symposium (TENSYMP)*, 2022, pp. 1–6, http://dx.doi.org/10.1109/tensymp54529.2022.9864342.

[21] X. Ye, L. Duan, and Q. Peng, "Spatiotemporal Prediction of Theft Risk with Deep Inception-Residual Networks," *Smart Cities*, vol. 4, no. 1, pp. 204–216, 2021, http://dx.doi.org/10.3390/smartcities4010013.

[22] R. C. Samant, S. H. Patil, R. N. Sinha, and A. K. Kadam, "A Systematic Ensemble Approach for Concept Drift Detector Selection in Data Stream Classifiers," *International Journal of Engineering Trends and Technology*, vol. 70, no. 9, pp. 119–130, 2022, http://dx.doi.org/10.14445/22315381/ijett-v70i9p212.

[23] S. Wares, J. Isaacs, and E. Elyan, "Data stream mining: methods and challenges for handling concept drift," *SN Applied Sciences*, vol. 1, no. 11, pp. 1–19, 2019, http://dx.doi.org/10.1007/s42452-019-1433-0.

[24] P. Wang, N. Jin, W. L. Woo, J. R. Woodward, and D. Davies, "Noise tolerant drift detection method for data stream mining," *Information Sciences*, vol. 609, pp. 1318–1333, 2022, http://dx.doi.org/10.1016/j.ins.2022.07.065.

[25] C. H. Tan, V. C. S. Lee, and M. Salehi, "Information resources estimation for accurate distribution-based concept drift detection," *Information Processing and Management*, vol. 59, no. 3, p. 102911, 2022, http://dx.doi.org/10.1016/j.ipm.2022.102911.

[26] S. G. T. C. Santos, R. S. M. Barros, and P. M. Gonçalves, "A differential evolution based method for tuning concept drift detectors in data streams," *Information Sciences journal*, vol. 485, pp. 376–393, 2019, http://dx.doi.org/10.1016/j.ins.2019.02.031.

[27] J. Tanha, N. Samadi, Y. Abdi, and N. Razzaghi-asl, "CPSSDS : Conformal prediction for semi-supervised classification on data streams," *Information Sciences*, vol. 584, pp. 212–234, 2022, http://dx.doi.org/10.1016/j.ins.2021.10.068.

[28] C. Manapragada, G. I. Webb, and M. Salehi, "Extremely fast decision tree," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, no. August, pp. 1953–1962, 2018, http://dx.doi.org/10.1145/3219819.3220005.

[29] L. Yang, D. M. Manias, and A. Shami, "PWPAE: An Ensemble Framework for Concept Drift Adaptation in IoT Data Streams," *2021 IEEE Global Communications Conference, GLOBECOM 2021 - Proceedings*, pp. 1–6, 2021, http://dx.doi.org/10.1109/globecom46510.2021.9685338.

[30] H. Guo, S. Zhang, and W. Wang, "Selective ensemble-based online adaptive deep neural networks for streaming data with concept drift," *Neural Networks*, vol. 142, pp. 437–456, 2021, http://dx.doi.org/10.1016/j.neunet.2021.06.027.

[31] Y. Sun, Z. Wang, H. Liu, C. Du, and J. Yuan, "Online Ensemble Using Adaptive Windowing for Data Streams with Concept Drift," *International Journal of Distributed Sensor Networks*, vol. 2016, 2016, http://dx.doi.org/10.1155/2016/4218973.

[32] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, and M. Woźniak, "Ensemble learning for data stream analysis: A survey," *Information Fusion*, vol. 37, pp. 132–156, 2017, http://dx.doi.org/10.1016/j.inffus.2017.02.004.

[33] H. M. Gomes, J. Read, and A. Bifet, "Streaming Random Patches for Evolving Data Stream Classification," in *2019 IEEE International Conference on Data Mining (ICDM)*, 2019, pp. 240–249, http://dx.doi.org/10.1109/icdm.2019.00034.

[34] H. M. Gomes, J. P. Barddal, A. F. Enembreck, and A. Bifet, "A survey on ensemble learning for data stream classification," *ACM Computing Surveys*, 50(2), 2017, http://dx.doi.org/10.1145/3054925.

[35] J. Montiel, J. Read, A. Bifet, and T. Abdessalem, "Scikit-multiflow: A Multi-output Streaming Framework," *Journal of Machine Learning Research*, vol. 19, pp. 1–5, 2018, https://doi.org/10.48550/arXiv.1807.04662.

Attribute Induction-oriented Excavation and Generalization Analysis of Site Archaeological Data

Jianbing Zhang
Mechanical and Electrical Engineering Department, Zhumadian Vocational and Technical College, Zhumadian 463000, China
E-mail: zjb19780815@163.com

Keywords: Site archaeology; Data excavation; Generalisation analysis; K-means; Apriori; Attribute induction

Received: March 14, 2023

A significant amount of digital archaeological data has emerged as a result of the recent increase in archaeological activity, which is crucial for the preservation of cultural heritage. However, redundant and repetitive archaeological data information often leads to difficulties in management. The study first enhances the Apriori algorithm, which is based on determining the artefact attributes of site archaeological data by applying the boosting degree with difference, in order to increase the effectiveness of archaeological research. A K-means algorithm with adaptive selection of initial clustering centres was then proposed as a means to generalise the archaeological data for analysis. The outcomes revealed that the enhanced Apriori algorithm's maximum runtime was only 0.33 seconds and its minimum runtime was 0.1 seconds. Due to the low impact of noise points on the dataset Flame, the revised K-means algorithm's standard deviation is only 2.537, with the majority of the error values being clustered around zero. After combining the two methods, the classification accuracy of the digitised resources of the site is concentrated around 92%, with high classification accuracy and data generalisation processing ability, which improves the processing efficiency and provides a more reliable method reference for site archaeological research efficiency improvement.

Povzetek: Študija uporablja izboljšan Apriori algoritem za določanje atributov artefaktov in prilagojen K-means algoritem za generalizacijo arheoloških podatkov.

1 Introduction

Human society has left behind a vast amount of material materials in production and life, and with the changes of the times, these valuable heritages have become the witness of social development, providing an important basis for human beings to realise their cultural heritage [1,2]. In the process of scientific investigation and excavation, archaeology is an important driving force in historical research by systematically and completely revealing and collecting the relics buried deep in the ground and uncovering the historical and cultural values and artistic values they contain. Digital archaeology is continuing to progress, and digital resources are expanding rapidly due to the research and technological advances that are occurring so quickly. These digital resources are used throughout the archaeological excavation process, and not only have a huge amount of data, but also a great variety, which needs to be managed and utilised effectively. At the same time, the majority of archaeologists are still in the manual management and searching stage and are unable to keep up with the demand for the enormous amount of archaeological data [3,4]. How to integrate various archaeological data sources to boost archaeology's efficacy is a significant challenge for archaeologists in the information era. Data mining technology is capable of uncovering latent patterns, and its integration of database, machine learning and other multi-disciplinary techniques is one of the effective methods for conducting massive random data mining nowadays. Data generalisation for site archaeology can help archaeologists to view data at different levels in a customised degree of abstraction, greatly improving the efficiency of archaeological data analysis [5,6]. Applying data mining techniques to archaeological data generalisation analysis has a high degree of feasibility. To further increase the effectiveness of archaeological labor, the study suggests using the Apriori algorithm and K-means algorithm to generalize archaeological data by identifying the characteristics of site artifacts.

2 Literature review

Archaeologists have recently paid a lot of attention to the excavation of archaeological data from sites. Previtali and Valente developed a framework for sharing archaeological data in order to maximise the impact of archaeological data, which utilises digital technology for the collection of raw data and realises the processing of raw data through image processing and scan alignment, and the results show that the method enriches the raw assemblage data with a high degree of interoperability

[7].

Korf et al. developed a non-targeted data analysis method based on mass spectrometric detection of compounds in archaeological samples, which extracted more compounds through isotopic analysis of residual molecules in the samples, and showed that the method reduced the researcher's bias against extracting too few compounds and accelerated the overall analysis time [8]. Tronicke et al. developed a multi-scale, analysis and visualisation method to address archaeological practice and research. This method produces a discrete redundant wavelet transform (RWT) using a à trous algorithm and the cubic spline filter, then computes a multiscale decomposition of 2D data via a sequence of 1D convolutions. The results demonstrate that the approach simplifies the understanding of archaeological geophysical datasets and is computationally efficient [9]. Kansa and Kansa address the issue of digital data in archiving, modelling and other issues in archaeological practice, proposes a data management approach that incorporates multiple techniques and explores the creation of archaeological data for a wider range of needs, providing a new reference for the creation of archaeological information systems [10]. Otárola-Castillo et al. addresses archaeologists' frequent excavation and investigation of archaeological data through null hypothesis significance tests and probability distributions, proposing a method for accomplishing this using Bayesian statistics. It creates a Bayesian statistical framework for handling archaeological data and employs a collection of data for a particular site, and the findings show the method's viability in a real-world setting [11]

The improvement of clustering algorithms has some reference value for site archaeological data excavation processing. In response to the sensitivity of the initial cluster center selection of the K-means algorithm, Li's team proposed an improved hybrid particle swarm optimisation algorithm for clustering centers. The outcomes demonstrated that the approach had a high accuracy rate and enhanced convergence speed [12,13]. Using the algorithm's pseudo-code and experimental validation on a standard dataset (Lris), to be sure that the K-value selection had no impact on the convergence of the K-means algorithm, four K-value selection strategies were investigated. The results demonstrate that different K-values can be selected to lessen the impact for various clustering ranges. Huang et al. established the FPK-mediterranean algorithm to find the most convergent results based on the iterative K-medoids clustering algorithm for immobile points, constructed immobile point equations for each cluster, and solved the set of equations. The findings revealed that this algorithm's clustering efficiency and clustering quality were much higher than those of the conventional K-medoids algorithm [14]. Qi et al. developed the FPK-medoids algorithm [15]. An adaptive kernel fuzzy C-mean algorithm based on the cluster structure was then proposed. The results show that this method has high converging efficiency.

In summary, site archaeological data excavations have used algorithms to process archaeological raw data, and clustering algorithms have been used less in archaeological research. The overview of the research status is shown in Table 1. However, traditional clustering algorithms have significant limitations, and the improvement of clustering algorithms becomes important. Therefore, the study aims to further advance the process of site archaeology by improving traditional clustering algorithms and combining them with archaeological data excavation and classification.

Table 1: Summary of the current state of research

Methods or datasets	Advantage	Disadvantage	Literature
Archaeological Data Sharing Framework, Digital acquisition technology	Semi automated workflow, Open Data, Interoperability	High complexity	Previtali M[7]
Non target data mining	Accelerate overall analysis time	Suitable for mass spectrometry detection of compounds only	Korf Ad[8]
Cubic spline filter, à trous algorithm, RWT	High computational efficiency	Method construction is complex	Tronicke J[9]
Archaeological Information Systems	Integrating multiple technologies	Method construction is complex	Kansa E[10]
Bayesian statistics	Plain language, assuming quantification, clarity, and transparency	The inference efficiency and precision are low	Otárola-Castillo E R[11]
Improved hybrid particle swarm, K-means	Improved convergence speed and accuracy	Method construction is complex	Li Y[12]
K-means, Lris dataset	Improved the selection method for K values	Low clustering accuracy	Yuan C[13]
K-medoids, FPK-mediterranean algorithm	Significant improvement in clustering efficiency and	The method is relatively complex	Huang X[14]

	quality		
An adaptive kernel fuzzy C-mean algorithm	High clustering accuracy	Low clustering efficiency	Qi G[15]

3 Site archaeological data excavation and data generalisation based on attribute induction

3.1 Site archaeological data excavation based on attribute induction

Site excavation data is divided into three areas: heritage management information, heritage information and heritage excavation information. In heritage excavation information, mechanical means such as exploratory boreholes and machine drilling must be used in order to start the excavation, and the complete information can only be retained to the greatest extent possible if all types of data resources are recorded and preserved in detail. These data resources come from the entire process of heritage collection, including documentation, various types of photographs from the excavation, GIS data, exploratory hole stratification data, etc. There is historical and cultural information and physico-chemical information in the heritage information. The physical and chemical information is obtained and recorded with the help of measuring instruments, while 3D models, drawings and photographs provide a more accurate description of all the physical information that is difficult to record in writing, such as shape and colour, and historical and cultural information must be obtained through in-depth identification by archaeologists [16]. Heritage management information contains a number of important activities related to cultural objects and is an important basis for their excavation, archaeology and restoration. Based on the different types of data storage, the study has analysed the attributes of site archaeological data in terms of two types of structured and unstructured recorded data. Unstructured information is divided into two main categories: pictures and videos, both of which are important supplements to textual materials and provide a visual record of the shape, colour and other attribute information of the artefacts. Structured data primarily records textual information on the excavation process and other pertinent information on the management process [17], as well as information on the provenance, characteristics, management, and conservation of cultural items. Table 2 displays a summary of the fundamental details recorded for the items.

Table 2: Partial basic information of cultural relics records

Field	Data type	Length	Description
RegNumber	Integer	50	Total registration number
Onname	Text	No limit	Original name of the collection
Name	Text	No limit	Collection Name
Source	Text	20	Source
Shape	Text	1000	Morphological character
Quality	Text	50	Quality
Grade	Text	20	Collection level
InDate	Text	10	Date of storage
CurrentStatus	Text	No limit	Current situation

The artefacts themselves contain attribute information divided into many aspects of information such as artefact type, texture, period, testing unit, testing person, and the name of the site where they are located. The study summarises the structured information of the artefacts and draws a structured schematic diagram of the artefacts' attributes, as shown in Figure 1.

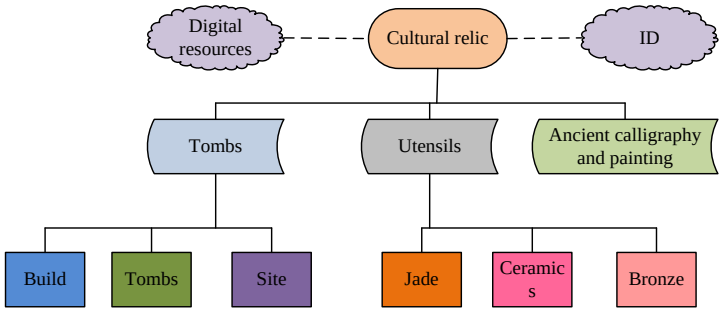


Figure 1: A structured schematic diagram of the attributes of cultural relics

The research is based on identifying the attributes of artefacts from site archaeological data and attribute generalisation through association rules. The association rule method has several uses in data mining and is capable of discovering useful associations in enormous amounts of data. The Apriori algorithm is straightforward and simple to use, making it one of the representative algorithms of association rules. Finding association rules that meet the minimal support and minimal confidence requirements for the frequent item set is the major objective [18]. The joining and trimming operations of the Apriori algorithm correspond to the identification of the frequent item set and the mining of association rules, respectively. In the join operation, the entire dataset is first traversed to obtain the frequent1 itemset L_1, and then the frequent $k-1$ itemset is joined with itself to obtain the candidate itemset C_k, to determine the final frequent k itemset L_k. In the pruning operation, since the candidate set C_k obtained under the concatenation contains elements that do not satisfy the condition, the frequent item set k and L_k must be found by traversing C_k again and deleting item sets smaller than $\min\_\sup$. After completing the above steps, the desired strong association rule is obtained by calculating the confidence of all items to remove items smaller than $\min\_conf$. Figure 2 depicts the Apriori algorithm's flow.

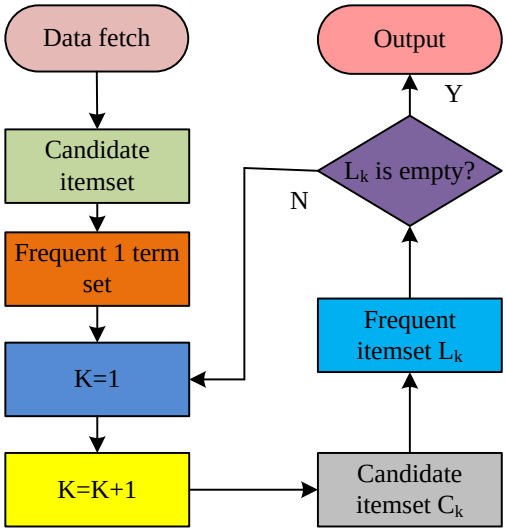


Figure 2: The basic process of Apriori algorithm

The traditional Apriori algorithm tends to obtain a large candidate set and is inefficient. Therefore, the study introduces a boosting degree to improve it. The boosting degree is used as the ratio of the probability of the presence of the posterior term Y to the probability of the occurrence of the posterior term X without X in the presence of the anterior term. Using the lift degree ensures that the association rules mined are positively correlated, as shown in equation (1).

$$Lift(X \Rightarrow Y) = \frac{Conf(X \Rightarrow Y)}{P(Y)} = \frac{P(X \cup Y)}{P(X)P(Y)} \quad (1)$$

In equation (1), the critical value of the lift is 1. $Lift(X \Rightarrow Y)$ is less than 1, the association rule shows negative correlation, when the two item sets exclude each other. There is no correlation with the association rule algorithm when $Lift(X \Rightarrow Y)$ is equal to 1. The association rule is now of scientific interest when $Lift(X \Rightarrow Y)$ is more than 1, which denotes that the association rule exhibits a strong positive connection. It is challenging to successfully discern between the

antecedent and consequent phrases of the rules and to compare various rules when using the lifting degree computation method. Therefore, the study designed a difference-based lifting degree to achieve association rule screening, as shown in equation (2).

$$Lift(X \Rightarrow Y) = \frac{Conf(X \Rightarrow Y) - Sup(Y)}{\max\{Conf(X \Rightarrow Y), Sup(Y)\}} \quad (2)$$

In equation (2), the determination of the influence of on X Y can be made by comparing the probability of the presence of the antecedent Y with the probability of the presence of Y itself when the antecedent X is present. $\max\{Conf(X \Rightarrow Y), Sup(Y)\}$ represents the normalisation factor, which makes the absolute value of $Lift(X \Rightarrow Y)$ less than 1. As $Lift(X \Rightarrow Y)$ is closer to 1, X has great impact on Y , and the closer it is to 0, the more useful the inverse rule of $(X \Rightarrow Y)$ is. The improved Apriori algorithm flowchart is shown in Figure 3.

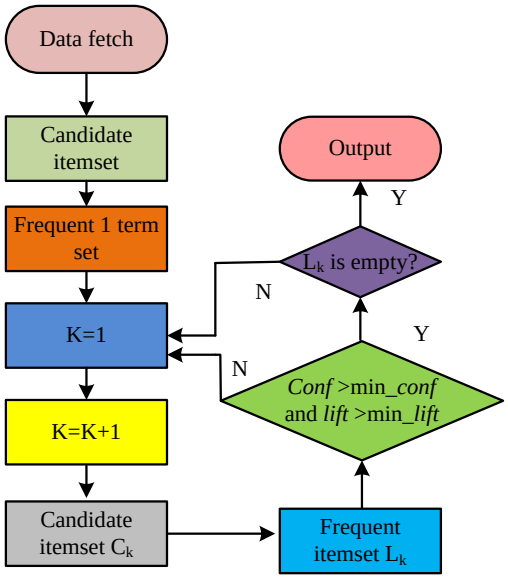


Figure 3: Improved apriori algorithm flowchart

3.2 Data generalisation analysis based on the K-means algorithm

The K-means algorithm is also used in the study to analyse site archaeological data in general. It is based on the introduction of a boosting degree improvement Apriori algorithm. To determine how similar the data are, the classic K-means approach largely analyses the distance between data points and the correlation coefficient between data indicators. In the process of effective clustering and ranking of data with high or low similarity, the correlation coefficient from the previous stage is also used as the basis for judging. In the assignment stage, the data with high similarity to the cluster centres are assigned to the same cluster [19-20]. Figure 4 depicts the fundamental flow of the conventional K-means algorithm.

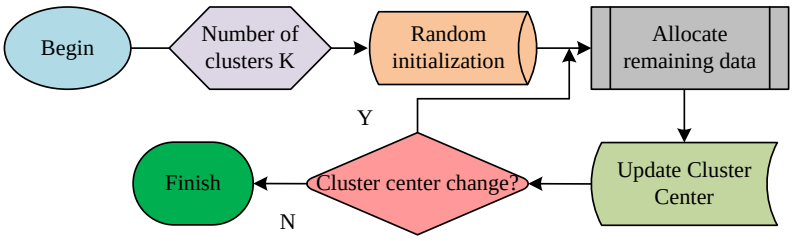


Figure 4: The basic process of traditional K-means algorithm

As can be apparent from equation (3), the Euclidean distance formula is used to determine how complex the data is.

$$d_{x,y} = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3)$$

In equation (3), x represents the data

($x=\{x_1,x_2,\cdots,x_n\}$) and y is also the data ($y=\{y_1,y_2,\cdots,y_n\}$). As shown in equation (4), the squared value of the error is typically employed in the traditional K-means algorithm as a standard function to evaluate the quality of the data clustering.

$$\begin{cases} C_k=\sum_{x_i\in C_k} x_i/|C_k| \\ E=\sum_{k=1}^{k}\sum_{x_i\in C_k}\|x_i-C_k\|^2 \end{cases} \quad (4)$$

In equation (4), C_k is the centre of clustering, which must ultimately be calculated to minimise the squared error. In this process, the final run of the algorithm ends when both the algorithm convergence and the minimum error squared conditions are satisfied, resulting in the desired classification result. According to equation (4), the traditional K-means algorithm has a relatively high probability of finding data values that fall into a local optimum solution [21-22]. Therefore, a method is devised that allows for the adaptive selection of initial clustering centres based on the distribution of different archaeological data features, in order to avoid local optima as much as possible. Figure 5 depicts the enhanced algorithm procedure.

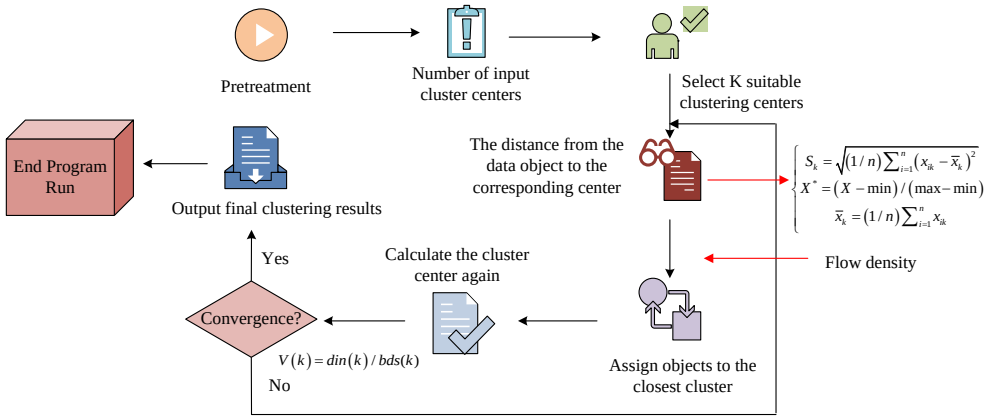


Figure 5: Basic process framework for improving K-means algorithm

The K clustering centres were first chosen by modified K-means algorithm for the input in Figure 4. Then, using the principle of proximity, it divides the data indicators into two groups: those with low similarity and those with high similarity. After a sample has been assigned, a resultant operation must be developed before the convergence of the data is judged and the output is finally obtained. The improved K-means algorithm runs in order to ensure that the gaps between the input data are small, and must be expanded pre-processing, as shown in equation (5).

$$\begin{cases} S_k=\sqrt{(1/n)\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2} \\ X^*=(X-\min)/(\max-\min) \\ \bar{x}_k=(1/n)\sum_{i=1}^{n}x_{ik} \end{cases} \quad (5)$$

X^* represents the greatest data value, the minimum data value, and the normalisation result in equation (5). Following the completion of the pre-processing, the entropy calculation is performed on the data having values between 0 and 1. The archaeological data set is set to $S=\{x_i\,|\,x_i\in R^m, i=1,2,3,\cdots,n\}$, which gives the convergence criterion and metric for clustering between data objects in the algorithm as revealed in equation (6).

$$\begin{cases} C=\sum_{j=1}^{k}\sum_{x_i\in c_j}|d(x_i,o_j)|^2 \\ d(x,y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \end{cases} \quad (6)$$

In equation (6), n stands for the overall number of data objects in the set, m for the spatial data's dimensional value, x,y for the attribute vector, and d for the Euclidean distance. $C_j=\{c_{j1},c_{j2},\cdots,c_{jn}\}^T$ represents the clustering centre. c_j represents the classification cluster, the average of all data points represented by o_j in the cluster c_j. The initial number of clusters is denoted by k. For the purpose of to determine the average distance of data objects in the data

set S and to calculate the density of a data object x_i in the set S, the study adds the idea of traffic density into the enhanced K-means algorithm, as indicated in equation (7).

$$\begin{cases} Mdts(S) = \left[2/n(n-1)\right] \cdot \sum_{i,j=1,i\neq j}^{n} d(x_i, x_j) \\ D(x_i) = \sum_{j=1}^{n} u\left(\rho \times Mdts(S) - d(x_i, x_j)\right) \end{cases} \quad (7)$$

In equation (7), ρ represents the radius factor, taking values in the interval [0.5, 2] and the function $u(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$. The mean density is then solved with the mean centres of the data points in the sub-clusters, as shown in equation (8).

$$\begin{cases} MD(S) = (1/n) \times \sum_{i=1}^{n} D(x_i) \\ \overline{x}_i = \left(1/|C_i|\right) \times \sum_{q=1}^{|C_i|} x_q, x_q \in C_i \end{cases} \quad (8)$$

In equation (8), S denotes the data set. $\overline{x}_i$ denotes the mean centres and $|C_i|$ denotes the number of objects clustered within the archaeological data in the cluster C_i. Equation (9) demonstrates that the value of the minimal distance between the mean centres of the data items in each cluster is the distance between clusters over the entire data set.

$$dbs(k) = \min_{1\leq i\leq k} \underset{1\leq j\leq k, i\neq j}{d} (\overline{x}_i, \overline{x}_j) \quad (9)$$

In equation (9), k denotes the quantity of data in the class cluster, and $\overline{x}_i, \overline{x}_j$ is the mean centres of the data points in the clusters C_i and C_j. The intra-cluster distance of the whole set S is the maximum of the intra-cluster distances of k, resulting in the calculation of the intra-cluster distance in the entire set as shown in equation (10).

$$din(k) = \max_{1\leq i\leq k} \left\{ \min_{1\leq j\leq |C_i|} \left\{ \frac{1}{|C_i|-1} \sum_{p=1,p\neq j}^{|C_i|} d(x_j, x_p) \right\} \right\}, x_j, x_p \in C_i \quad (10)$$

In equation (10), $din(k)$ refers to the intra-cluster distance. the gap between data objects in the set that belong to various clusters gets wider the bigger the resulting inter-cluster distance. In order to comply with the above requirements, the study further introduces the clustering effect determination function as shown in equation (11).

$$V(k) = din(k) / bds(k) \quad (11)$$

In equation (11), $V(k)$ is the inter-cluster clustering of the set. The degree of similarity between the data objects in data set S is at its highest when the discriminant function has a minimal value, which implies that the differences between the data items in various clusters also have a maximum value k. In these circumstances, it is possible to choose the cluster value that will have the best clustering effect by choosing an appropriate value as the minimum value of the discriminant function $V(k)$.

4 Effectiveness of site archaeological data excavation and generalisation analysis

The work makes use of association rules to improve the K-means algorithm by generalising qualities to site archaeological data. This uses the Apriori algorithm, which is based on the lifting degree of differences, to filter association rules. First, it is confirmed that the revised Apriori algorithm is valid. This part of the experiment was carried out on different datasets. Three datasets were chosen from the University of California, Irvine (UCI) database for the study: The Agaricus-lepiota (poisonous mushroom) dataset, the groceries dataset, and Voting Records of the United States Congress in 1984. The Agaricus-lepiota dataset contains a total of 8124 samples with 22 different attributes.The dataset describes different sample characteristics of mushrooms, such as color, odor, smoothness, etc.; as well as the sample's categorical labels, edible or poisonous.The Groceries dataset records a daily transaction record of a German supermarket, containing 9835 consumption records, 169 different items, with each purchase transaction corresponding to a list of items purchased by the purchaser.The 1984 U.S. Congressional Voting Records record the results of legislators' votes on 16 different policy issues according to Republican versus Democratic parties, and the dataset contains 435 records.

Table 3 shows that the Institute's revised Apriori algorithm does, however, include lift as a judgement criterion in terms of algorithm running time. The frequent item set obtained by this algorithm only includes the frequent item set of A, which accelerates the algorithm's processing speed. At the same time, the method is more

efficient than the traditional Apriori algorithm in terms of execution as the posterior term only contains the target term in the process of rule formation and no longer requires screening of redundant rules. In terms of the number of association rules, the traditional Apriori algorithm outputs all the rules in the same data set, while the improved Apriori algorithm obtains all the association rules whose back term is A. It has been discovered that the revised Apriori algorithm can eliminate the duplicate rules, resulting in a set of rules that are all tightly connected and positively correlated, increasing the method's effectiveness even more.

Table 3: Comparison of results between apriori algorithm and improved apriori algorithm in three datasets

Dataset name	Support	Confidence level	Evaluating indicator	Number of association rules	Run time/s
Groceries	0.2	0.02	Sup-conf-lift	59	2.83
			sup-conf	341	2.95
Agaricus-lepiota	0.9	0.4	Sup-conf-lift	16	1.76
			sup-conf	718	1.78
Integrated	0.9	0.3	Sup-conf-lift	286	0.36
			sup-conf	2978	0.37

Based on the data in Table 3, further comparison experiments were done, setting the boosting degree to 1 and the minimum support degree to 0.02. Table 4 displays the number of rules found by the two algorithms at various minimum levels of confidence. Table 4 demonstrates an inverse link between the minimal confidence of the Apriori algorithm before and after the enhancement and the number of rules. At a minimum support of 0.1, the traditional Apriori algorithm obtains 3556 rules, but the required target rules are only 583, indicating the existence of 2973 redundant rules. The number of rules for the original Apriori algorithm and the modified Apriori method are 1213 and 270, respectively, at a minimal confidence level of 1, and there are 943 duplicate rules. In this case, applying the traditional Apriori algorithm to attribute summarisation for archaeological data excavation would require a significant amount of time to sift through the redundant rules. In contrast, the study's improved Apriori algorithm is able to cope with the generation of unnecessary rules in order to provide more objective association rules.

Table 4: Number of rules obtained by two algorithms under different minimum confidence levels

Minimum confidence	Improved Apriori algorithm	Apriori algorithm
0.1	583	3556
0.2	583	3556
0.3	583	3556
0.4	583	3556
0.5	547	3378
0.6	369	2203
0.7	314	1958
0.8	297	1732
0.9	288	1520
1.0	270	1213

To further validate the improved Apriori method, the FP-growth algorithm and Node-list Pre-order Size Fuzzy Frequent (NPSFF) Algorithm Based on Fuzzy Association Rule Mining [23] was included to the experiment for comparison. Figure 6 displays the running times of FP-growth algorithm, NPSFF algorithm and before and after the improvement of Apriori algorithm. Figure 6 illustrates how the overall runtime of the four techniques decreased as the confidence level rose. The highest value of the traditional Apriori algorithm is close to 1.0 s, the highest value of the FP-growth algorithm is close to 0.75 s, the highest value of the NPSFF algorithm is close to 0.55 s, and the lowest values of both are 0.22 s, 0.25 s and 0.23 s, respectively, while the highest value of the runtime obtained by the improved Apriori algorithm is only 0.33 s and the lowest value is 0.1 s, which is lower than the other three methods. It can also be found that the improved Apriori algorithm is lower than the other three

methods in the whole process of confidence increase, and the execution efficiency has been significantly improved.

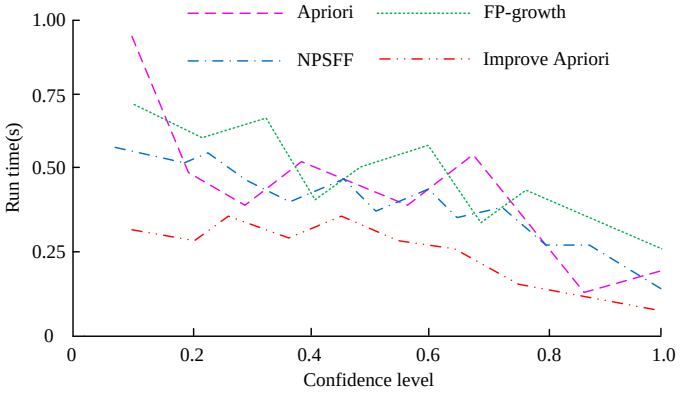
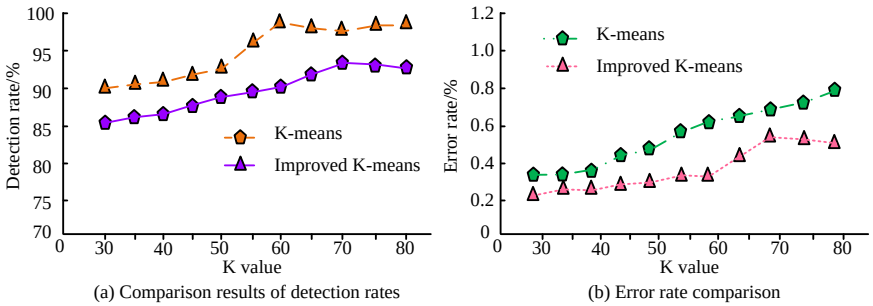


Figure 6: Comparison of running time of FP-growth, apriori algorithm and improved priori algorithm

The study subsequently evaluated the enhanced K-means algorithm after completing the performance validation of the revised Apriori algorithm. Figure 7 displays the findings of the study, which first contrasted it with the conventional K-means algorithm and assessed it using the metrics of data detection rate and error rate. Figure 7(a) and (b) display the outcomes. It has been discovered that as the value of K increases, the detection rate and error rate of the K-means algorithm before and after the improvement show varying degrees of increase. Figure 7(a) shows that the modified K-means algorithm greatly improves upon the classic K-means clustering algorithm in terms of data detection rate. With a detection rate of 98.21% at K=60, the upgraded K-means algorithm's detection rate starts to stabilise. The traditional, enhanced K-means algorithm had a 93.92% detection rate at K=70. Figure 7(b) demonstrates that the modified algorithm has reduced false alarm rate than the conventional algorithm. The enhanced algorithm's error rate at K=70 was 0.402%. The classic algorithm had an error rate of 0.623%. The comparison demonstrates that the optimised research approach can achieve the global ideal answer by having a high detection rate and a low error rate.



(a) Comparison results of detection rates (b) Error rate comparison

Figure 7: Comparison of detection and error rates of K-means algorithm before and after improvement

Figure 8 illustrates how the study contrasts the error of the upgraded K-means algorithm's clustering results with those of the conventional K-means method. The error histogram of the conventional K-means algorithm is shown in Figure 8(a), and the results of the upgraded K-means algorithm's error histogram are shown in Figure 8(b). While there are significant swings in the error range in Figure 8(a), the standard deviation of the K-means method is 4.498. The revised K-means algorithm's standard deviation in Figure 8(b) is 2.537, and a large portion of the error is clustered around 0, with a reduced range of error fluctuation. The upgraded K-means algorithm provides greater classification accuracy and stability as compared to the pre-improvement.

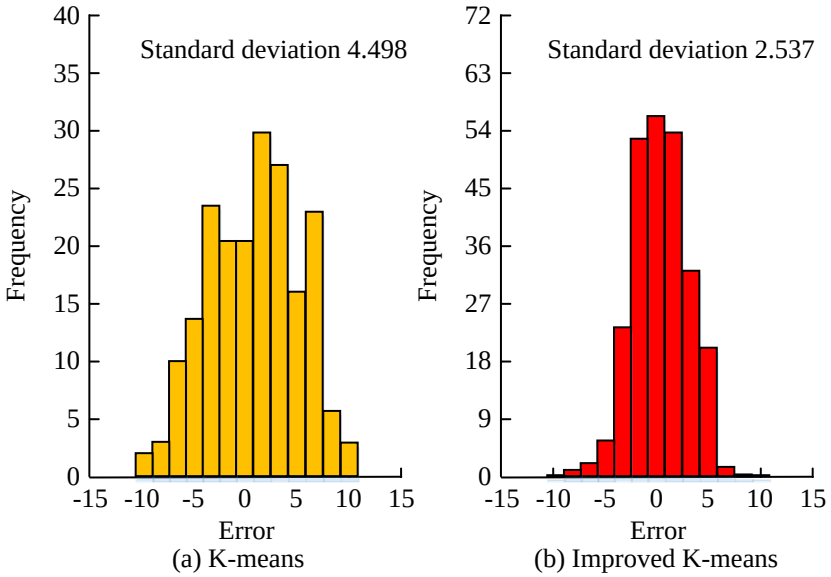


Figure 8: Histograms of prediction error distribution for different algorithms

The study included the Mean Shift clustering method to compare with the enhanced K-means algorithm in order to further demonstrate its superior performance. The dataset used is Flame, which is an artificial dataset and has clusters with ambiguous boundaries, which can verify the effectiveness of the algorithm for noisy data processing. The Flame dataset is a forest fire detection dataset based on aerial images made public by Northern Arizona University and others, and contains 177 different categories. Figure 9 displays the outcomes of the three techniques on the dataset Flame. Figures 9(a) and (b) show that the Mean Shift clustering algorithm and the K-means algorithm both provide outputs that contain a significant amount of noisy and ambiguous data points. And there is a more complex relationship between the real clusters, both fail to accurately identify the basic shape of the clusters, and the clustering effect is poor. In Figure 9(c), the enhanced K-means algorithm accomplishes accurate clustering of this dataset with a better clustering effect and better performance while being minimally impacted by the noise points.

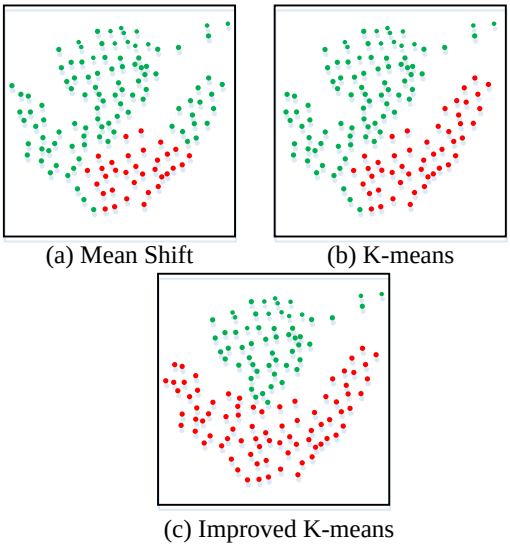


Figure 9: The clustering performance of three algorithms in the artificial dataset flame

The current advanced clustering models Attribute Spectral Clustering (ASC) [24] and Fuzzy decision tree-based clustering algorithm (FDTC) [25] are selected, and Xie Beni (XB) and Davies-Bouldin Index (Davies-Bouldin Index, DB) indicators for evaluation, and the experimental results are shown in Figure 10. As

seen in Figure 10(a), the DB value of the improved K-means algorithm shows a decreasing trend with the increase of the number of iterations and is at the lowest level of 0.07. As seen in Figure 10(b), the XB value of the different clustering methods shows a decreasing trend, and the improved K-means algorithm shows a more obvious advantage of taking the value of the XB, and the clusters of the clustering results have the smallest intraclass distances and the largest interclass distances.

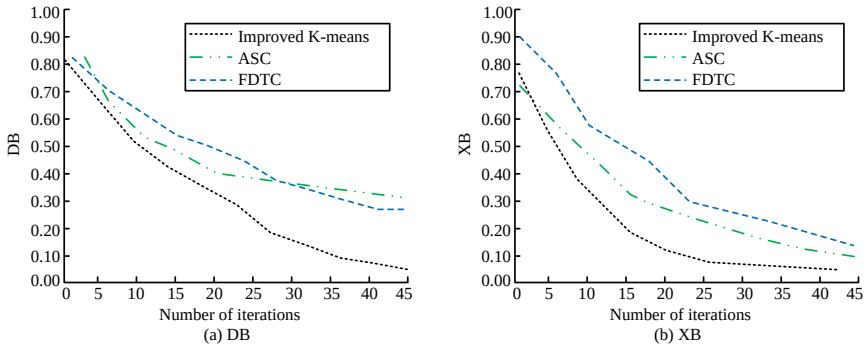


Figure 10: Comparison of statistical results of evaluation indicators of different clustering models

In order to evaluate the scalability and noise tolerance of different clustering methods, the study selects the distributed processing ability, clustering stability and noise point misclassification rate indexes for evaluation, and the experimental results are shown in Figure 11. From Fig. 11 (a), it can be seen that the improved K-means algorithm has the optimal distributed processing ability, and the processing ability evaluation value rises the fastest with the increase of iteration number. As seen in Figure 11(b) and (c), the clustering stability and noise point misclassification rate of the improved K-means algorithm perform optimally. The highest value of clustering stability reaches 92.56%, and the clustering results still have high consistency in the presence of noise. The noise point misclassification rate of the improved K-means algorithm converges to the lowest value of 19.85%, and the lower noise point misclassification rate improves the accuracy of the algorithm.

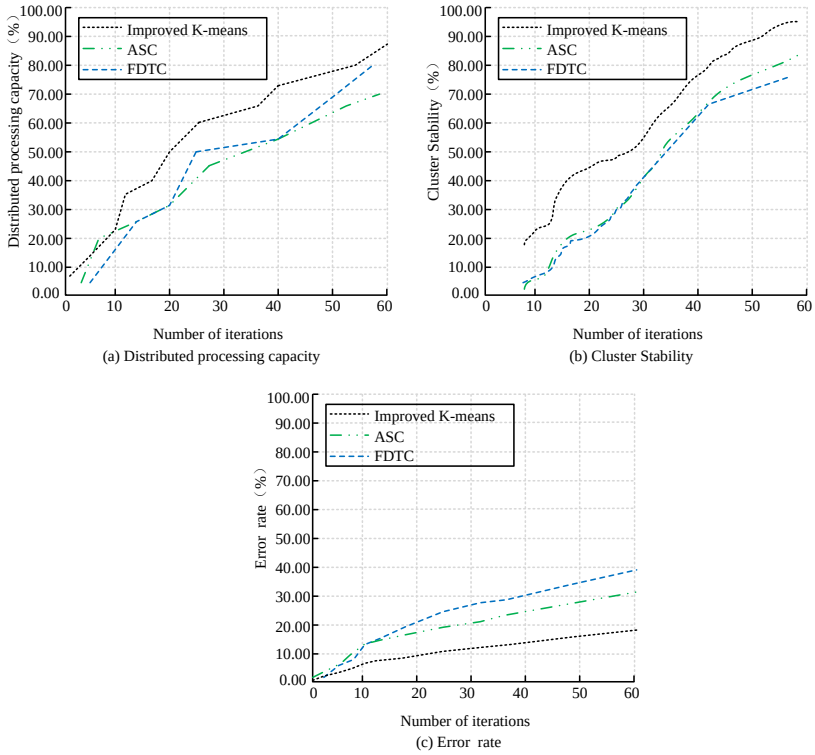
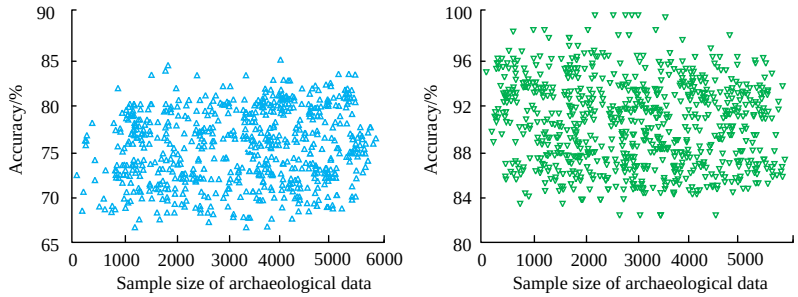


Figure 11: Scalability and noise tolerance of different clustering methods

The study assessed the enhanced Apriori algorithm and the enhanced K-means algorithm and showed that both performed better than expected. They are now jointly applied in practice to validate the results of site archaeological data generalisation analysis. The study selected the digitized resources of Liangzhu Culture Site, one of the southeast Chinese Neolithic culture, Hongshan Site, a noble burial site of the Yue State in Wuxi City, Jiangsu Province, and Sanxingdui Site, a site of the Late Neolithic to Bronze Age in Guanghan City, Sichuan Province, as experimental objects. The heritage data processing of this site was done using the combined technique, and the average results of the data processing for the three different sites are shown in Figure 12. The results of the overall classification accuracy of archaeological data is represented by Figures 12(a) and (b) before and after the combined application of the two methods, respectively. As can be seen from Figure 12, before the joint application of the methods, the classification accuracy of the archaeological data was overwhelmingly below 85% and concentrated at around 75%. After the joint application, the classification accuracy was mostly over 84% and concentrated around 92%, which is a significant increase compared to the pre-application period, proving the effectiveness of the method.



(a) Before joint application

Figure 12: Accuracy of archaeological data classification before and after joint application

5 Discussion

The integration of archaeological digital resources can help researchers access, analyze and compare archaeological resources more conveniently, and understand and excavate the development patterns of ancient civilizations, cultural changes and the laws of human activities. The traditional manual management, electronic documents, shared files and other methods have gradually failed to meet the needs of archaeological work, and the integration of archaeological digital resources needs to adapt to the development of the times and gradually change in the direction of informationization and intelligence. The process of data mining involves a series of steps from data collection to visualization, aiming to accurately search for effective information from massive information. Data mining technology can be used for the description and prediction of target data sets, classification and clustering of data, and identification of abnormal data. Therefore, in the face of massive archaeological data information, the use of data mining can be used to realize the classification and clustering of data, according to the characteristics of cultural relics and sites, mining the connection between different cultural relics and sites, and discovering the trend and cyclical changes of cultural evolution.

In this regard, the study takes data mining as the technical core, adopts the difference enhancement degree improvement Apriori algorithm, and utilizes this method to complete the mining of archaeological data. Meanwhile, the K-means algorithm with adaptive selection of initial clustering centers is designed to further realize the generalization analysis of archaeological data. The improved Apriori algorithm and the improved K-means algorithm perform better in terms of computational efficiency, clustering accuracy, clustering effect, as well as scalability and generalization. The clustering accuracy of the research-designed method is improved by 7.3 percentage points compared to the adaptive kernel fuzzy C-mean clustering algorithm studied by Qi et al [15]. Huang et al [14] designed K-Medoids clustering algorithm based on the immobile point iteration fluctuates in the range of values of the normalized mutual information metrics between 0.7 and 0.8, and the similarity level between clustered labeled clusters is average. In contrast, the research-designed method has reached the optimal value level on DB and XB metrics, and the interclass and intracluster distance status of the clustered clusters has reached the optimal status, and the advantages of the research's improved clustering method are obvious. In conclusion, the study synthesizes the advantages of existing research shown in Table 1 and considers the necessity of information technology in processing large-scale archaeological data. At the same time, the study embodies the research idea of integrating multiple techniques with reference to the current research status quo, and introduces the ideas of difference enhancement degree, adaptive selection, and flow density to improve the traditional Apriori algorithm and K-means algorithm. Compared with the K-means algorithm that introduces intelligent optimization

algorithm, the model complexity is reduced to a certain extent, the model computation is reduced, and the model complexity and computational efficiency are improved.

Comprehensively, the improved Apriori algorithm and improved K-means algorithm proposed by the study perform better in archaeological data mining and greatly improve the management of archaeological data. Based on the analysis and clustering results of data mining, archaeologists can discover more hidden patterns and correlations of cultural relics and sites, providing deeper insights into the study of historical civilizations and promoting the development of the field of archaeology. Considering the development needs of the archaeological industry, on the basis of the research results, future research can introduce more informatization and automation technologies to realize the intelligence of archaeological excavation work, such as the use of computer vision technology to realize the identification and detection of cultural relics, the construction of knowledge maps of cultural relics and sites, and the prediction and extrapolation of cultural relics' historical periods. The development of the field of archaeology will be promoted through in-depth cross-disciplinary cooperation.

6 Conclusion

The large-scale application of digital archaeological data has facilitated the further development of archaeological research. The study summarises the corresponding types of artefact attribute induction for the problem of extensive and complex archaeological data at sites, and applies the improved Apriori algorithm with the K-means algorithm to archaeological data management. This result show that with a minimum support of 0.1, the traditional Apriori algorithm obtained 3556 rules, but the required target rules were only 583, proving that the improved Apriori algorithm can cope with the generation of unnecessary rules. As the confidence level increases, the traditional Apriori algorithm approaches a maximum of 0.1 s, the FP-growth algorithm approaches a maximum of 0.75 s, and the improved Apriori algorithm achieves a maximum runtime of only 0.33 s. In the performance test of the improved K-means algorithm, the data detection rate of the algorithm starts to stabilise when K=60, at which point the detection rate is 98.21%. When K=70, the error rate of the improved K-means algorithm was 0.402%. The error rate for the conventional K-means algorithm was 0.623%. In the error comparison, the standard deviation of the K-means algorithm was 4.498 and that of the improved K-means algorithm was 2.537. With the joint application of the two improved methods, the classification accuracy of archaeological data was mostly over 84%, concentrated around 92%, which greatly improved the management of archaeological data. However, when the proposed method was validated, the study only analysed one selected excavated Neolithic culture site, which may have some errors, and therefore needs to be further optimised by extending the actual validation scope.

Reference

[1] C. Ward, “Excavating the archive/archiving the excavation: Archival processes and contexts in archaeology,” Advances in Archaeological Practice, vol. 10, no. 2, pp. 160-176, 2022. https://doi.org/10.1017/aap.2022.1

[2] I. Sipiran, A. Mendoza, A. Apaza, and C. Lopez, “Data-driven restoration of digital archaeological pottery with point cloud analysis,” International Journal of Computer Vision, vol. 130, no. 9, pp. 2149-2165, 2022. https://doi.org/10.1007/s11263-022-01637-1

[3] S. G. Ortman, and J. H. Altschul, “What north american archaeology needs to take advantage of the digital data revolution,” Advances in Archaeological Practice, vol. 11, no. 1, pp. 90-103, 2023. https://doi.org/10.1017/aap.2022.42.

[4] G. D. Malaperdas, “Practical methods of GIS for archaeologists: Viewshed analysis-the kingdom of pylos example. Geoplanning: Journal of Geomatics and Planning, vol. 8, no. 1, pp. 1-22, 2021. https://doi.org/10.14710/geoplanning.8.1.1-22

[5] D. A. Contreras, Z. Batist, C. Zogheib, and T. Carter, “Matching pragmatic lithic analysis and proper data architecture: The QuARI R shiny database interface,” Advances in Archaeological Practice, vol. 9, no. 4, pp. 299-311, 2021. https://doi.org/10.1017/aap.2021.11

[6] M. A. Jubair, S. A. Mostafa, A. Mustapha, Z. Baharum, M. A. Salamat, and A. Erianda, “A multi-agent K-Means algorithm for improved parallel data clustering,” JOIV: International Journal on Informatics Visualization, vol. 6, no. 1-2, pp. 145-150, 2022. https://doi.org/10.30630/joiv.6.1-2.934

[7] M. Previtali, and R. Valente, “Archaeological documentation and data sharing: digital surveying and open data approach applied to archaeological fieldworks,” Virtual Archaeology Review, vol. 10, no. 20, pp. 17-27, 2019. https://doi.org/10.4995/var.2019.10377

[8] A. Korf, S. Hammann, R. Schmid, M. Froning, H.

Hayen, and L. J. Cramp, "Digging deeper-A new data mining workflow for improved processing and interpretation of high resolution GC-Q-TOF MS data in archaeological research," Scientific Reports, vol. 10, no. 1, pp. 1-9, 2020. https://doi.org/10.1038/s41598-019-57154-8

[9] J. Tronicke, N. Allroggen, F. Biermann, F. Fanselow, J. Guillemoteau, C. Krauskopf, and E. Lück, "Rapid multiscale analysis of near-surface geophysical anomaly maps: Application to an archaeogeophysical data set," Geophysics, vol. 85, no. 4, pp. 109-118, 2020. https://doi.org/10.1190/geo2019-0564.1

[10] E. Kansa, and S. W. Kansa, "Digital data and data literacy in archaeology now and in the new decade," Advances in Archaeological Practice, vol. 9, no. 1, pp. 81-85, 2021. https://doi.org/10.1017/aap.2020.55

[11] E. R. Otárola-Castillo, M. G. Torquato, J. Wolfhagen, M. E. Hill, and C. E. Buck, "Beyond chronology, using bayesian inference to evaluate hypotheses in archaeology," Advances in Archaeological Practice, 2022, 10(4): 397-413. https://doi.org/10.17605/OSF.IO/54F62

[12] Y. Li, J. Qi, X. Chu, and W. Mu, "Customer segmentation using k-means clustering and the hybrid particle swarm optimization algorithm," The Computer Journal, vol. 66, no. 4, pp. 941-962, 2023. https://doi.org/10.1093/comjnl/bxab206

[13] C. Yuan, and H. Yang, "Research on K-value selection method of K-means clustering algorithm," J - Multidisciplinary Scientific Journal, vol. 2, no. 2, pp. 226-235, 2019. https://doi.org/10.3390/j2020016

[14] X. Huang, M. Ren, and Z. Hu, "An improvement of K-Medoids clustering algorithm based on fixed point iteration," International Journal of Data Warehousing and Mining, vol. 16, no. 4, pp. 84-94, 2020. https://doi.org/10.4018/IJDWM.2020100105

[15] G. Qi, W. Guan, Z. He, and A. Huang, "Adaptive kernel fuzzy C-Means clustering algorithm based on cluster structure," Journal of Intelligent and Fuzzy Systems, vol. 37, no. 2, pp. 2453-2471, 2019. https://doi.org/10.3233/JIFS-182750

[16] A. Sabir, H. A. Ali, and M. A. Aljabery, "ChatGPT tweets sentiment analysis using machine learning and data classification," Informatica, vol. 48, no. 7, pp. 103-112, 2024. https://doi.org/10.31449/inf.v48i7.5535

[17] M. H. Santoso, "Application of association rule method using apriori algorithm to find sales patterns case study of indomaret tanjung anom," Brilliance: Research of Artificial Intelligence, vol. 1, no. 2, pp. 54-66, 2021. https://doi.org/10.47709/brilliance.v1i2.1228

[18] F. Moodi, and H. Saadatfar, "An improved K-means algorithm for big data," IET Software, vol. 16, no. 1, pp. 48-59, 2022. https://doi.org/10.1049/sfw2.12032

[19] A. R. Khan, S. Khan, M. Harouni, R. Abbasi, S. Iqbal, and Z. Mehmood, "Brain tumor segmentation using K-means clustering and deep learning with synthetic data augmentation for classification," Microscopy Research and Technique, vol. 84, no. 7, pp. 1389-1399, 2021. https://doi.org/10.1002/jemt.23694

[20] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data," Quality and Quantity, vol. 56, no. 3, pp. 1283-1291, 2022. https://doi.org/10.1007/s11135-021-01176-w

[21] R. Raymond, and M. A. Savarimuthu, "Retrieval of interactive requirements for data intensive applications using random forest classifier," Informatica, vol. 47, no. 9, pp. 35-50, 2023. https://doi.org/10.31449/inf.v47i9.3772

[22] T. Mahmood, and Z. Ali, "Prioritized muirhead mean aggregation operators under the complex single-valued neutrosophic settings and their application in multi-attribute decision-making," Journal of Computational and Cognitive Engineering, vol. 1, no. 2, pp. 56-73, 2022. https://doi.org/10.47852/bonviewJCCE2022010104

[23] T. T. T. Tran, T. N. Nguyen, T. T. Nguyen, G. L. Nguyen, and C. N. Truong, "A fuzzy association rules mining algorithm with fuzzy partitioning optimization for intelligent decision systems," International Journal of Fuzzy Systems, vol. 24, no. 5, pp. 2617-2630, 2022. https://doi.org/10.1007/s40815-022-01308-w

[24] K. Berahmand, M. Mohammadi, A. Faroughi, and R.

P. Mohammadiani, “A novel method of spectral clustering in attributed networks by constructing parameter-free affinity matrix,” Cluster Computing, vol. 25, no. 2, pp. 869-888, 2022. https://doi.org/10.1007/s10586-021-03430-0

[25] L. Jiao, H. Yang, Z. Liu, and Q. Pan, “Interpretable fuzzy clustering using unsupervised fuzzy decision trees,” Information Sciences, vol. 611, no. 8, pp. 540-563, 2022. https://doi.org/10.1016/j.ins.2022.08.077

Unmanned Logistics Vehicle Control based on Path Tracking Control Algorithm

Menglin Wu*, Zhenyu Liu
Department of Economic Management, Zhengzhou Vocational College of Finance and Taxation, Zhengzhou 450000, China
E-mail: 17335743056@163.com
*Corresponding author

Keywords: unmanned logistics vehicles, path tracking control algorithms, model predictive control algorithms, logistics industry, variable structure control

Received: March 20, 2024

The logistics industry has made significant progress in recent years. However, there are still issues with low operational efficiency and high costs. Unmanned logistics vehicles have gained attention as an efficient and intelligent mode of transportation with the rapid development of the industry. The study utilizes an advanced path tracking control algorithm, in combination with model predictive control technology, to monitor and adjust the path, speed, and direction of unmanned logistics vehicles in real-time. The aim is to enhance the stability, safety, and efficiency of travel. The experiments revealed that the average accuracy of path deviation prediction of the proposed model on two different datasets is 88.33% and 82.1%, which is 3.96% and 4.72% higher than that of the control model, respectively. The control accuracy of the proposed model reached 94.19% on the KITTI Vision Benchmark Suite dataset and 95.61% on the CARLA Simulator dataset, which are both higher than the other control models. In addition, the study also tested the proposed model for energy consumption, controller switching frequency, lateral error and other indexes, and the findings revealed that the proposed model of the study exhibits high stability and efficiency. This research not only provides new ideas for the control of unmanned logistics vehicles, but also verifies the effectiveness of the control strategy through experiments.

Povzetek: Študija vpelje napredni algoritem za sledenje poti v kombinaciji s prediktivnim modelom za nadzor brezpilotnih logističnih vozil, kar izboljšuje stabilnost, varnost in učinkovitost prevoza z visoko kvaliteto nadzora.

1 Introduction

In traditional logistics transportation, manual operation and scheduling are essential, but manual operation and scheduling have defects such as high labor cost, inefficient operation, prone to human error, and lack of real-time monitoring and data analysis [1-2]. With the rapid development of electronic, communication and computer technologies, unmanned logistics vehicles (ULV) control has become an important way to solve the efficiency and safety problems in the logistics industry [3]. Compared to traditional logistics, ULVs have improved in path tracking, speed control and safety, which improves logistics efficiency and reduces operational costs. ULV control refers to the use of path tracking control (PTC) technology and driverless technology to realize autonomous operation and control of logistics vehicles [4]. By introducing driverless technology and ULV algorithms, the transportation efficiency of logistics vehicles can be improved, the risk of accidents can be reduced, and the pressure on human resources can be alleviated. In addition, research on ULV algorithms has benefited from advances in computer vision, sensor technology and artificial intelligence [5]. With the help of high-precision maps, LIDAR, cameras, and other sensors, ULVs can sense and understand their surroundings in real time to better plan their paths and avoid obstacles. There are two innovations in this research: first, the linear six-degree-of-freedom (L6DOF) vehicle dynamics model (V-MPC) is optimized based on variable structure control (VSC) to simplify the output structure of the dynamics model. The second point is that the objective function and output function of the model are optimized using model predictive control (MPC), which improves the stability of the model path tracking. The structure of this article is divided into six parts: The first part is related work, which will review the literature to summarize the development status, application scenarios, etc. of ULV technology. The second part is the methodology, which constructs the unmanned vehicle operation model through MPC and VSC algorithms. The performance test, which is the third section, is how this experiment verifies the suggested model's functionality. The fourth part is the discussion, which compares the design method with relevant literature and analyzes its

advantages. The study's weaknesses and findings are compiled in the conclusion, which makes up the fifth section. The sixth part is Ethical and Safety Considerations, which introduces the ethical and safety standards followed by the design method.

ULV is the current hotspot in the research of driverless technology, and many researchers have explored for the realization method of ULV. Hang et al. proposed a novel four-wheel steering electric vehicle as an automatic ground vehicle. To realize the automatic driving of this automatic ground vehicle, they constructed a linear variable parameter system model to adapt the ULV apparatus to different longitudinal speeds and road friction coefficients. The experimental results indicated that the linear variable-parameter system model constructed by the study exhibited better path tracking performance [6]. Path tracking is one of the main responsibilities of self-driving automobiles, as discovered by Chen et al. They consequently used deep neural network techniques to optimize path tracking. In this study, a combination of proximal policy optimization and pure optimization was chosen to build the vehicle controller architecture. The combination of the two algorithms makes the overall operation of the control system more robust and improves the additivity of the controller. The results indicated that the optimized control system's path tracking capability was significantly enhanced under low-speed driving conditions [7]. Lin et al. used the ULV algorithm to optimize a linear three-degree-of-freedom V-MPC and proposed an integrated control method combining the MPC and ULV algorithms. The MPC was typically used to avoid yaw, while the ULV algorithm maintains vehicle roll stability by controlling the braking force of each tire. The study's comparison of the suggested model with popular models like CarSim revealed that the enhanced control system, which uses the ULV algorithm, produces smoother outcomes [8]. Sun et al. proposed a fast non-singular terminal sliding mode control strategy with a double hidden layer output feedback neural network. The strategy improved the sliding mode control using feedback neural network and ULV algorithm, and constructed vehicle kinematics and dynamics models. Simulation experiments yielded that the control algorithm of the proposed model has significant advantages such as high tracking accuracy, fast convergence, and robustness compared with the traditional sliding mode controller [9].

MPC is one of the commonly used algorithms in current dynamics research, and more extended applications of this algorithm have emerged. Pan et al. proposed a new multilayer graph architecture based on the MPC algorithm to achieve scalability of interaction networks. The architecture established formation control laws for autonomous formation, formation maintenance, collision and obstacle avoidance using the MPC algorithm on the basis of multilayer graphs. Finally, the experiments adopted the proposed framework to accomplish the formation maintenance and trajectory tracking tasks in the constraint space, which verified the feasibility of the proposed framework [10]. Çimen et al. combined the MPC algorithm with the firefly optimization algorithm, thus proposing a new optimization algorithm. The optimization algorithm was used to study single-peak, multi-peak, composite and CEC-C06 2019 benchmark optimization and optimal design parameter determination problems. It was experimentally verified that the proposed model possesses lower loss and higher performance efficiency than the current common models of the same type [11]. Beus et al. proposed a load/frequency controller for hydraulic turbine governor based on MPC. This controller was updated by linearly predicting the operating points of the model parameters, thus greatly improving the stability and operating efficiency of the frequency controller. The proposed controller was experimentally compared with several state-of-the-art controllers such as particle swarm optimization-based PI controller. Additionally, the findings showed that the research's suggested model benefits from having a straightforward structure and quick calculation [12]. A hierarchical non-linear MPC method for cooperative control of vehicle-vehicle networks was proposed by Liu et al. The algorithm employs a multilayer structure and solves its optimization problem by continuous/generalized minimum residual method. To enhance tracking performance, the algorithm also included a controller with a double-loop structure. According to experimental findings, the suggested method outperforms the most advanced vehicle cooperative control models in terms of stability and error rate, proving its usefulness in real-world scenarios [13].

To clarify the advantages and disadvantages of PTC technology and MPC technology, the research summarized the literature review in Table 1. Further optimization is needed based on current research due to the low computational efficiency and implementation difficulties found in previous studies.

Table 1: Summary of related work content

Reference	Improvement/Application direction	Results
[6]	This study was based on a linear variable parameter system model to modify four-wheel steering electric vehicles	The linear variable parameter system model has better path tracking performance
[7]	This study utilized deep neural network technology to optimize path tracking algorithms	Under low-speed driving conditions, the path tracking ability of the optimized control system is significantly improved
[8]	This study optimized the linear three degree of freedom V-MPC using the ULV algorithm	The enhanced control system using ULV algorithm produces smoother results
[9]	This study was based on optimizing the path tracking algorithm using a dual hidden layer output feedback neural network	The optimized control algorithm has the characteristics of high tracking accuracy and fast convergence speed
[10]	This study proposed a multi-layer graph architecture based on the MPC algorithm	This architecture diagram completes formation maintenance and trajectory tracking tasks in the constrained space
[11]	This study optimized the MPC algorithm based on fireflies	The optimized algorithm has lower loss and higher performance efficiency
[12]	This study proposed a load/frequency controller for a water turbine governor based on MPC	Proposed model has a simple structure and fast computing power
[13]	This study proposed a hierarchical nonlinear MPC method for vehicle network collaborative control	The effectiveness of this method in real-life scenarios has been verified through experiments

In summary, ULV has now achieved greater results in small unmanned vehicle applications after recent years of development. However, the current commonly used ULV algorithms still have defects such as high computational complexity and difficult to realize. Therefore, the research tries to use VSC algorithm and MPC algorithm to optimize the more complex L6DOF, aiming to construct a stable and efficient ULV control algorithm.

2 Improved control of unmanned logistics vehicles based on VSC and MPC algorithms

ULV has steadily come to dominate the logistics and transportation fields due to the industry's rapid development as well as the ongoing advancements in science and technology. The traditional manual driving logistics vehicles have some limitations in terms of efficiency, safety and environmental protection. To overcome these shortcomings, the study uses advanced control algorithms, such as MPC and vehicle stability control, to construct a stable and efficient ULV control model. This will help promote technological innovation and sustainable development in the logistics industry.

2.1 Path tracking model based on VSC

The VSC algorithm is a control strategy whose core idea is to utilize sliding modal hyperplanes to achieve fast, non-linear switching of the dynamic characteristics of the system [14-15]. This algorithm is robust in dealing with uncertainties and perturbations because the sliding modal hyperplane can be adaptively adjusted according to the changes in the system state. In ULV, the VSC algorithm is used to design controllers that enable the system to track the desired path quickly and accurately [16]. The VSC approach has garnered a lot of interest in the ULV sector because of its benefits, which include quick response times and insensitivity to changes an system parameters [17]. In order to better study the path selection condition of logistics vehicles on the way of transporting goods, it is necessary to construct a V-MPC. A L6DOF is used in this study. It is assumed that c denotes the spacing from the right wheel of the vehicle to the center of mass and d denotes the spacing from the left wheel to the center of mass. In order to facilitate the analysis in hand, the study establishes a new temporary coordinate system for each analysis point, and the temporary coordinate system is $x-y$. Then F_x denotes the force received in the direction of x of the temporary coordinate system, and F_y denotes the force received in the direction of y of the temporary coordinate system. The structure of the model is shown in Figure 1.

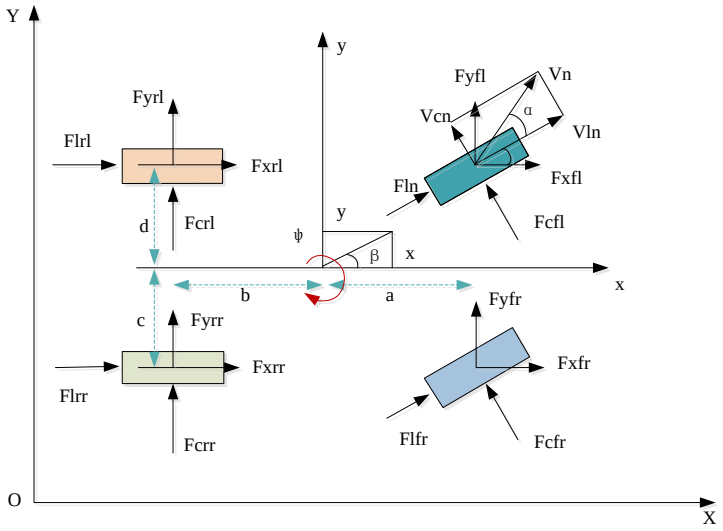


Figure 1: Hand analysis of linear six-degree-of-freedom vehicle dynamics model

F_c in Figure 1 shows the lateral force on the vehicle tires, which has a significant impact on the handling, stability and safety of the vehicle. F_l, on the other hand, denotes the longitudinal force, which acts similarly to the lateral force and are both forces that maintain stable vehicle motion. In addition, α, β is the center of mass and the lateral deflection angle of the tire in the figure, respectively, and δ denotes the tire deflection angle. If the vehicle's power wheel deflection angle stays constant, the vehicle's motion will likewise be constant according to this model [18]. Currently, equation (1) displays the relationship between the steering wheel rotation angle and the transverse angular velocity at the vehicle's center of mass. This ratio can reflect the stability of the current state of the vehicle.

$$G_\omega = \frac{v_x}{i_{sw} L\left(1+\alpha v_x^2\right)} \tag{1}$$

In equation (1), ω denotes the pendulum angular velocity at the center of mass, and L denotes the horizontal distance between the front and rear axles of the vehicle. v_x denotes the velocity of the horizontal motion of the vehicle, and i_{sw} denotes the ratio of the front wheel angle to the control angle. K denotes the stabilization factor, and G_ω denotes the ratio of the pendulum angular velocity at the center of mass to the steering wheel turning angle. The value of K is related to the mass of the vehicle, the distance between axles, etc., and its computational expression is shown in equation (2) in s^2/m^2.

$$K = \frac{m}{L^2}\left(\frac{b}{C_f} - \frac{a}{C_r}\right) \tag{2}$$

In equation (2), C_f denotes the front tire lateral deflection stiffness and m denotes the mass. b stands for the distance between the vehicle's front axle and center, while C_r stands for the rear tires' stiffness in terms of lateral deflection. The a represents the separation between the vehicle's center and rear axle. In addition, the vehicle center of mass lateral deflection (MLD) also has a large effect on vehicle stability, the center of MLD affects the stability of vehicle motion by increasing the steering wheel stability [19]. The gain coefficient of center of MLD on steering wheel stability is shown in equation (3).

$$G_\beta = \frac{b + \frac{mv^2 a}{LC_r}}{i_{sw} L\left(1+Kv_x^2\right)} \tag{3}$$

In equation (3), G_β denotes the gain coefficient of the center of MLD on steering wheel stability. β denotes the center of MLD angle. If the vehicle transverse swing angle and the horizontal direction of the running state remain unchanged, at this time the instantaneous speed of the vehicle doing circular motion can be approximated as the tangential direction of the speed. At this time, the instantaneous velocity and motion path of the vehicle are shown in Figure 2.

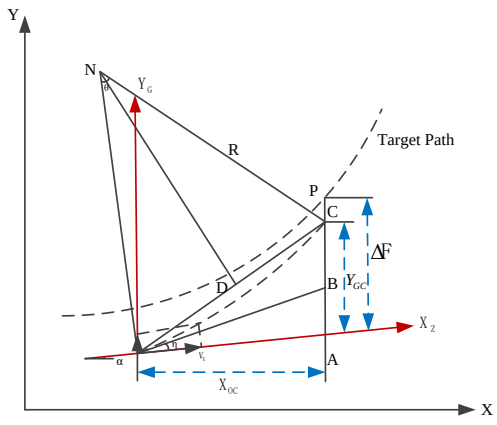


Figure 2: Instantaneous velocity and motion path diagram of the vehicle

In Figure 2, $\angle CGB$ denotes the chordal tangent angle of the vehicle's trajectory, and therefore $\angle CGB$ is equal to the $\frac{1}{2}$ of the circumferential angle of the vehicle's trajectory. Therefore, the position of the vehicle in the direction of Y_G changes as shown in equation (4).

$$y_G = \tan\left(\frac{\theta}{2}+\beta\right)x_G \tag{4}$$

In equation (4), y_G represents the displacement in the Y_G direction on the $X_G - Y_G$ A coordinate system established with the vehicle's last moment of motion point G as the origin, and x_G represents the displacement in the X_G direction. When the vehicle is in the process of uniform speed steering, the relationship between the angles in equation (4) can be deduced from the ideal state of the steering wheel angle of rotation angle of the expression, which becomes larger as shown in equation (5).

$$\delta_{sw} = \frac{2\arctan\left(\frac{\Delta f}{v_x t_p}\right)}{G_\omega t_P + 2G_\beta} \tag{5}$$

In equation (5), Δf denotes Y_{GC} in Figure 2 and t_p denotes the time period. The kinetic model also serves to predict the vehicle motion deviation. The algorithm's prediction of the deviation is related to the projected distance L_{ag} from the point of motion to the predicted point at the previous moment, and the expression for the calculation of the motion deviation is shown in equation (6).

$$\Delta f = \frac{d_{la}}{\cos\Delta\varphi} = \frac{d_{la}}{\cos\left(\arcsin\left(\frac{d_{la}}{L_{ag}}\right)\right)} \tag{6}$$

In equation (6), d_{la} denotes the spacing between the predicted path to the projection point. Δf denotes the motion deviation of the vehicle, and $\Delta\varphi$ denotes the angle between the target path and the actual path. At this point, the non-L6DOF V-MPC is constructed, so the study needs to optimize this model according to the VSC algorithm. Convergence law design is an important task in the design of VSC algorithm. Convergence law is a control strategy designed to make the system state gradually converge to the target state or desired value [20]. The role of convergence law design is to improve the stability, tracking performance and robustness of the system by gradually adjusting the control signals so that the system state gradually converges to the target state. The study uses the exponential convergence law to optimize the model, and the mathematical expression of the exponential convergence law is shown in equation (7).

$$s = -\eta\,\mathrm{sgn}(s) - ks \tag{7}$$

In equation (7), $\mathrm{sgn}(.)$ denotes the sign function, η, k both denote the parameters of the exponential convergence law, and both are rational numbers greater than zero. s denotes the convergence law value. In addition, vibration elimination is also a more important part. Vibration elimination requires the definition of the saturation function, and the expression of the conservation function is shown in equation (8).

$$sat(s) = \begin{cases} \mathrm{sgn}(s), |s| > \psi \\ s, |s| \le \psi \end{cases} \tag{8}$$

In equation (8), ψ denotes the width of the boundary layer. $|.|$ denotes the absolute value taken. In the dynamics model, the setting of the boundary layer can affect the stability and convergence of the simulation. Reasonable setting of the boundary layer can reduce the numerical error and oscillation phenomenon, improve the simulation effect.

2.2 Path tracking model construction based on MPC improvement

Based on the above, the study constructed a non-L6DOF path-tracking model based on VSC, but the switching process of VSC usually introduces a control pulse, i.e., a sudden change of the control input on the sliding mode surface [21]. This impulse signal may lead to a non-ideal response of the system, and therefore new methods need to be introduced to further optimize the model. The MPC algorithm is an optimization control method that is widely used in fields such as industrial process control and motor vehicle control. Utilizing a dynamic model of the system, it makes predictions and solves an optimization problem in each control cycle to determine the ideal control inputs. Its basic steps include system modeling, prediction, optimization problem definition, optimization solution and application of control inputs

[22]. The algorithm is able to take into account the non-linear characteristics and constraints of the system as well as dynamic optimization during each control cycle. The control schematic of the MPC algorithm is shown in Figure 3.

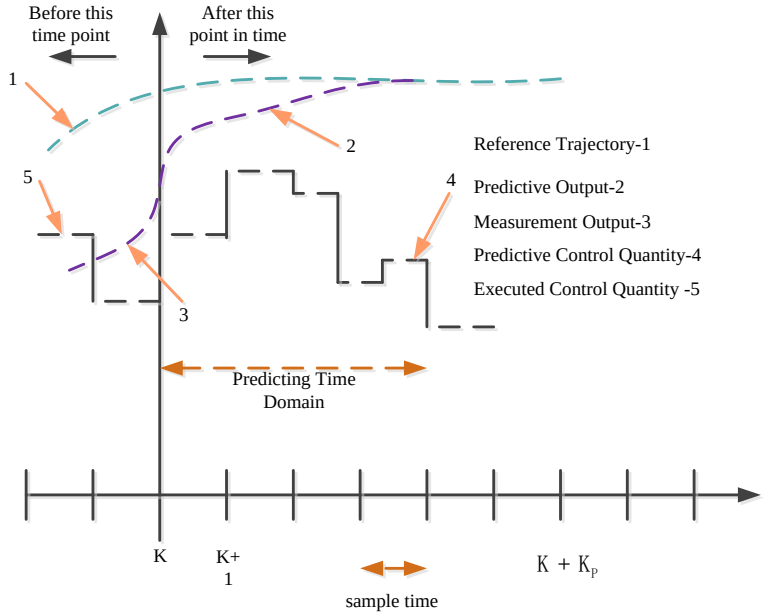


Figure 3: MPC algorithm control principal diagram

The MPC algorithm redefines the instantaneous states, control parameters, and outputs of the above V-MPC, which are shown in equation (9).

$$\begin{cases} \xi(t) = f\left(\xi(t), u(t)\right) \\ \zeta(t) = h\left(\xi(t), u(t)\right) \end{cases} \tag{9}$$

In equation (9), $\xi(t)$ is the state vector of the model and $\zeta(t)$ denotes the output of the model. t denotes the time and $u(t)$ denotes the control vector of the model. Discreteization of the dynamics model is also required when instantaneous states, control parameters, and outputs have been redefined. The original non-linear model is to be transformed into a discrete linear time-varying model using the discretization method. Equation (10) displays the state vector expression following the discretization process.

$$A\xi(t) = B\xi(t) + Cu(t) \tag{10}$$

In equation (10), A, B, C are all discretized process variables, and the specific value of A, B, C can be derived by solving equation (9) in a generalized way. Moreover, the discretized state vector of $t+1$ moment is calculated as shown in equation (11).

$$\xi(t+1) = B'\xi(t) + C'u(t) + d_t \tag{11}$$

In equation (11), B', C' is the discretized process variable and d_t is the special solution of equation (9) at moment t. The kinetic model can be iterated after discretization so as to calculate the discrete values at each moment. Objective function design, constraint design, and prediction function design are needed after the model is discretized. Among them, the objective function design plays a key role, and the reasonable selection and definition of the objective function can clarify the problem objective, guide the problem modeling, provide measures and constraints, and affect the behavior and final results of the optimization algorithm [23]. Assuming that N_p denotes the prediction step and Y_p denotes the output prediction value. Equation (12) displays the expression for the MPC algorithm's objective function.

$$J\left(\xi(t), \Delta U(t), \varepsilon\right) = \sum_{i=1}^{N_p} \| Y_p(k+i) - Y_{ref}(k+i) \|_Q^2 + \sum_{i=1}^{N_c} \| \Delta u(k+i) \|_R^2 + \rho\varepsilon^2 \tag{12}$$

In equation (12), N_c denotes the control step size and ε denotes the relaxation factor. $\|.\|$ denotes the Euclidean paradigm operation, Y_{ref} denotes the control output ideal, and Q, R, ρ denotes the corresponding weight matrix. Path tracking accuracy and stability can be increased by the system by tracking the intended path more precisely and by decreasing the goal function. In addition to guaranteeing that the state of the system is always within the safety range, the model constraints also ensure the viability of the control inputs and state variables [24]. The performance of the model and its ability to adapt to different demands can be effectively improved by setting constraints reasonably. The constraints of the model of this research are mainly for four aspects: output, control weight, weight change rate, and relaxation factor. The specific constraints are shown

in equation (13).

$$\begin{cases} Y_{\min} \le Y(t) \le Y_{\max} \\ u_{\min} \le u(t) \le u_{\max} \\ \Delta U_{\min} \le \Delta U \le \Delta U_{\max} \\ \varepsilon_{\min} \le \varepsilon \le \varepsilon_{\max} \end{cases} \tag{13}$$

In equation (13), $Y(t)$ denotes the vehicle output quantity and $u(t)$ denotes the core control quantity of the model. ΔU denotes the weight change of the control quantity and ε denotes the relaxation factor. The wrong iteration loop will end earlier after the model is given the constraints, so the computational efficiency of the model is further improved.

Currently, the model's output function differs from the pre-optimization, and equation (14) displays the optimized prediction function's expression.

$$y(t) = \varepsilon\xi(t) + \tau\Delta U(t) + \mu_t \tag{14}$$

In equation (14), μ_t denotes the amount of model control at the current moment, and the acquisition of the $\xi(t)$ value relies on the sensors of the simulation model. The introduction of the MPC algorithm allows the model to perform a comprehensive path deviation analysis and correct the route in a timely manner through multi-point prediction. The structure of multi-point deviation prediction is shown in Figure 4.

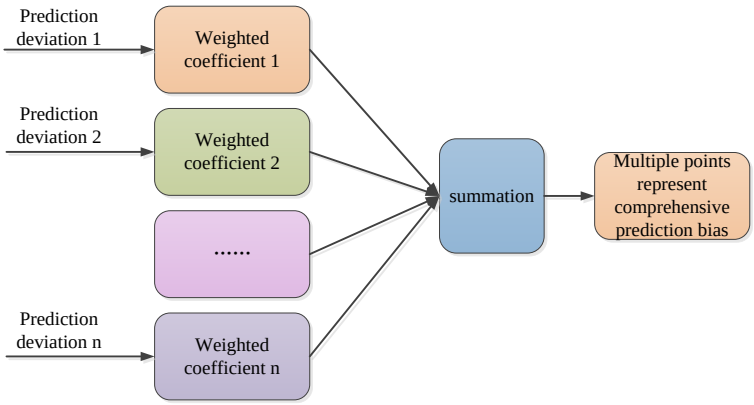


Figure 4: Structure diagram of multi-point deviation prediction

After the optimization of VSC algorithm and MPC algorithm, the V-MPC can be improved in terms of control performance, stability, tracking performance and energy consumption optimization. These optimizations can improve the vehicle handling, safety and energy efficiency performance. The flow chart of the optimized V-MPC operation is shown in Figure 5.

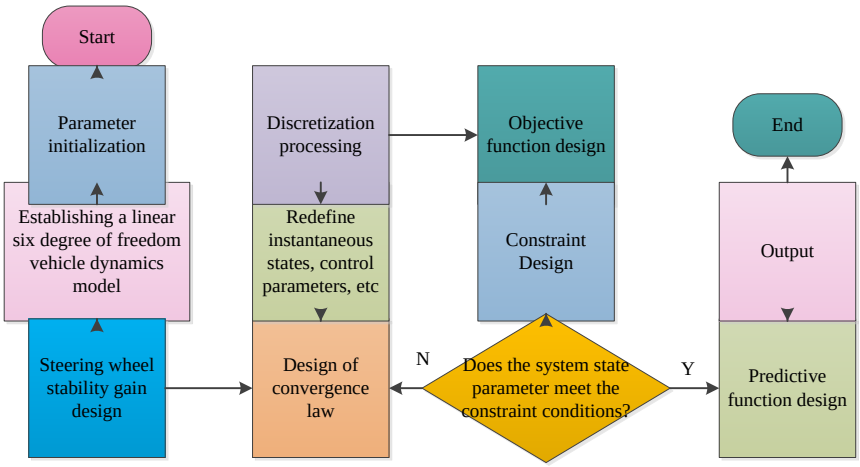


Figure 5: V-MPC model operation flow chart

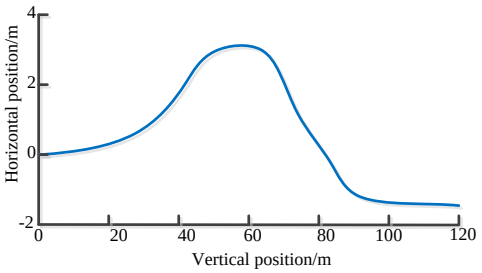
In practical logistics scenarios, the proposed model addresses the issue of low accuracy in traditional PTC algorithms by enhancing the path tracking algorithm. This improvement concept also has implications for other

map recognition and automation control. However, it should be noted that this research method still has limitations in more precise motion control scenarios. Consequently, the proposed enhanced PTC algorithm is unable to handle path control tasks with low accuracy, such as subways and light rail.

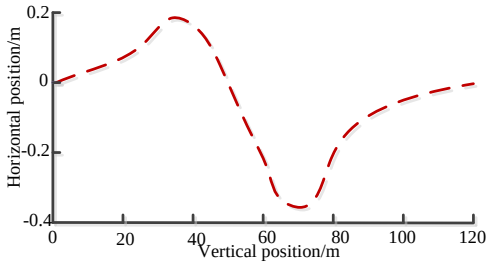
3 Testing and analyzing the effect of V-MPC model application

The performance test of the V-MPC model is conducted on a desktop computer with i7-13700K CPU, GeForce RTX 3090 GAMER OC graphics card, and CentOS 7. The experiment utilizes the KITTI Vision Benchmark Suite (KITTI) and CARLA Simulator (CARLA) data sets, which contain sensor data from autonomous vehicles collected on real roads, including camera images and laser radar data. Although the KITTI and CARLA datasets provide a wealth of sensor and environmental data, there are some limitations to both datasets. KITTI data sets are mainly collected under clear weather conditions, which may limit their applicability in severe weather conditions such as rain and snow. Because CARLA is a simulation environment, its data may not fully simulate the complexity and variability of the real world. Therefore, to ensure control of experimental variables, all models participating in the performance analysis experiment will undergo a 2-hour pre-training session. The pre-training dataset will consist of randomly selected data samples from KITTI, while ensuring consistency between the pre-training system environment and hardware environment.

The control models used in this experiment are the particle swarm optimization-based PI controller model (PSO-PI) and the L6DOF. The double-shifted reference trajectories of the L6DOF model and the V-MPC model on the CARLA dataset are shown in Figure 6. Figure 6(a) represents the double-shifted line reference trajectory of the L6DOF model, from which it can be seen that the transverse position of the L6DOF model has a large magnitude of variation, indicating that the stability of the L6DOF model is poor. Figure 6(b) represents the double-shifted line reference trajectory of the V-MPC model, and the transverse position variation of the V-MPC model is smaller, indicating that the model is more stable.



(a) L6DOF model double shift line reference trajectory



(b) V-MPC model double shift line reference trajectory

Figure 6: Reference trajectories of double moving lines for different models

Figure 7 compares the transverse errors of the V-MPC model and the L6DOF model at various speeds using the KITTI dataset. The lateral error findings of the

L6DOF model and the V-MPC model are shown in Figures 7(a) and 7(b), respectively. The error variation in the first half of the route is essentially the same for the two models, according to the data; nevertheless, the V-MPC model's error value is lower than the L6DOF models. Moreover, the error variation of V-MPC model in the second half of the journey is obviously better than the control model.

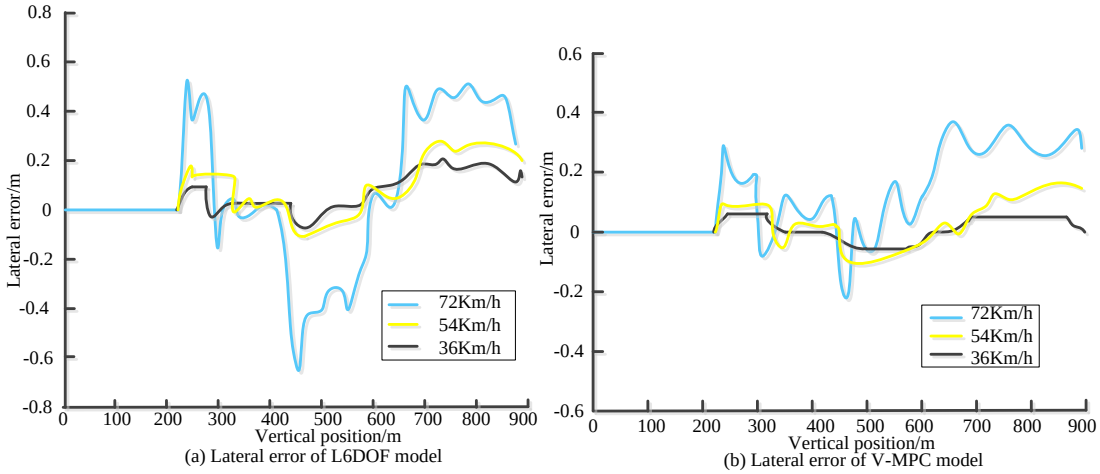


Figure 7: Comparison of lateral errors of different models

The comparison of path deviation prediction accuracy of PSO-PI model and V-MPC model is shown in Figure 8. The accuracy of each model's path deviation prediction using the CARLA dataset is shown in Figure 8(a). On this dataset, the V-MPC model's average accuracy is 88.33%, greater than the PSO-PI model's 84.64%. Figure 8(b) represents the path deviation prediction accuracy of different models on the KITTI dataset. Based on Figure 8(b), the V-MPC model has an average accuracy of 82.1%, which is 4.72% greater than the PSO-PI model.

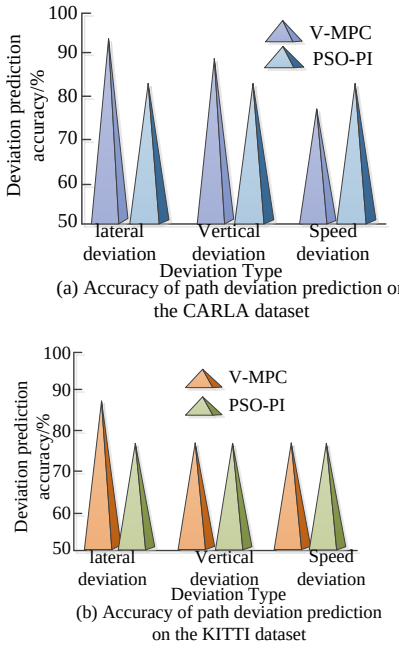


Figure 8: Comparison of accuracy of path deviation prediction among different models

The energy consumption metrics are employed to assess the optimized kinetic models' energy usage efficiency. Figure 9 shows a comparison of the PSO-PI and V-MPC models' energy consumption across various datasets. The energy usage of the models using the KITTI dataset is shown in Figure 9(a). The energy usage of the two models on the CARLA dataset is shown in Figure 9(b). In Figure 9, the change in energy consumption of V-MPC model with increasing distance traveled is smoother, so it is concluded that V-MPC model has better stability during operation.

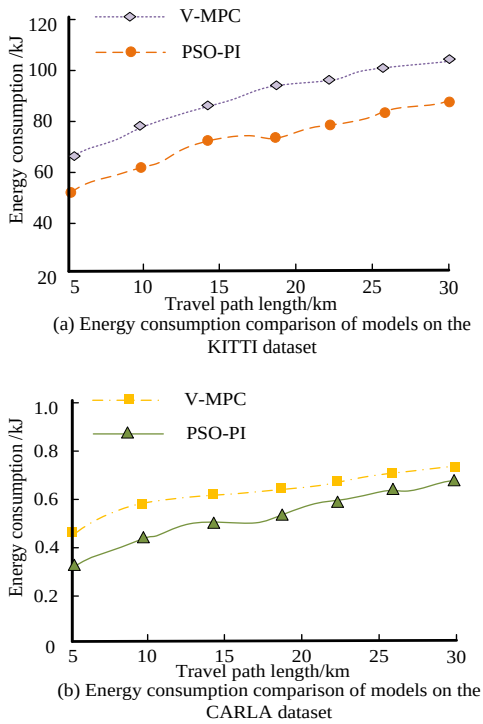
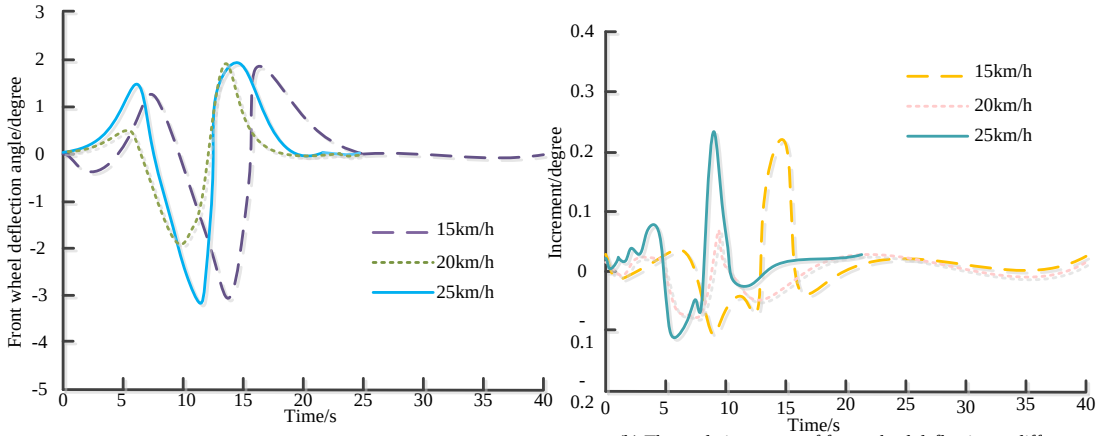


Figure 9: Comparison of energy consumption of different models

Figure 10 represents the variation of front wheel deflection angle and deflection increment versus time when the driverless car equipped with the V-MPC model is doing uniform circular motion. Figure 10(a) shows the

front wheel deflection angle versus time at different vehicle speeds, while Figure 10(b) shows the angular increment of front wheel deflection at different vehicle speeds. The sub-experiment tests the stability of the model steering by circular motion. In Figure 10, the curve fluctuation is minimized at a speed of 20km/h. Therefore, it is concluded that the steering stability of the vehicle segment is optimal when the initial speed is 20km/h in the experimental environment, while the speed is too fast or too slow will lead to the reduction of stability.

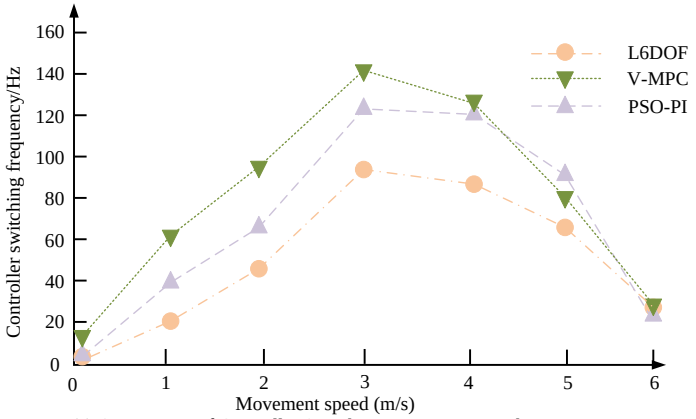


(a) The relationship between front wheel deflection angle and time at different vehicle speeds

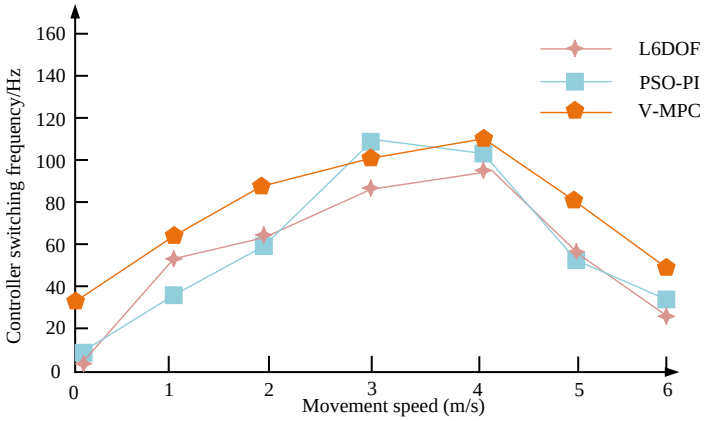
(b) The angle increment of front wheel deflection at different vehicle speeds

Figure 10: Changes in the relationship between front wheel deflection angle and deflection increment with time

For the VSC algorithm, the controller switching frequency is an important index, and the controller switching frequency affects the response speed, stability and energy consumption. Figure 11 displays the findings of this study, which analyzes the controller switching frequencies of the three models—PSO-PI, V-MPC, and L6DOF—using various datasets. The controller switching frequencies of the three models on the KITTI and CARLA datasets are shown in Figures 11(a) and 11(b), respectively. The figures demonstrate that there is a considerable difference in switching frequency between the KITTI and CARLA datasets. It can also be concluded that the controller switching frequency decreases gradually when the speed is too large or too small, and peaks at a speed of 3-4 m/s.



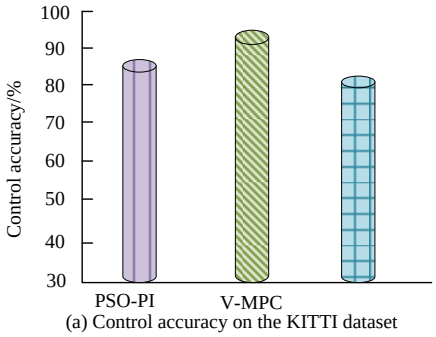
(a) Comparison of Controller Switching Frequencies on the KITTI Dataset



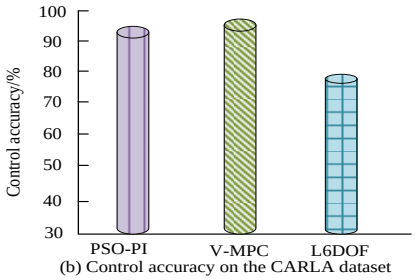
(b) Comparison of Controller Switching Frequencies on the CARLA Dataset

Figure 11: Comparison of switching frequencies of different model controllers

The control accuracy indicates the accuracy of the optimized dynamics model for the control inputs. Figure 12 displays a comparison of the three models' control accuracy using the KITTI and CARLA datasets: PSO-PI, V-MPC, and L6DOF. The control accuracy of the V-MPC model in Figure 12(a) is 94.19%, which is 8.73% higher than the L6DOF model and 4.70% higher than the PSO-PI model. Figure 12(a) shows the control accuracy of the three models on the KITTI dataset. Figure 12(b) represents the control accuracy of the three models on the CARLA dataset. Based on the experimental findings, the V-MPC model has a higher control precision with an accuracy of 95.61%, compared to the PSO-PI model with an accuracy of 91.76% and the L6DOF model with an accuracy of 81.62%.



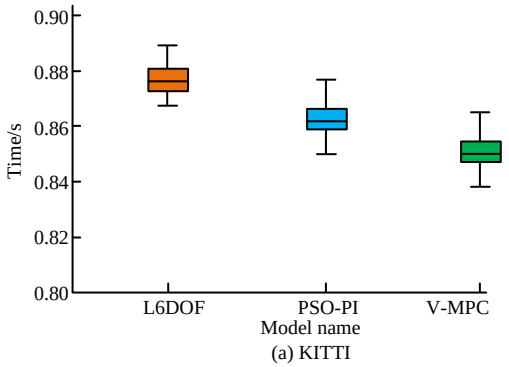
(a) Control accuracy on the KITTI dataset



(b) Control accuracy on the CARLA dataset

Figure 12: Comparison of control accuracy of different models

To assess the complexity of the V-MPC, PSO-PI, and L6DOF algorithms, the KITTI and CARLA datasets are used as inputs, and the output time of the corresponding models can intuitively reflect the complexity of different algorithms. Figure 13 illustrates a comparison of the output times for each algorithm. Figure 13 (a) compares the complexity of the V-MPC, PSO-PI, and L6DOF models on the KITTI dataset, while Figure 13 (b) compares the complexity of different models on the CARLA dataset. The figures show that the V-MPC model has a shorter output time on both datasets (0.856s and 0.818s, respectively) compared to the PSO-PI and L6DOF models. Therefore, the proposed model is more progressive in terms of algorithm complexity. Based on the experiment results, it is evident that the proposed model has high control accuracy, making it suitable for scenarios with long driving times and high path accuracy requirements, such as unmanned express delivery. Additionally, the proposed model exhibits strong scalability. The MPC algorithm can introduce dynamic parameter optimization algorithms to accelerate the convergence speed and computational efficiency of the model because it predicts through a system dynamic model.



(a) KITTI

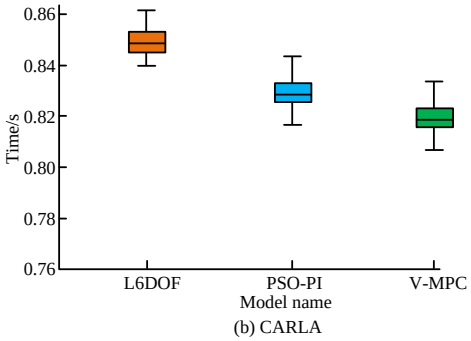


Figure 13: Schematic diagram of comparing the complexity of path tracking control using different algorithms

The study analyzes and compares errors generated by different models on the KITTI dataset. To eliminate accidental errors, the experiment is repeated three times. The results of the error comparison are presented in Figure 14, which shows that the V-MPC model has the lowest path judgment error rate among the three models. Additionally, the V-MPC model had an average error rate of 1.17% across all three experiments, while the PSO-PI model had an average error rate of 1.55%, and the L6DOF model had an average error rate of 1.69%. Based on the error analysis results, it can be concluded that the proposed model has minimal errors and is suitable for high-accuracy automated path selection scenarios.

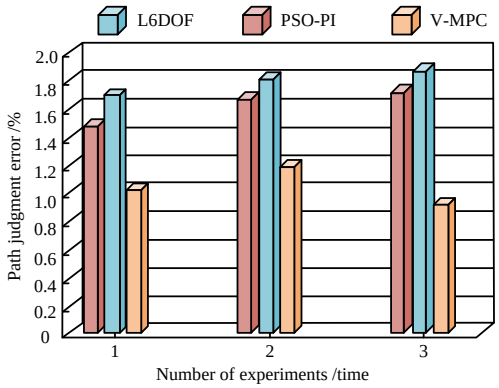


Figure 14: Comparison diagram of error analysis for different models

Based on the experiment results, it is evident that the proposed model has high control accuracy and energy utilization efficiency, making it suitable for scenarios with long driving times and high path accuracy requirements, such as unmanned express delivery. Additionally, the proposed model exhibits strong scalability. The MPC algorithm can introduce dynamic parameter optimization algorithms to accelerate the convergence speed and computational efficiency of the model because it predicts through a system dynamic model.

4 Discussion

In recent years, with the rapid development of the logistics industry, there has been a great deal of interest in free vehicles due to their high efficiency and intelligence. Firstly, in terms of PTC, existing research has predominantly employed algorithms based on deep neural networks and sliding mode control, which have yielded satisfactory tracking outcomes, such as the deep neural network optimization control system proposed by Chen et al [7]. and the fast nonsingular terminal sliding mode control strategy proposed by Sun et al [9]. Secondly, previous studies have demonstrated the efficacy of MPC as a control strategy. This included the multi-layer graph architecture based on MPC proposed by Pan et al [10]. and the load/frequency controller based on MPC proposed by Beus et al [12]. By optimizing the objective function and constraint conditions of the MPC algorithm, the control system can operate in an efficient manner. Finally, existing research has demonstrated the advantages of the ULV algorithm in improving vehicle roll stability and control smoothness. For instance, the integrated control method combining MPC and ULV algorithm proposed by Lin et al [8] has been shown to be effective. However, current research still faces challenges in terms of computational complexity and implementation. Consequently, the research combines the VSC and MPC algorithms and optimizes the 6-degree-of-freedom vehicle dynamics model in order to overcome the shortcomings of existing research and achieve more stable and efficient ULV control. The optimized model has improved significantly in terms of computational complexity, output time, and control accuracy. This is due to the VSC algorithm simplifying the model's structure, which avoids unnecessary calculations. Additionally, the MPC algorithm optimizes the objective and output functions of the model, allowing for accurate analysis and path selection. After introducing the VSC and MPC algorithms, the performance of the linear six-degree-of-freedom vehicle dynamics model has significantly improved. This model can be applied to industries such as unmanned logistics and unmanned food delivery due to its fast output capability, which enhances its adaptability and expands its application space.

5 Conclusion

This research delves into ULV control based on ULV algorithms. High-precision navigation and control of ULVs is achieved through advanced ULV algorithms and dynamics modeling techniques. The technique not only offers a fresh approach to ULV control, but it also establishes the framework for the automated and intelligent creation of next-generation logistics systems. The experimental findings showed that on two distinct datasets, the suggested model performed better than the control model. In terms of path deviation prediction, the average accuracy of the proposed model on the two

datasets was 88.33% and 82.1%, which was 3.96% and 4.72% better than the control model, respectively. This result proved the advantage of the proposed model in terms of prediction accuracy. Also the control accuracy of the proposed model reached 94.19% on the KITTI dataset and 95.61% on the CARLA dataset, both of which are higher than other control models. In addition to the above two metrics, the study also tested other aspects of the model, including energy consumption, controller switching frequency and lateral error. The experimental results verify that the suggested model exhibits a decent level of sophistication and performs comparably across various criteria.

Furthermore, in consideration of the complex and unpredictable nature of real-world logistics scenarios, the study employed a substantial number of data samples to train Jupiter, ensuring that the training data samples encompass the majority of potential scenarios in real-world logistics. Concurrently, in actual unmanned logistics workshops, the range of activities performed by different transfer vehicles is relatively limited. Consequently, the study employs path tracking algorithms to ensure that the model possesses the requisite dynamic and unpredictable capabilities to cope with real-world scenarios following autonomous selection.

6 Ethical and safety considerations

The use of UAVs or drones in modern logistics is becoming increasingly prevalent. As such, it is of paramount importance to ensure that their use is safe and ethical. Research in this field adheres to the highest ethical standards in order to minimize the potential risks that UAVs may face in their autonomous decision-making process. Throughout the entire development process of UAVs, relevant safety protocols have been developed to prevent potential accidents. At the same time, a protection mechanism is set up in UAVs, which can switch to manual control in case of system failure, ensuring safety during transportation. Finally, the study will implement rigorous privacy protection measures in accordance with global data security standards.

Acknowledgement

The research is supported by 2023 Henan University Humanities and Social Sciences Research General Project, Research on the innovative development of urban green logistics distribution under the "Carbon peaking and Carbon neutrality" goals, 2023-ZDJH-108.

References

[1] Z. Sun, Q. Wang, L. Chen, and C. Hu, "Unmanned technology-based civil-military intelligent logistics system: from construction to integration," Journal of Beijing Institute of Technology, vol. 31, no. 2, pp. 140-151, 2022. https://doi.org/10.15918/j.jbit1004-0579.2022.010

[2] A. Kamat, S. Shanker, and A. Barve, "Assessing the factors affecting implementation of unmanned aerial vehicles in Indian humanitarian logistics: a g-DANP approach," Journal of Modelling in Management, vol. 18, no. 2, pp. 416-456, 2023. https://doi.org/10.1108/jm2-02-2021-0037

[3] J. Gu, "Risk prediction of enterprise credit financing using machine learning," Informatica, vol. 46, no. 7, pp. 145, 2022. https://doi.org/10.31449/inf.v46i7.4247

[4] K. L. Choi, M. J. Kim, and Y. M. Kim, "On safety improvement through process establishment for SOTIF application of autonomous driving logistics robot," International Journal of Internet, Broadcasting and Communication, vol. 14, no. 1, pp. 209-218, 2022. https://doi.org/10.7236/IJIBC.2022.14.1.209

[5] G. Iannace, G. Ciaburro, and A. Trematerra, "Acoustical unmanned aerial vehicle detection in indoor scenarios using logistic regression model," Building Acoustics, vol. 28, no. 1, pp. 77-96, 2021. https://doi.org/10.1177/1351010X20917856

[6] P. Hang, and X. Chen, "Path tracking control of 4-wheel-steering autonomous ground vehicles based on linear parameter-varying system with experimental verification," Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, vol. 235, no. 3, pp. 411-423, 2021. https://doi.org/10.1177/0959651820934572

[7] I. M. Chen, and C. Y. Chan, "Deep reinforcement learning based path tracking controller for autonomous vehicle," Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, vol. 235, no. 2-3, pp. 541-551, 2021. https://doi.org/10.1177/0954407020954591

[8] F. Lin, S. Wang, Y. Zhao, and Y. Cai, "Research on autonomous vehicle path tracking control considering roll stability," Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, vol. 235, no. 1, pp. 199-210, 2021. https://doi.org/10.1177/0954407020942006

[9] Z. Sun, J. Zou, D. He, and W. Zhu, "Path-tracking control for autonomous vehicles using double-hidden-layer output feedback neural network fast nonsingular terminal sliding mode," Neural Computing and Applications, vol. 34, no. 7, pp. 5135-5150, 2022. https://doi.org/10.1007/s00521-021-06101-8

[10] Z. Pan, Z. Sun, H. Deng, and D. Li, "A multilayer graph for multiagent formation and trajectory tracking control based on MPC algorithm," IEEE Transactions on Cybernetics, vol. 52, no. 12, pp.

13586-13597, 2021. https://doi.org/10.1109/TCYB.2021.3119330

[11] M. E. Çimen, and Y. Yalçın, "A novel hybrid firefly–whale optimization algorithm and its application to optimization of MPC parameters," Soft Computing, vol. 26, no. 4, pp. 1845-1872, 2022. https://doi.org/10.1007/s00500-021-06441-6

[12] M. Beus, and H. Pandžić, "Practical implementation of a hydro power unit active power regulation based on an mpc algorithm," IEEE Transactions on Energy Conversion, vol. 37, no. 1, pp. 243-253, 2021. https://doi.org/10.1109/TEC.2021.3094059

[13] W. J. Liu, H. F. Ding, M. F. Ge, and X. Y. Yao, "Cooperative control for platoon generation of vehicle-to-vehicle networks: a hierarchical nonlinear MPC algorithm," Nonlinear Dynamics, vol. 108, no. 4, pp. 3561-3578, 2022. https://doi.org/10.1007/s11071-022-07400-y

[14] G. Veselov, A. Tselykh, A. Sharma, and R. H. Huang, "Applications of artificial intelligence in evolution of smart cities and societies," Informatica, vol. 45, no. 5, 2021. https://doi.org/10.31449/inf.v45i5.3600

[15] H. Wang, T. Zhang, X. Zhang, and Q. Li, "Observer-based path tracking controller design for autonomous ground vehicles with input saturation," IEEE/CAA Journal of Automatica Sinica, vol. 10, no. 3, pp. 749-761, 2023. https://doi.org/10.1109/JAS.2023.123078

[16] N. Wang, Y. Zhang, C. K. Ahn, and Q. Xu, "Autonomous pilot of unmanned surface vehicles: Bridging path planning and tracking," IEEE Transactions on Vehicular Technology, vol. 71, no. 3, pp. 2358-2374, 2021. https://doi.org/10.1109/TVT.2021.3136670

[17] S. Feng, Y. Qian, and Y. Wang, "Collision avoidance method of autonomous vehicle based on improved artificial potential field algorithm," Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, vol. 235, no. 14, pp. 3416-3430, 2021. https://doi.org/10.1177/09544070211014319

[18] M. Labbadi, and M. Cherkaoui, "Adaptive fractional-order nonsingular fast terminal sliding mode based robust tracking control of quadrotor UAV with Gaussian random disturbances and uncertainties," IEEE Transactions on Aerospace and Electronic Systems, vol. 57, no. 4, pp. 2265-2277, 2021. https://doi.org/10.1109/TAES.2021.3053109

[19] S. Wallace, and G. Lăzăroiu, "Predictive control algorithms, real-world connected vehicle data, and smart mobility technologies in intelligent transportation planning and engineering," Contemporary Readings in Law and Social Justice, vol. 13, no. 2, pp. 79-92, 2021. https://doi.org/10.22381/CRLSJ13220216

[20] A. Williams, "Human-centric functional computing as an approach to human-like computation," Artificial Intelligence and Applications, vol. 1, no. 2, pp. 118-137, 2023. https://doi.org/10.47852/bonviewAIA2202331

[21] G. Bandewad, K. P. Datta, B. W. Gawali, and S. N. Pawar, "Review on discrimination of hazardous gases by smart sensing technology," Artificial Intelligence and Applications, vol. 1, no. 2, pp. 86-97, 2023. https://doi.org/10.47852/bonviewAIA3202434

[22] J. Jin, and J. Gong, "An interference-tolerant fast convergence zeroing neural network for dynamic matrix inversion and its application to mobile manipulator path tracking," Alexandria Engineering Journal, vol. 60, no. 1, pp. 659-669, 2021. https://doi.org/10.1016/j.aej.2020.09.059

[23] P. M. Marusak, "A numerically efficient fuzzy MPC algorithm with fast generation of the control signal," International Journal of Applied Mathematics and Computer Science, vol. 31, no. 1, pp. 59-71, 2021. https://doi.org/10.34768/amcs-2021-0005

[24] D. Wang, Q. Chen, X. Zhang, C. Gao, B. Wang, X. Huang, and J. Crittenden, "Multipollutant control (MPC) of flue gas from stationary sources using SCR technology: A critical review," Environmental Science & Technology, vol. 55, no. 5, pp. 2743-2766, 2021. https://doi.org/10.1021/acs.est.0c07326

Edge Computing Based Multi-Objective Task Scheduling Strategy for UAV with Limited Airborne Resources

Xiaoqiang Wang
Chinese-German College of Engineering, Shanghai Technical Institute of Electronics & Information, Shanghai, 201411, China
Email of corresponding author: wxqfyc@163.com

Keywords: unmanned aerial vehicle, edge computing, limited airborne capacity, multi-objective, task scheduling

Received: March 11, 2024

The unmanned aerial vehicles often suffer from insufficient computing power due to the limited onboard resources, resulting in task delays under heavy tasks. A system based on edge computing was constructed to solve this problem, which involved task allocation center, unmanned aerial vehicle group, data node, and power supply station. A mathematical optimization framework based on task, resource, and scheduling models was proposed, and the non-dominated sorting genetic algorithm III was used. The objective optimization was efficiently processed through genetic operations, non-dominated sorting, and reference point-based selection mechanisms. These results confirmed that the non-dominated sorting genetic algorithm III performed well in comprehensive performance evaluation, with an MS index of 0.881 in large-scale map tests and an AQ index of 0.133 in medium-sized maps. The calculation time was 58.9 seconds, 140.5 seconds, and 545.3 seconds in small, medium, and large map tests, respectively, leading other algorithms. Therefore, the designed model has excellent performance in task quality, time extension, and computational efficiency, which has application value.

Povzetek: Študija uporablja sistem temelječ na robnem računalništvu, za načrtovanje večnivojskih nalog za brezpilotna letala z omejenimi viri je uveden sistem, temelječ na robnem računalništvu, ki dosega visoko točnost in učinkovitost pri razporejanju nalog.

1 Introduction

In the era of highly developed technology, Unmanned Aerial Vehicle (UAV) has become an important research field, with applications ranging from military, investigation, daily delivery to ecological research, and more. The onboard resource management of the UAV operation becomes an important technical challenge. This includes many aspects of UAV power management, load scheduling, flight path design, etc. [1-3]. Especially, how to effectively carry out task scheduling to ensure optimal operational efficiency and task completion quality becomes an important research topic under limited airborne resource conditions. In task scheduling, UAV needs to ensure optimal allocation of system resources while executing tasks to achieve maximum work efficiency [4-6]. With the development of big data and cloud computing, edge computing becomes a hot research field in recent years. Edge computing can solve the high data transmission delay, data loss, and security in cloud computing. Edge computing is an important technical strategy to support efficient operation of UAV in complex environments. Previous studies have mostly focused on single task scheduling strategies, with less attention paid to multi-task scheduling problems under limited resource conditions [7-9]. Therefore, this study aims to analyze the edge computing-based multi-target task scheduling strategy under the condition of limited UAV airborne resources. It is hoped to provide new theoretical support and practical reference for the task scheduling strategy of UAV. The research mainly includes four parts. Firstly, the research objective is proposed. Then, a multi-objective task scheduling strategy for UAV is designed. Next, model validation is conducted. Finally, a conclusion is drawn.

2 Research background

As UAV continues to evolve, the task scheduling research gradually deepens. You W et al. designed an optimization model aimed at minimizing the total energy consumption of user UAV. Meanwhile, an iterative algorithm using block coordinate descent method was proposed, which had high efficiency [10]. Halder et al. proposed a novel clustering method that enabled UAV to achieve dynamic task scheduling. The throughput optimality of its scheduling algorithm was determined through the Lyapunov drift analysis framework. These experiments confirmed that the proposed method surpassed existing solutions in terms of energy consumption, cluster overhead, throughput, end-to-end latency, flow success rate, and packet loss rate [11]. Niu Z et al. discussed how to better utilize UAV for task

scheduling in disaster scenarios. They suggested using a decentralized computing network consisting of UAV and ground mobile devices. These results confirmed that their algorithm reduced the energy consumption of the entire system by more than 50% while ensuring queue stability [12]. Wang Y et al. proposed a mixed integer nonlinear programming model. Meanwhile, an alternating optimization algorithm was proposed based on differential evolution and greedy Hongya algorithm to obtain suboptimal solutions. These experiments confirmed that the total benefit of this scheme was approximately 50% higher than existing methods [13].

On the other hand, the application of Non-dominated Sorting Genetic Algorithm III (NSGA-III) gradually diversifies and deepens. Khettabi I et al. used the new dynamic Non-dominated Sorting Genetic Algorithm II (NSGA-II) and the new NSGA-III for evaluation. The effectiveness of the proposed method was demonstrated through three measurement criteria. Finally, TOPSIS was used to assist decision-makers in evaluating and selecting the best process plan [14]. Awad M et al. aimed to use NSGA-II and its extended version to solve single objective and Multi-objective Optimization (MOO) problems and find Pareto optimal solutions in stock investment portfolio management. These experiments confirmed that NSGA-II was effective for portfolio problems with two objective functions, while NSGA-III was effective for problems with three objective functions [15]. Johnson N N et al. investigated the influence of welding parameters on solder joint quality and combined Kriging and the new NSGA-III for MOO of RSW. These experiments confirmed that optimized welding parameters effectively improved the welding quality. When using the optimized welding parameters, the concave diameter, tensile shear strength, and peel strength of the welded specimens increased by about 9.21%, 4.95%, and 7.69%, respectively [16]. Harif S et al. proposed four conflicting optimization objectives using NSGA-III to improve the ideal position of sensors. These experiments confirmed that as the sensors increased, the Pareto front became more effective [17]. The literature summary is shown in Table 1.

Table 1: Literature summary

Author	Research dimensions	Research contents	Research conclusion	Literature
You et al.	UAV task scheduling	Designed an optimization model aimed at minimizing the total energy consumption of user UAV	This method has higher efficiency compared to existing solutions	[10]
Halder et al.		Proposed a clustering method for UAV dynamic task scheduling	This method performs excellently in terms of energy consumption, cluster overhead, throughput, end-to-end latency, and packet loss rate	[11]
Niu et al.		Analyzed UAV task scheduling methods in disaster scenarios	The model reduces the energy consumption of the entire computing system by more than 50% while ensuring queue stability	[12]
Wang et al.		A mixed integer nonlinear programming model was proposed.	The overall return of this plan is 50% higher than that of existing methods	[13]
Khettabi et al.	Application of NSGA-III model	Proposed evaluation methods for new dynamic NSGA-II and new NSGA-III	The effectiveness of the method is demonstrated through three measurement standards	[14]

Awad et al.	Expanded and applied the NSGA-II model to solve single objective and multi-objective optimization problems.	Solved the combination problem of two objective functions and three objective functions	[15]
Johnson et al.	Multi-objective optimization of RSW using Kriging and the new NSGA-III model	After using optimized welding parameters, the concave diameter, tensile shear strength, and peel strength of the welded samples all increased.	[16]
Harif et al.	Proposed four optimization objective functions based on NSGA-III to improve the placement of sensors	Obtained better monitoring results	[17]

In conclusion, many researchers focus on the task scheduling strategy on UAV, mobile edge computing and solving optimization problems. Some researchers have designed optimization problems to reduce the total energy consumption of UAV by constructing layered systems. Some have explored how to use UAV for task scheduling in disaster scenarios. Some scholars have also emphasized the role of UAV in dynamic task scheduling, optimal UAV deployment, and mobile device location determination. Meanwhile, the application of NSGA-III has also been promoted in many fields, such as reconfigurable manufacturing systems, portfolio management, manufacturing, and so on. This study further considers the UAV task allocation that UAV may face under limited loads on the basis of the existing research mentioned above, providing a new approach for this field. This study not only solves the limited airborne problem, but also solves the task scheduling problem under limited computing power. A comprehensive model performance evaluation is conducted, providing a more efficient solution for real-time task scheduling under limited resources.

3 Design of unmanned aerial vehicle multi-objective task scheduling strategy

A MOO-UAV scheduling mathematical model is proposed to address the computational power shortage and task delay caused under limited airborne conditions. This model consists of tasks, resources, and scheduling models and is optimized and solved using NSGA-III.

3.1 Architecture of task scheduling model

The MOO-UAV system is built on the edge computing framework. This system mainly includes the following main components: task allocation center, UAV group, data node, and power supply station. The MOO-UAV system also includes some abstract elements, including the study of expected optimization goals, various constrainable conditions, and strategy algorithms for solving problems. Figure 1 shows this system.

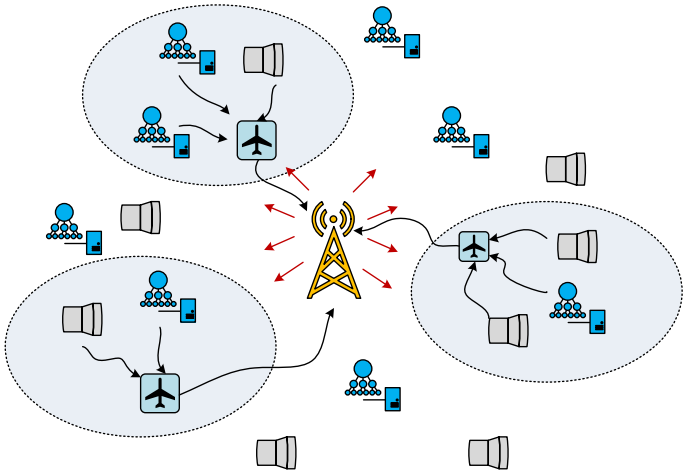


Figure 1: System model

The control center is responsible for supervising all UAV and assigning tasks to them in real-world physical elements. When a data node generates a request, the control center will analyze information such as node location and data size based on this. Then tasks are assigned to each UAV. The control center also serves as a power supply station, as well as takeoff and landing points. UAV is the carrier and processor of data tasks. Each UAV is equipped with a complete set of equipment such as flight control system, wireless communication system, navigation system, etc. The equipment is used for data processing and information transmission during navigation. The data node is responsible for providing data tasks and transmitting tasks to the UAV during hovering. The power supply station is responsible for supplying UAV energy. In non-physical elements, the optimization objective is to optimize multiple objectives. As a result, the total range of all UAV, the total completion time of data tasks, and the total duration of all overtime tasks can be minimized. Attention should be paid to constraints such as UAV flight speed, power consumption, and range during the solving process. It is crucial to choose the appropriate optimization algorithm for this issue. Figure 2 shows the complete workflow.

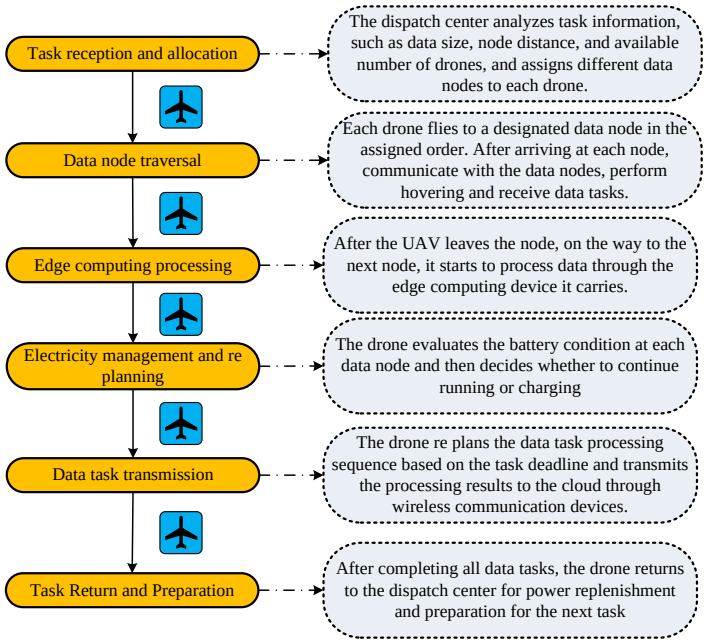


Figure 2: Workflow

The dispatch center will allocate tasks based on various factors such as task information and the number of UAV after receiving a batch of pending data tasks. The UAV traverses' nodes in the order assigned, hovers over

each node and receives tasks, and then processes data tasks while flying to the next node. If the battery is low, UAV needs to go to the power supply station to replace the battery before continuing the task. UAV returns to the dispatch center to prepare for the next task after completing the task. It is necessary to optimize the route of UAV to minimize the task completion time and voyage during this process, while completing as many tasks as possible.

3.2 Construction of task scheduling operation model

The biggest challenge when scheduling UAV is to find a solution that is suitable for assigning UAV to a large amount of data tasks. Therefore, the timely completion of all tasks can be ensured, and the scheduling cost of UAV can be reduced as much as possible. Therefore, a redesigned MOO-UAV scheduling mathematical model is adopted in this study. It mainly consists of tasks, resources, and scheduling models.

Firstly, a task model is designed, and a scheduling hub is established as the initial stopping point for all UAV, with many data nodes. Each data node has its own unique task data volume. The data scale for each node is generated through normal distribution to simulate the real-world environment. Then, a resource model is established, including all UAV and edge computing devices. All UAV stay at the dispatch hub at the beginning. The flight speed, data transmission speed, processing speed, and total endurance time of UAV are all known. The endurance time is a key parameter as it determines the number of tasks that a UAV can perform at once. Then, a scheduling model is studied and constructed. Charging stations are set up in the data node area considering the endurance capability of UAV in this model. Meanwhile, UAV generates specific flight paths based on task sequences and base station positions. Then constraints are defined to ensure that the remaining endurance time of UAV at the data node cannot be negative. UAV must have sufficient energy to return to the dispatch center or transfer to the charging station. The optimization objective of this study is to minimize the completion time of data tasks, the total flight path of all UAV, and the total duration of all timeout tasks. Figure 3 shows task scheduling assumptions and constraints.

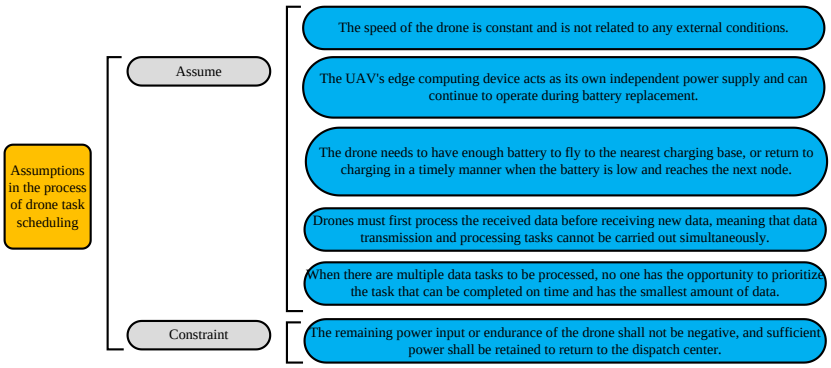


Figure 3: Task scheduling assumptions and constraints

It is assumed that formula (1) is the flight path of UAV.

$$L_j = \left\{ l_{j,1}, l_{j,1}, \cdots, l_{j,\tau} \right\} \tag{1}$$

In formula (1), τ refers to the number of destinations. The constraints of the model are represented by formula (2).

$$t_{j,\tau}^{r,e} \geq \frac{D\left(l_{j,q}, B_{j,q}\right)}{v} + h_{j,\tau} \tag{2}$$

In formula (2), $l_{j,q}$ refers to the destination of the UAV. v refers to the speed. j refers to the number of UAV. $B_{j,q}$ refers to the nearest charging station. $h_{j,\tau}$ refers to the estimated electricity consumption from the node to the dispatch center. D refers to the distance. Further, formula (3) can be obtained.

$$t_{j,\tau}^{r,e} = \begin{cases} t^{\max} \\ t_{j,\tau-1}^{r,e} - h_{j,\tau-1} - \dfrac{D\left(l_{j,q}, B_{j,q-1}\right)}{v} \end{cases} \tag{3}$$

In formula (3), $t^{\max}$ refers to the maximum endurance time. Formula (4) is used for constraints due to the fact that the ultimate destination of the UAV is a data node, which is not a charging station.

$$N_j \notin B \tag{4}$$

In formula (4), B refers to the charging station. The study adopts MOO methods suitable for three objectives and higher to evaluate the final benefits. Therefore, the flight path of UAV, dwell time at all nodes,

and other data processing time can be minimized after the end of the last data task. The completion time of UAV tasks is represented by formula (5).

$$t_j^{fin} = \sum_{q=2}^{N_j} \left(\frac{D\left(l_{j,q-1}, B_{j,q}\right)}{v} + h_{j,q} \right) + \sum_{\rho=1} Z_{N_{j,\tau,\rho}} \quad (5)$$

In formula (5), N_j refers to the total destinations. $D\left(l_{j,q-1}, B_{j,q}\right)$ refers to the distance between different destinations. $h_{j,q}$ refers to the hover time. $Z_{N_{j,\tau,\rho}}$ refers to the pending task for the last data point. The objective function is represented by formula (6).

$$f(x) = \min\left[f_1(x), f_2(x), f_3(x) \right]^T \quad (6)$$

In formula (6), $f_1(x)$ refers to the highest completion time. $f_2(x)$ refers to the total flight distance. $f_3(x)$ refers to the duration of task timeout. $f_1(x)$ is represented by formula (7).

$$f_1(x) = \max\left[t_1^{fin}, t_2^{fin}, \cdots, t_n^{fin} \right] \quad (7)$$

In formula (7), t_n^{fin} refers to the completion time of the task. $f_2(x)$ is represented by formula (8).

$$f_2(x) = \sum_{j=1}^{n} \sum_{q=2}^{N_j} D\left(l_{j,q-1}, l_{j,q}\right) \quad (8)$$

$f_3(x)$ is represented by formula (9).

$$f_3(x) = \sum_{i=1}^{m} t_i^{out} \quad (9)$$

In formula (9), t_i^{out} refers to the timeout time.

3.3 Design of multi-objective solving strategies

The MOO algorithm demonstrates strong superiority in handling single objective and multi-objective problems. However, a more powerful MOO tool, namely NSGA-III, is required for increasingly common problems involving three objectives or MOO. NSGA-III is a multi-objective evolutionary algorithm based on reference points, which emphasizes non-dominant population members more than NSGA-II. This method also adds a set of individual related reference points to select the next population individual. As a result, better convergence and diversity can be achieved to find non-dominant solutions. Figure 4 shows the process of the model.

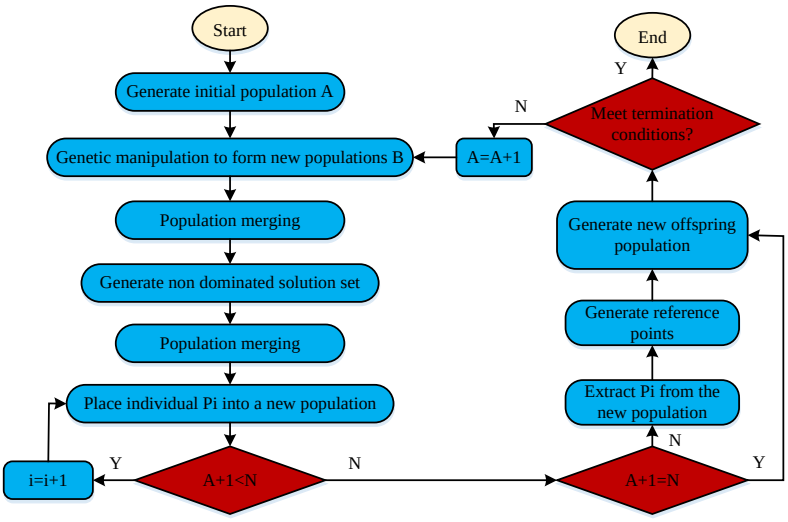


Figure 4: Model process

The execution of NSGA-III requires generating an initial solution using the roulette wheel algorithm, which is the initial population P0 of size H. Then NSGA-III begins genetic operations on the initial solution population, including mutation, crossover, and individual selection, to generate an iterative population algebra gmax. Genetic operations, non-dominated sorting operations, reference point generation, and the association between individuals and reference points are important components of NSGA-III. Genetic

manipulation involves two types of mutations and one crossover approach. Genetic manipulation aims at maintaining genetic diversity while preserving excellent gene fragments. Figure 5 shows the mutation and crossover processes.

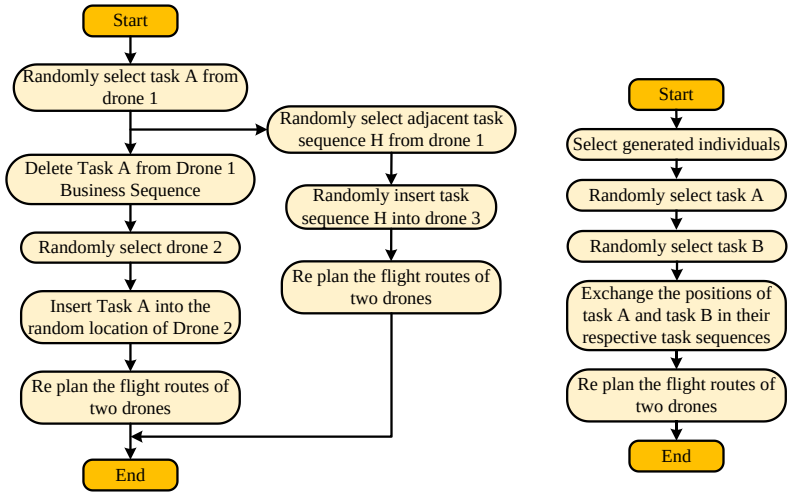


Figure 5: Variations and cross processes

This step preserves certain UAV flight sequences of outstanding individuals and adds them to the sequences of other UAV. Meanwhile, it is crucial to determine the reference point in NSGA-III. The generated new population and parent population form a mixed population after mutation and crossover. All individuals are added to the offspring population after non-dominated sorting is performed on the mixed population to generate multiple non-dominated layers. This step stops until the offspring size exceeds half of the mixed population size. In this process, the study does not use congestion ranking in NSGA-II, but instead uses a reference point-based method for ranking. The association between individuals and reference points is a crucial step in forming a new population. This solution is to associate all individuals with a reference point, starting from the ideal point and extending infinitely to all reference points. Each line represents a reference line corresponding to a reference point. Figure 6 is a schematic diagram of the reference line.

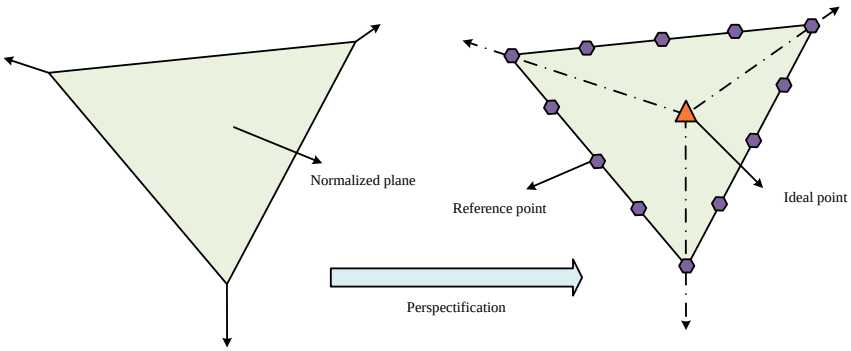


Figure 6: Reference line diagram

Then the closest distance between each individual and all reference lines is calculated. The closest reference point is the individual's associated reference point. Sometimes, a reference point may be associated with one or more individuals, or there may be no individual associated with it. The total reference points in the target problem are represented by formula (10).

$$H = \begin{pmatrix} M + p - 1 \\ p \end{pmatrix} \tag{10}$$

In formula (10), M represents the number of objective problems. If a three-objective problem is adopted, M is 3. p represents population. The

research prioritizes selecting reference points with fewer associated individuals to ensure the diversity of individual populations. Meanwhile, the associated individuals are added to the offspring population. The minimum value of the total duration is represented by formula (11).

$$MIN = \begin{cases} z_1^{\min} \\ z_2^{\min} \\ z_3^{\min} \end{cases} \quad (11)$$

In formula (11), $z_1^{\min}$ represents the maximum and minimum completion time. $z_2^{\min}$ represents the shortest flight length. $z_3^{\min}$ represents the total timeout duration. Therefore, the reference plane zero point in formula (12) can be formed.

$$Z_0 = \left(z_1^{\min}, z_2^{\min}, z_3^{\min} \right) \quad (12)$$

When the population is adaptively normalized, the maximum objective value in the three objective problem is normalized and represented by formula (13).

$$f_i'(x) = f_i(x) / z_i^{\max}, (i = 1,2,3) \quad (13)$$

In formula (13), $z_i^{\max}$ represents the maximum target value. The individual target dimensions are fixed and the extremum is determined, represented by formula (14).

$$ASF(x,\omega) = \max_{i=1}^{M} f_i'(x) / \omega_i, x \in S_x, \omega = (\tau, \cdots, \tau, 1, \tau, \cdots, \tau), \tau = 10^{-6} \quad (14)$$

After determining the maximum values of all target dimensions, the minimum value is selected as the extremum point, represented by formula (15).

$$Z_i^{\max} = \min\left\{ ASF(x_1, w_i), ASF(x_2, w_i), ASF(x_3, w_i) \right\} \quad (15)$$

The individuals are associated with the reference point after normalizing all individuals with the reference point. Priority is given to selecting reference points with fewer associated individuals in the set to add to the offspring population, thereby improving individual diversity of the population.

4 Verification of the effectiveness of traffic flow prediction models

The study first conducted parameter analysis when testing the effectiveness of traffic flow prediction models. Afterwards, an overall model efficiency analysis was conducted. Four multi-objective indicators: Average Quality (Aq), Maximum Spread (Ms), Maximum Distance (Md), and Average Distance (Ad) were used.

4.1 Parameter analysis

In this experiment, the UAV underwent three different configuration settings. Firstly, the study set constants such as UAV endurance time and flight speed. Table 2 shows the experimental setup.

Table 2: Experimental setup

	Drone parameters	Option 1	Option 2	Option 3
Drone settings	Battery life (seconds)	800	1000	1200
	Flight speed (meters/second)	20	30	40
	Data transmission speed (minutes)	Θ*0.20	Θ*0.15	Θ*0.10
	Data processing speed (minutes)	Θ*0.50	Θ*0.33	Θ*0.25
Map node settings	Map type		Aponym	Number of nodes
	Small-sized		att48	48
	Medium-sized		bier127	127
	Large-sized		att532	532
Map Type Settings	Map type		Describe	
	Middle		The logistics dispatch center is located in the center of the map	
	Side		The logistics dispatch center is located on the edge of the map	
	Inside		The logistics dispatch center is located at a random location within	

	the map

In Table 2, three maps of different sizes and distinguishes map features were constructed. Figure 7 shows the analysis of A-type and V-type parameters under Aq and Ms.

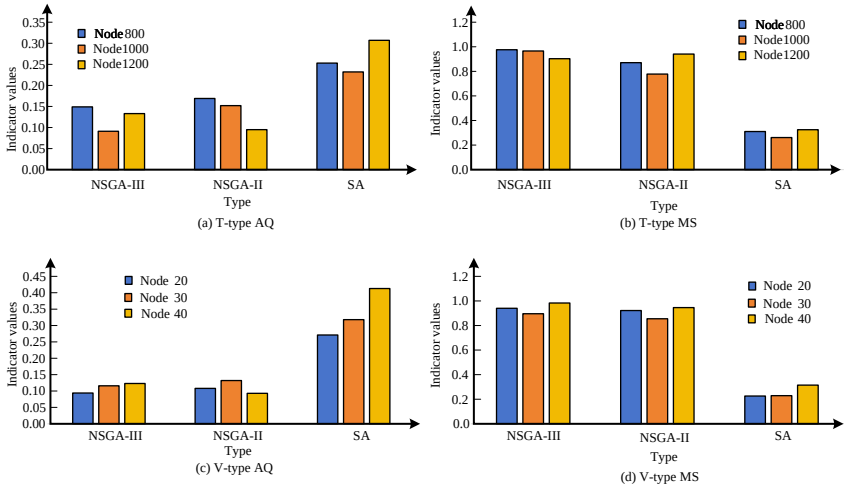


Figure 7: Analysis of A-type and V-type parameters under AQ and MS

In Figure 7, the comparison models are NSGA-II and Improved Simulated Annealing Algorithm (SA), respectively. The study analyzed the effectiveness of UAV task allocation models based on the provided data. The study focused on comparing the performance of different algorithms under specific multi-objective indicators. The indicators considered include Aq and Ms, while the analyzed algorithms are NSGA-III, NSGA-II, and SA. When the nodes were 800, NSGA-III showed a high efficiency of 0.149 on Aq through the analysis of T-shaped parameters, which was better than NSGA-II's 0.169 and SA's 0.253. Therefore, NSGA-III had better processing ability for UAV task allocation. On MS, NSGA-III was 0.976, leading NSGA-II at 0.871 and SA at 0.31. This trend was also maintained when the parameters increased to 1000, with AQ and MS of NSGA-III being 0.091 and 0.966, respectively, while NSGA-II and SA performed less well. In Table 3, the remaining indicators under Aq and Ms were compared.

Table 3: Comparison of remaining indicators under AQ and MS

Parameter types	Nodes	Aq			Ms		
		NSGA-III	NSGA-II	SA	NSGA-III	NSGA-II	SA
F-type parameter	Middle piece	0.1	0.146	0.307	0.974	0.813	0.263
	Broadside	0.089	0.135	0.299	0.964	0.938	0.314
	Internal	0.124	0.141	0.221	0.906	0.873	0.348
Map parameters	48	0.126	0.103	0.333	0.92	0.921	0.336
	127	0.133	0.096	0.411	0.902	0.937	0.303
	532	0.137	0.17	0.358	0.881	0.87	0.299
T-trans	0.2	0.082	0.166	0.41	0.925	0.93	0.224
	0.15	0.126	0.162	0.336	0.943	0.948	0.3
	0.1	0.105	0.126	0.28	0.848	0.925	0.326
T-pro	0.5	0.145	0.15	0.322	0.875	0.883	0.241
	0.33	0.131	0.116	0.312	0.9	0.888	0.191
	0.25	0.114	0.126	0.293	0.934	0.97	0.266

Merge average	-	0.119	0.133	0.316	0.926	0.901	0.283

In Table 3, NSGA-III performed better with an Aq of 0.1 and a MS of 0.974 for the F-type parameter at the middle node, which were higher than NSGA-II and SA. This indicated that NSGA-III was superior in balancing quality and scalability when dealing with intermediate node tasks. The analysis of side nodes and internal nodes also showed a similar trend, with the Aq of NSGA-III for side nodes being 0.089 and MS being 0.964, both in a leading position. In the evaluation of map parameters, an increase of nodes had a negative impact on the Aq of NSGA-II and SA. However, NSGA-III maintained relatively stable performance, with Aq values gradually increasing from 0.126 to 0.137. This indicated the better robustness of NSGA-II compared to other algorithms. The MS of NSGA-III was 0.881, especially when the nodes were 532, which was better than NSGA-II and SA. In addition, as the parameter values decreased from 0.2 to 0.1, the Aq performance of NSGA-III remained stable, ranging from 0.082 to 0.105, while MS increased from 0.41 to 0.848. This reflected the algorithmic adaptability to various task scheduling scenarios. When the parameter value decreased from 0.5 to 0.25, NSGA-III also exhibited more robust Aq performance compared to NSGA-II and SA for the T-pro parameter category, with a decrease from 0.145 to 0.114. NSGA-III also maintained an advantage in Ms, rising from 0.875 to 0.934. NSGA-III exhibited overall better performance based on comprehensive analysis of all parameter categories. Its comprehensive average Aq was 0.119, which was significantly lower than NSGA-II's 0.133 and SA's 0.316. The comprehensive average of MS was 0.926, which was the highest among all algorithms. Table 4 shows the comparison of indicators under Davg and Dmax.

Table 4: Comparison of indicators under Davg and Dmax

Parameter types	Nodes	Md			Ad		
		NSGA-III	NSGA-II	SA	NSGA-III	NSGA-II	SA
F-type parameter	Middle piece	0.018	0.015	0.539	0.099	0.066	0.654
	Broadside	0.009	0.016	0.72	0.104	0.083	0.663
	Internal	0.012	0.022	0.553	0.115	0.148	0.732
T-trans	0.21	0.01	0.021	0.688	0.138	0.078	0.857
	0.16	0.009	0.015	0.663	0.098	0.091	0.783
	0.11	0.012	0.018	0.585	0.089	0.123	0.834
T-pro	0.51	0.009	0.022	0.563	0.133	0.148	0.874
	0.34	0.009	0.015	0.611	0.106	0.132	0.866
	0.26	0.008	0.024	0.742	0.085	0.123	0.823
V-shaped	21	0.008	0.022	0.536	0.135	0.064	0.698
	31	0.013	0.015	0.672	0.117	0.089	0.713
	41	0.015	0.026	0.597	0.133	0.083	0.765
T-shaped	805	0.027	0.021	0.361	0.141	0.083	0.791
	1005	0.017	0.016	0.572	0.084	0.096	0.816
	1205	0.009	0.024	0.664	0.124	0.134	0.893
Map parameters	49	0.01	0.017	0.713	0.096	0.073	0.856
	128	0.015	0.021	0.627	0.133	0.138	0.71
	533	0.021	0.022	0.558	0.132	0.12	0.726
Merge average	-	0.013	0.02	0.609	0.115	0.104	0.781

In Table 4, NSGA-III performed better for F-type parameters than the other two algorithms in terms of Ad index. The values of NSGA-III were 0.018, 0.009, and 0.012, respectively, whether in the middle, side, or interior, which were much lower than SA's 0.539, 0.72, and 0.553. This indicated that the average delay of NSGA-III was lower, and the task allocation efficiency of UAV was higher under F-type parameters. NSGA-III performed for the T-trans parameter better on Ad than the

other two algorithms, whether the node was 0.21, 0.16, or 0.11. The values of NSGA-III were 0.01, 0.009, 0.012, and the SA values were 0.021, 0.015, and 0.018, respectively. However, the performance of these three algorithms was almost the same in terms of Md, with only slightly higher SA, at 0.857, 0.783, and 0.834, respectively. The situation was different for T-pro. Although NSGA-III still outperformed the other two algorithms on Davg, the values of SA on Md were 0.874, 0.866, and 0.823, all higher than those of NSGA-III. This indicated that SA had a higher maximum delay under T-pro, and the efficiency of UAV in executing tasks needed to be improved. NSGA-III still showed a leading advantage in Ad for the V-shaped parameter group, with values of 0.008, 0.013, 0.015, and SA values of 0.536, 0.672, and 0.597. However, on Md, SA exhibited higher latency. Whether the node was 805, 1005, or 1205, NSGA-III had a better value on Ad than SA for the T-type parameter group. However, SA had a higher value than NSGA-III on Md. Finally, whether the node was 49, 128 or 533, NSGA-III had better values on Ad than SA for the map parameter group. SA had higher values than NSGA-III on Md. The NSGA-III UAV multi-objective task allocation model performs better than the other two models in most scenarios and parameters through the above comparison, with particularly outstanding performance in average latency.

4.2 Efficiency comparison of models

Six different optimization models were compared, including Improved Ant Colony Optimization (IAC), Improved Genetic Algorithm (IGA), Improved Particle Swarm Optimization (IPSO), and Improved Deep Learning Optimization (IDLO). The difference in computation time for completing map data tasks of the same scale is shown in Figure 8.

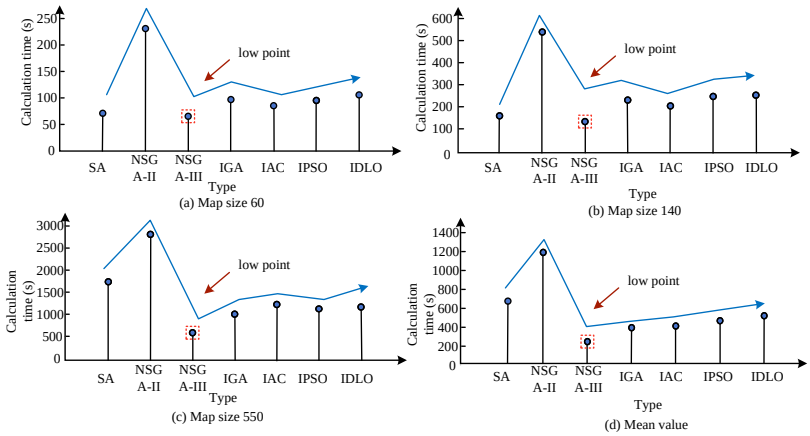


Figure 8: Calculation time difference

In Figure 8, NSGA-III showed a significant efficiency advantage among all algorithms, with the lowest average computation time across all map sizes, only 248.2 seconds. NSGA-III exhibited significant computational efficiency compared to other algorithms. The computation time of NSGA-III was 58.9 seconds on the minimum map size. The closest IAC and SA had a computation time of 70.2 and 85.3 seconds, respectively. The time for NSGA-III was 140.5 seconds on a medium map size, while the time for other algorithms ranged from 158.9 to 252.5 seconds. NSGA-III maintained the shortest computation time of 545.3 seconds on the maximum map size. NSGA-III significantly reduced computation time by more than 40% compared to SA's 950.5 seconds for the second shortest time. On average, NSGA-III was 192.6, 165.0, 230.3, and 254.8 seconds faster than IGA, SA, IPSO, and IDLO, respectively. This obvious time-saving effect revealed that NSGA-III's algorithm efficiency was higher when facing complex task scheduling with multiple objectives. Figure 9 shows a comparison of task execution times.

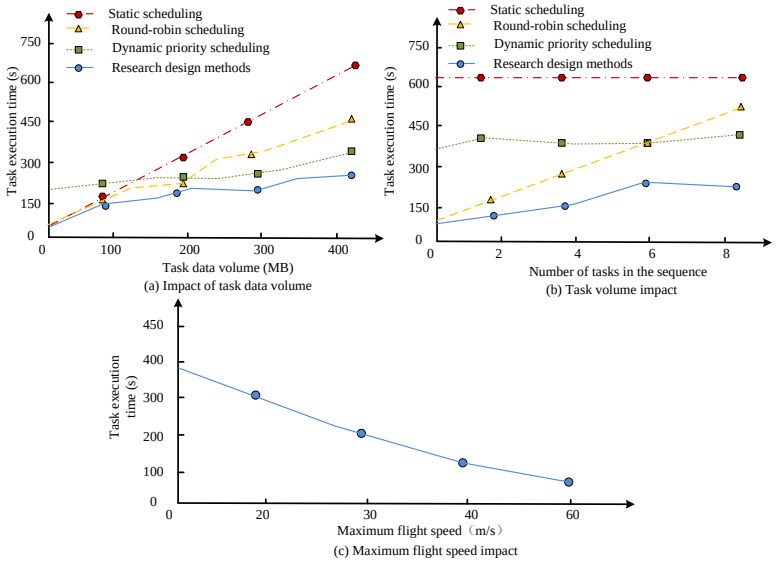


Figure 9: Comparison of task execution time

In Figure 9 (a), the design method always had the shortest execution time under changes in task data volume. From Figure 9 (b), the research design method dynamically adjusted the execution time under changes in task volume. This method maintained a minimum and ultimately decreased, indicating optimization of task execution sequence. From Figure 9 (c), the faster the flight time, the faster the task completion speed.

Table 5: Comparison of average operation time (s)

Optimization model	Minimum map size	Medium map size	Maximum map size	Average calculation time
NSGA-III	55.0	130.4	500.0	228.4
IDRLO	58.9	140.5	545.3	248.2
ENNO	60.5	125.7	540.2	242.1

From Table 5, the NSGA-III model had the shortest average computation time and was the most superior. The superiority of this computation time was reflected in large, medium, and small maps.

5 Discussion

This study is based on the NSGA-III algorithm and focuses on the task configuration and path planning problems of UAV under limited resource conditions. In the results, the processing efficiency and scheduling effect of multi-objective scheduling problems were significantly improved under limited airborne conditions of UAV. In similar fields, the Thomas T team solved the routing and scheduling problems of single truck multi-UAV delivery systems using mixed integer linear programming and RF-RRO heuristic methods, mainly optimizing delivery time and cost [18]. Sun F et al. applied the dragonfly algorithm for UAV task scheduling in agricultural plant protection environments, focusing on the timeliness of task execution [19]. The Ms index of this study reached 0.881 in large-scale map testing, and the performance of the model in practical applications was better than NSGA-II and SA. Meanwhile, the designed model had a shorter computation time, demonstrating higher computational efficiency. Meanwhile, the computational efficiency of the research model on large-scale maps is excellent. Therefore, this designed model had advantages in computational efficiency and MOO. This method not only improved the efficiency and quality of UAV scheduling systems, but also provided a new perspective for solving similar problems.

This study provides effective solutions for the application of UAV systems in commercial logistics, disaster response and rescue, environmental monitoring, and agricultural management. Therefore, flight path planning, task allocation, and charging station management can be addressed. This study not only improves the efficiency of drone scheduling, but also significantly improves computational efficiency. Therefore, UVA can minimize resource consumption and flight risks while ensuring task execution under limited onboard conditions. This study can be used for UAV disaster assessment, wounded search, and emergency supplies transportation in natural disaster rescue operations. The research method can be used for environmental monitoring and precision agriculture

monitoring, allowing operators to reduce operating costs and collect data more efficiently. In UVA warfare, this method can improve the efficiency of drone military reconnaissance, surveillance, and strike. Then task execution speed and accuracy efficiency can be improved in complex and resource scarce environments. Meanwhile, the real-time response capability of combat units can be enhanced. In terms of scalability, in the future, injection task allocation, path planning, charging management, and other functions can be modularized. At the same time, more improvements can be made for special environments to enhance the overall flexibility of the system.

6 Conclusion

The study proposed an integrated mathematical model framework for more efficient planning of UAV flight paths, task allocation, and management of charging stations. This framework adopted NSGA-III to implement scheduling strategies based on factors such as task characteristics, UAV resource capabilities, map size, and flight parameters. These results confirmed that the designed model exhibited superiority in multiple key performance indicators. In large maps, NSGA-III achieved a Ms index of 0.881, higher than NSGA-II and SA. On a medium-sized map, its Aq index reached 0.133. In addition, the calculation time of this research model on small maps was 58.9 seconds, which was lower than other models. On a medium-sized map, its calculation time was only 140.5 seconds, compared to an IGA of 158.9 seconds. Specifically, NSGA-III had a computation time of 545.3 seconds on large maps, which had an advantage in computational efficiency. In summary, the research model not only shows significant efficiency in dealing with UAV multi-objective task scheduling problems, but also performs equally well in scheduling quality and time management.

References

[1] H. Sun, B. Zhang, X. Zhang, Yu. Ying, K. W. Sha, and W. S. Shi, "FlexEdge: Dynamic task scheduling for a UAV-Based on-demand mobile edge server," IEEE Internet of Things Journal, vol. 9, no. 17, pp. 15983-16005, 2022. https://doi.org/10.1109/JIOT.2022.3152447

[2] Z. Liu, Y. Cao, P. Gao, X. H. Hua, D. C. Zhang, and T. Jiang, "Multi-UAV network assisted intelligent edge computing: Challenges and opportunities," China Communications, vol. 19, no. 3, pp. 258-278, 2022. https://doi.org/10.23919/JCC.2022.03.019

[3] U. A. Bukar, M. S. Sayeed, S. F. A. Razak, S. Yogarayan, and O. A. Amodu, "An exploratory bibliometric analysis of the literature on the age of information-aware unmanned aerial vehicles aided communication," Informatica, vol. 47, no. 7, pp. 91-114, 2023. https://doi.org/ 10.31449/inf.v47i7.4783

[4] C. Zhu, G. Zhang, and K. Yang, "Fairness-aware task loss rate minimization for multi-UAV enabled mobile edge computing," IEEE Wireless Communications Letters, vol. 12, no. 1, pp. 94-98, 2022. https://doi.org/10.1109/LWC.2022.3218035

[5] J. Zhang, M. Li, Z. Chen, and B. Lin, "Computation offloading for object-oriented applications in a UAV-based edge-cloud environment," The Journal of Supercomputing, vol. 78, no. 8, pp. 10829-10853, 2022. https://doi.org/10.1007/s11227-021-04288-0

[6] J. Chen, X. Qing, F. Ye, K. Xiao, K. You, and Q. Sun, "Consensus-based bundle algorithm with local replanning for heterogeneous multi-UAV system in the time-sensitive and dynamic environment," The Journal of Supercomputing, vol. 78, no. 2, pp. 1712-1740, 2022. https://doi.org/10.1007/s11227-021-03940-z

[7] Z. Wei, M. Zhu, N. Zhang, L. Wang, Y. Y. Zou, Z. Y. Meng, H. C. Wu, and Z. Y. Feng, "UAV-assisted data collection for internet of things: A survey," IEEE Internet of Things Journal, vol. 9, no. 17, pp. 15460-15483, 2022. https://doi.org/10.1109/JIOT.2022.3176903

[8] W. Liu, K. Quijano, and M. M. Crawford, "YOLOv5-Tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 8085-8094, 2022. https://doi.org/10.1109/JSTARS.2022.3206399

[9] S. Pal, A. Roy, P. Shivakumara, and U. Pal, "Adapting a swin transformer for license plate number and text detection in drone images," Artificial Intelligence and Applications, vol. 1, no. 3, pp. 145-154, 2023. https://doi.org/10.47852/bonviewAIA3202549

[10] W. You, C. Dong, Q. Wu, Y. B. Qu, Y. L. Wu, and R. He, "Joint task scheduling, resource allocation, and UAV trajectory under clustering for FANETs," China Communications, vol. 19, no. 1, pp. 104-118, 2022. https://doi.org/10.23919/JCC.2022.01.009

[11] S. Halder, A. Ghosal, and M. Conti, "Dynamic super round-based distributed task scheduling for UAV networks," IEEE Transactions on Wireless Communications, vol. 22, no. 2, pp. 1014-1028, 2022. https://doi.org/10.1109/TWC.2022.3200366

[12] Z. Niu, H. Liu, X. Lin, and J. Z. Du, "Task scheduling with UAV-assisted dispersed computing for disaster scenario," IEEE Systems Journal, vol. 16, no. 4, pp. 6429-6440, 2022. https://doi.org/10.1109/JSYST.2021.3139993

[13] Y. Wang, X. Wei, H. Wang, J. H. Fan, J. Chen, K. Zhao, and Y. Y. Hu, "Joint UAV deployment, SF placement, and collaborative task scheduling in heterogeneous multi-UAV-empowered edge intelligence," IET Communications, vol. 17, no. 5, pp. 641-657, 2023. https://doi.org/10.1049/cmu2.12570

[14] I. Khettabi, L. Benyoucef, and M. Amine Boutiche, "Sustainable multi-objective process planning in reconfigurable manufacturing environment: Adapted new dynamic NSGA-II vs new NSGA-III," International Journal of Production Research, vol. 60, no. 20, pp. 6329-6349, 2022. https://doi.org/10.1080/00207543.2022.2044537

[15] M. Awad, M. Abouhawwash, and H. N. Agiza, "On NSGA-II and NSGA-III in portfolio management," Intelligent Automation and Soft Computing, vol. 32, pp. 1893-1904, 2022. https://doi.org/10.32604/iasc.2022.023510

[16] N. N. Johnson, V. Madhavadas, B. Asati, A. Giri, S. A. Hanumant, N. Shajan, and K. S. Arora, "Multi-objective optimization of resistance spot welding parameters of BH340 steel using Kriging and NSGA-III," Transactions of the Indian Institute of Metals, vol. 76, no. 11, pp. 3007-3020, 2023. https://doi.org/10.1007/s12666-023-03051-8

[17] S. Harif, G. Azizyan, M. Dehghani Darmian, and G. Mohammad, "Selecting the best location of water quality sensors in water distribution networks by considering the importance of nodes and contaminations using NSGA-III (case study: Zahedan water distribution network, Iran)," Environmental Science and Pollution Research, vol. 30, no. 18, pp. 53229-53252, 2023. https://doi.org/10.21203/rs.3.rs-1938809/v1

[18] T. Thomas, S. Srinivas, and C. Rajendran, "Collaborative truck multi-drone delivery system considering drone scheduling and en route operations," Annals of Operations Research, pp. 1-47, 2023. https://doi.org/10.1007/s10479-023-05418-y

[19] F. Sun, X. Wang, and R. Zhang, "Task scheduling system for UAV operations in agricultural plant protection environment," Journal of Ambient Intelligence and Humanized Computing, pp. 1-15, 2020. https://doi.org/10.1007/s12652-020-01969-1

Image Stitching Technology for Police Drones Using an Improved Image Registration Method Incorporating ORB Algorithm

Yingzi Cong
School of Continuing Education, Criminal Investigation Police University of China, Shenyang 110854, China
E-mail: ccpcyz@163.com

Keywords: fusion ORB algorithm, image registration, police drones, drone imaging, image stitching

Received: March 8, 2024

Drone aerial photography technology can assist public security departments in conducting on-site investigations and evidence collection. The single-aerial image captured by drones has limited field of view. Due to the influence of reconnaissance scenes, there are various shooting angles and overlapping areas. An improved image registration algorithm was proposed based on ORB. This algorithm combined segmentation and denoising processing to evenly distribute ORB feature points throughout the entire drone image. On this basis, Laplace fusion algorithm was used to match the overlapping regions. The study compared the improvement of ORB with several other methods. Three sets of experiments were compared and analyzed with the other three different fusion algorithms. These experiments confirmed that the improved image registration method had significant improvements in accuracy and speed compared with existing methods. The improved ORB registration algorithm was superior to traditional ORB in registration accuracy and registration speed. This algorithm had a registration rate increase of 7.4% and a time reduction of 153.97 seconds. This method effectively improves the splicing effect. It also handles the details of the splicing seam well, providing a basis for police to analyze the scene.

Povzetek: Študija uporablja izboljšan algoritem registracije slik, ki vključuje ORB za tehnologijo združevanja slik pri policijskih dronih, izboljšuje natančnost in hitrost registracije ter zagotavlja boljše rezultate spajanja slik.

1 Introduction

The drone aerial photography system uses unmanned aerial vehicles, carrying image acquisition devices. Meanwhile, wireless transmission technology is utilized to quickly obtain low altitude high-resolution images [1]. With the development of science and technology, drones are widely used in various aspects of society, such as in industry. They can be used for safety monitoring in hazardous areas such as chemical sites and high-voltage lines [2]. Police drones are a new type of police equipment that has the advantages of convenience and efficiency in aerial reconnaissance. Meanwhile, police drones can provide key technical support for police [3]. Traditional investigation methods can no longer meet the growing demand for investigation and evidence collection. Because there are challenges such as road traffic congestion, uncertainty in the scope of investigation, and the increasing use of anti-investigation methods by criminals [4]. So in recent years, drones are well applied in areas such as counter-terrorism and stability maintenance, air defense surveillance, and target tracking [5-6]. Drone aerial photography technology can assist public security departments in on-site investigation and evidence collection, real-time command and dispatch. Drone aerial photography technology has important practical significance and broad application prospects in public security. Oriented Fast and Rotated Brief (ORB) is an algorithm used in image processing, mainly for feature detection and description. By combining FAST feature detectors and Brief descriptors, feature points in images can be quickly and accurately detected and described [7]. Compared to other feature detection and description algorithms, ORB has significant advantages in speed and is widely used in fields such as object detection, tracking, recognition, and stitching. At present, the data collected by police drones are limited to general video surveillance scenes. The construction of panoramic images is relatively scarce. For this purpose, the study intends to carry out research on drone aerial image stitching technology and apply it to evidence collection and investigation. This has significant practical significance for alleviating the current situation of insufficient police force in public security organs. Meanwhile, this paper can improve the efficiency of investigation and evidence collection.

The study consists of four parts. Firstly, a discussion and analysis of previous research are conducted. Secondly, an innovative drone image stitching technology is proposed in combination with ORB. Then, the proposed technology is tested and analyzed to verify its application value in police drones. Finally, a conclusion is drawn and future work prospects are presented.

ORB is an efficient, accurate, and widely used image

processing algorithm, which has a wide range of applications in the field of image registration. Image registration is also an important technical means in various image processing, which is also an excellent image processing technique.

Tian et al. proposed a novel similarity measure that combined cosine and other methods. It was combined with the directional features and ORB feature extraction of accelerated segmentation testing to achieve rapid image forgery detection. This image was segmented into overlapping image blocks, and each image block's feature points were extracted using ORB. Similarity and match text were calculated by similarity means. These two blocks having max similarity were identified. Combining these two methods could demonstrate good robustness to lighting and others [8]. Batista et al. used image analyzing means to extract features and established an animal thermal comfort model based on thermal image classification of 7/8 Holstein-Gil heifers. Using an ORB detector to extract features from thermal images could verify the differences in the image's pixel intensity. PCA indicated that Tcan, Tbac, Thea, and Tski had a better correlation in characterizing animal thermal comfort [9]. Wu et al. proposed a 3D measuring method using feature correspondence. Only a frame of sine wave stripe pattern needed to be projected onto the tested object. Image correcting means were used to correct pixel size. A camera was used to capture five frames of deformation patterns, and the corresponding modulation patterns were extracted. The modified algorithm could assess mismatches and achieve pixel matching well [10]. Lin et al. proposed a CapsNetORB framework to achieve distorted target recognition and suppress deformation effects in the final prediction. The previous encoded vectors for the latter, while the latter detected interval dimensions with unchanged spatial scales to bridge the correlation between the source standard image and the distorted image. Therefore, the source standard image's category with the topmost correspondence was the final result [11]. Ma et al. proposed an ORB using a modified Quadtree Oriented Fast and Rotated Brief (QTORB). A new calculating method was put forward to elevate the algorithm's ability in uniform areas. QTORB managed and optimized feature points to remove overly intensive and coinciding points. QTORB could elevate features' distributing uniformity better [12].

With the continuous development of technology, image stitching technology has been widely applied in various fields. For police drones, it also plays an important role. To achieve better image stitching results, many researchers have improved image registration algorithms, which can be widely applied in multiple fields. Tang et al. proposed a fast sonar image stitching method, which included denoising, feature extraction, initial matching, stitching, and optimization. Based on the Euclidean distance between initial matching points and the inclination angle of the connecting line, poorly matched feature point pairs were removed to avoid incorrect matching [13]. Laarousi et al. proposed a new method. A new method was used in this experiment to distinguish moving objects. A map was created, where even in homogeneous regions, each pixel had the one-of-a-kind value within its surrounding environment. This method combined the early calculated mapping to quickly and effectively find the optimal seam [14]. Combining aerial imaging and visual analysis in open-pit mines, Winkelmaier et al. provided a new method for monitoring tension cracks, which might happen on workbenches or capture platforms excavated based on computer-aided design models. The size, location, and evolution of tension cracks were commonly used to predict slope failure and ensure safe mining operations. Clear photos were utilized in this experiment to depict cracks. Controllable filters, ENet, and UNet deep learning models were used to depict tension cracks. ENet adopted a left one cross-validation means to generate the excellent curve [15]. Hahn proposed a hyperspectral snapshot imaging method based on intermediate image plane diffraction. In this intermediate image, the planar diffraction microstructure was used to deflect light to the aperture where spectral filtering was performed. This method was related to imaging spectroscopy with reduced spectral resolution. Compared with filter-based hyperspectral snapshot imaging, it avoided the manufacturing difficulties of mosaic filter arrays, but spatial and spectral resolutions became coupled. This led to the uncertainty product of spatial and spectral resolution [16]. Legleiter et al. evaluated this method's potential to enhance retrieval by utilizing field observations of water depth and helicopter footage of clear flowing rivers. These results indicated that the depth inferred from the average image was more exact than that of a single image. The regression R-2 between observed and predicted values increased. IBARI elevated image depth maps' texture, making the representation of waterway morphology smoother and more coherent [17]. The relevant literature research is shown in Table 1.

Table 1: Related literature research

Reference	Method	Result	Limitation	Critical evaluation
Tian et al. [8]	Cosine and Jaccard similarity measures are combined with ORB feature extraction	High robust feature matching	Large consumption of computing resources	Computing efficiency needs to be improved
Batista et al. [9]	Animal thermal comfort model is established by ORB detector and PCA	Effectively classify thermal images	Image quality-dependent	The generalization ability needs to be verified
Wu et al. [10]	On-line 3D measurement and improved motion statistical feature algorithm	Robust pixel matching	Environment and equipment sensitive	Practicality needs to be enhanced
Lin et al. [11]	The CapsNetORB framework suppresses deformation effects	Improve the accuracy of target recognition	High computational cost	Framework needs to be simplified
Ma et al. [12]	QTORB algorithm optimizes feature point extraction	The feature points are evenly distributed	Application scenarios need to be adjusted	Experimental verification needs to be increased
Tang et al. [13]	Fast sonar image stitching	Improve stitching quality	Feature point selection is sensitive	The automation needs to be increased
Laaroussi et al. [14]	Dynamic object detection is combined with A* algorithm	Improve detection accuracy	High algorithm complexity	Practical applications need to be simplified
Winkelmaier et al. [15]	Deep learning models monitor tension cracks	Excellent model performance	Training data are in high demand	Adaptability needs improvement
Hahn et al. [16]	New method of hyperspectral snapshot imaging	Avoid filter manufacturing difficulties	Image quality is affected	Resolution needs to be optimized
Legleiter and Kinzel [17]	Evaluate deep retrieval potential	Extrapolating depth is more accurate	Subject to water conditions	Environmental adaptability needs to be enhanced

In summary, it is of great significance to study the image stitching technology of improved image registration methods based on ORB in police drones. In the future, ORB can be further improved on the existing basis. Therefore, more efficient and accurate feature point extraction and matching methods can be explored to achieve higher quality image stitching results. Meanwhile, ORB can further expand the application fields of image stitching technology and explore its application value in other fields.

2 Image registration based on improved ORB

Image registration is an important technical means. The main goal of image registration is to establish a transformation model between two images with overlapping regions and convert them into the same coordinate system. On this basis, a new feature-based registration method is proposed. SIFT, SURF, ORB, BRISK, etc. are compared and the optimal algorithm is selected.

2.1 Analysis of multiple algorithms based on feature extraction in image registration

Common image features used for feature extraction include spots, corners, inflection points, contours, and edges. Speckles are generally areas that differ in color or grayscale from surrounding pixels. Corner points are often the intersection of a corner or a line of an object in an image. Therefore, the noise resistance of corner points is stronger than corner points [18-19]. In spot detection, there are two classic spot detection methods: SIFT and SURF, while corner spot detection methods such as ORB and BRISK. In SIFT, firstly, the scale space of the image is established and the extreme values of the image are extracted. Secondly, interpolation is utilized to locate the pixels in the image and eliminate false points. When constructing a scale space, it is achieved through scale transformation, represented by formula (1).

$$G(x,y,\sigma)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (1)$$

In formula (1), a Laplace transform is applied to it,

represented by formula (2).

$$\nabla^2 g = \frac{\partial^2 g}{\partial^2 x} + \frac{\partial^2 g}{\partial^2 y} \tag{2}$$

The result obtained through Laplace transform is represented by formula (3).

$$\nabla^2_{norm} = \sigma^2 \left(\frac{\partial^2 g}{\partial^2 x} + \frac{\partial^2 g}{\partial^2 y} \right) = -\frac{1}{2\pi\sigma^2} \left(1 - \frac{x^2 + y^2}{\sigma^2} \right) e^{-\frac{x^2+y^2}{2\pi\sigma^2}} \tag{3}$$

The Gaussian difference scale space of the image is represented by formula (4).

$$D(x, y, \sigma) = G(x, y, k\sigma) * I(x, y) - G(x, y, \sigma) * I(x, y) \tag{4}$$

In formula (4), $I(x, y)$ is a two-dimensional image. σ is the scale coordinate, which determines the smoothness of the image A smaller σ indicates a clearer image, while a larger σ indicates a more blurry image. Each pixel in the middle two frames of each group is detected in the degree of freedom space. The adjacent pixels consist of 8 adjacent pixels in this image. There are 18 pixels in the adjacent image, meaning 26 pixels need to be detected. Figure 1 shows the process of detecting extreme points in the scale space.

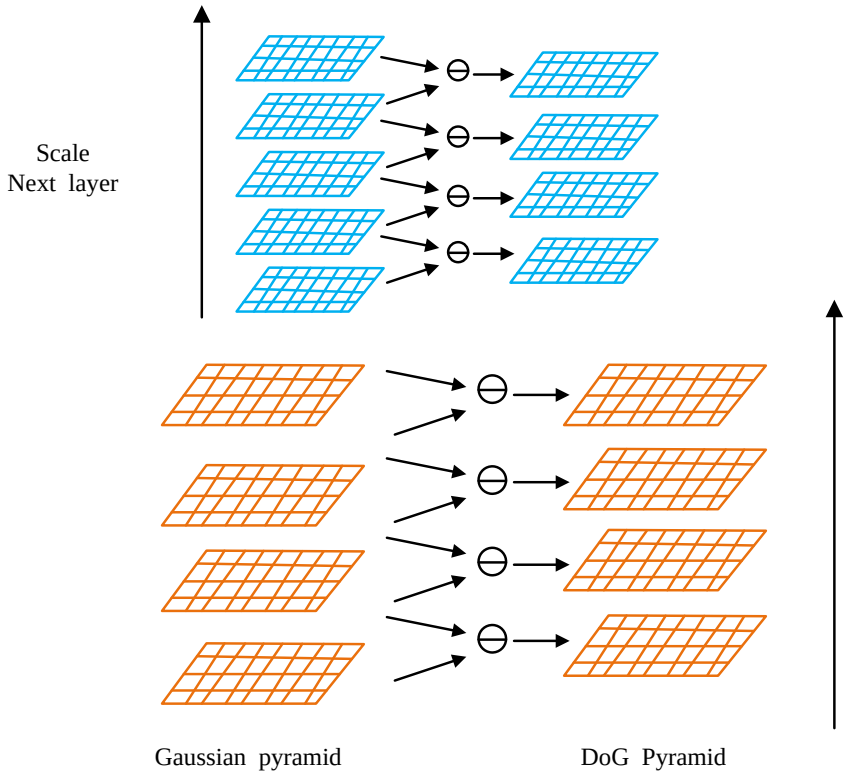


Figure 1: Scale space detection of extreme point process

From Figure 1, if this point is the maximum or minimum of 26 adjacent points, it can be considered as a feature point within this range. Extreme points are detected and pseudo points were removed in the scale space. The Gaussian difference scale space of the image is expanded using the Taylor formula, represented by formula (5).

$$D(X) = D + \frac{\partial D^T}{\partial X} + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \tag{5}$$

In formula (5), $\hat{X} = (x, y, \sigma)^T$ is the offset of the extreme value. The partial derivative of $D(X)$ is calculated and defined as 0. The position of the extreme point is represented by formula (6).

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X} \tag{6}$$

From this, formula (7) can be obtained.

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial x} \hat{x} \tag{7}$$

If the deviation of the extreme point on any dimension exceeds 0.5, it is shifted along the offset direction until the offset between the sub-pixel points does not exceed 0.5. Then the point is considered a feature point. When $|D(\hat{x})| \le 0.03$, it indicates that the response value is relatively small and susceptible to noise. Meanwhile, this

pole should be removed. Firstly, the feature points are rotated, and then their neighborhoods are rotated to rotate in the direction of the feature points. Firstly, the 16*16 region where the feature points are located is segmented. Each sub-region is segmented to obtain the gradient values of each sub-region. This algorithm uses Gaussian weight function to calculate the weight of gradients in the neighborhood. The closer the algorithm is to the feature points, the greater its weight in Figure 2.

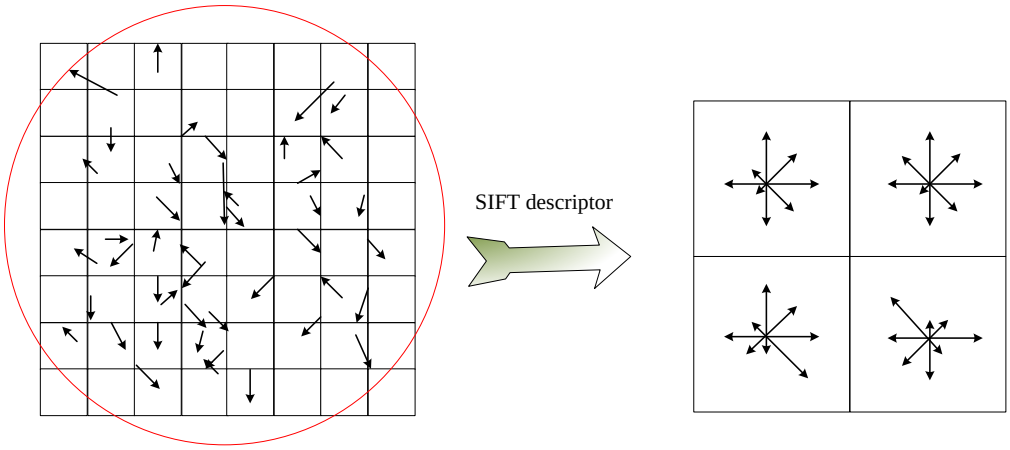


Figure 2: A schematic representation of the SIFT descriptor generation

In Figure 2, there are 16 sub-regions, each with 8 different gradients, totaling 16×8=128 values, forming a 128-dimensional SIFT descriptor. ORB applies the feature extraction method at FAST corners to feature extraction and optimizes this method. The FAST corner extraction method involves selecting 16 pixels from any pixel in an image on three discrete circles around it in Figure 3.

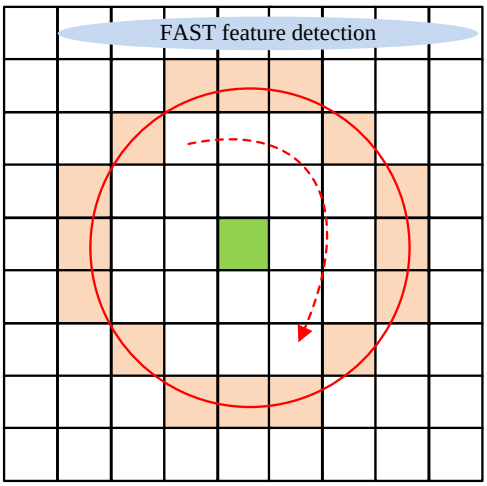


Figure 3: A schematic diagram of the FAST feature detection procedure

From Figure 3, if all n pixels out of the 16 pixels on the circle have pixel values larger than $I_p + t$ or smaller than $I_p - t$, then the point is considered as a corner point. I_p is the pixel value at the p point. t is a threshold. Usually, n takes 12 or 9. A grayscale histogram decomposition method based on wavelet transform is proposed. This method normalizes the grayscale histogram to obtain the grayscale histogram and normalizes the grayscale histogram. Firstly, FAST is utilized to sort the feature points in each layer of cone images. The Harris corner response values of feature points are calculated and sorted. The top N points are retained and expressed using formula (8).

$$\begin{aligned} R &= derM - \alpha(trM)^2 \\ M &= \sum_w \begin{bmatrix} I_x(x,y)^2, I_x(x,y)I_y(x,y) \\ I_x(x,y)I_y(x,y), I_y(x,y)^2 \end{bmatrix} \\ &= \begin{bmatrix} \sum_w I_x(x,y)^2, \sum_w I_x(x,y)I_y(x,y) \\ \sum_w I_x(x,y)I_y(x,y), \sum_w I_y(x,y)^2 \end{bmatrix} \end{aligned} \tag{8}$$

In formula (8), $I_x(x,y)$ and $I_x(x,y)$ are the partial derivatives of image $I(x,y)$. Based on the grayscale centroid of ORB, the orientation of feature points is obtained by utilizing the grayscale of pixels and the displacement of their adjacent centroids. The moment of one feature point's domain is represented by formula (9).

$$m_{pq} = \sum_{x,y} x^p y^q I(x,y) \tag{9}$$

The centroid of its corresponding neighborhood is represented by formula (10).

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \tag{10}$$

The angle between the feature point and the mass center is the principal direction of feature points, which is represented by formula (11).

$$\theta = a\tan 2(m_{01}, m_{10}) \tag{11}$$

The BRIEF descriptor is essentially a binary encoding that selects a set of point pairs centered on feature points and compares their grayscale values [20]. A binary comparison criterion function is defined for the neighborhood space and expressed using formula (12).

$$\tau(p;x,y) = \begin{Bmatrix} 1, p(x) < p(y) \\ 0, p(x) > p(y) \end{Bmatrix} \tag{12}$$

A binary string with a length of n can be obtained, represented by formula (13). This binary string is obtained by selecting multiple pairs of point sets around the feature points and comparing the judgment functions

$$f_n(p) + \sum_{1 \le i \le n} 2^{i-1} \tau(p; x_i, y_i) \tag{13}$$

ORB takes the main direction of feature points as the feature descriptor. Meanwhile, ORB adds a rotation matrix to its point-to-point matrix to ensure that the rotation of the descriptor remains unchanged. Firstly, n points near the feature points are combined to form a $2 \times n$ matrix, which is represented by formula (14).

$$S = \begin{pmatrix} x_1, x_2, ..., x_n \\ y_1, y_2, ..., y_n \end{pmatrix} \tag{14}$$

The rotation matrix corresponding to the direction of feature points is represented by formula (15).

$$R_\theta = \begin{bmatrix} \cos\theta, \sin\theta \\ -\sin\theta, \cos\theta \end{bmatrix} \tag{15}$$

The corresponding feature point pair matrix is represented by formula (16).

$$S_\theta = R_\theta S \tag{16}$$

The final feature descriptor after determining the direction is represented by formula (17).

$$g_n(p,\theta) = f_n(p) \big| (x_i, y_i) \in S_\theta \tag{17}$$

ORB uses a greedy search method to select the largest and smallest 256 pairs from all candidate point sets as feature descriptors. Therefore, the discriminability of BRIEF descriptors can be improved. An initial mask is first created for the image to be registered during mask construction and feature point positioning. The size of the mask is the same as that of the image, and the initial value is all zero. The improved ORB is used to locate the feature points on the initial mask. ORB quickly identifies the corner points in the image through the FAST corner detector. ORB is combined with the BRIEF descriptor for the feature description. An adaptive threshold is introduced in the algorithm to adjust the feature points extracted. Therefore, the distribution uniformity of feature points can be improved, and the uniform distribution of feature points can be ensured in the entire image range. After the feature points are located, the mask is updated to mark the positions of these feature points as 1 and keep the remaining positions as 0. In this way, the mask only retains the feature point information, which provides the basis for the subsequent image matching and fusion. When extracting feature points, the whole image is divided into several small blocks. Then the feature points are extracted independently in each small block. The minimum denoising algorithm is used to remove the clustered feature points and reduce the possibility of mismatching. The denoising process preserves scattered and distinguishable feature points by setting a threshold to identify. Meanwhile, those points that are too clustered are removed. For retained feature points, the ORB algorithm is used to generate descriptors.

2.2 Image stitching based on improved ORB in police drones

ORB only uses a fixed value to determine the required number of feature points. Therefore, ORB is not suitable for general situations and improves the speed and efficiency of matching. This project focuses on the characteristics of public security drone image acquisition and the problems existing in ORB. The uneven distribution of feature points in ORB is solved based on the method of image block segmentation. The clustering degree of feature points in ORB is improved through denoising methods. Figure 4 shows an optimized image matching method.

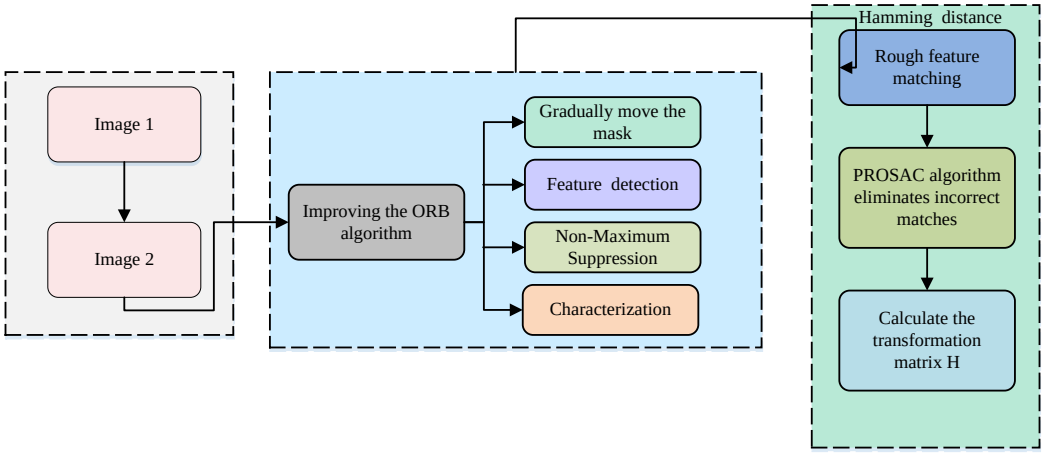


Figure 4: Process diagram for improving image registration methods

In Figure 4, first, the improved ORB algorithm is adopted for feature extraction in the registration of the two images. For the images to be registered, the study constructs a mask that will be used to identify and extract key feature points in the images. The construction of the mask is an iterative process. The feature points are accurately detected in different areas by gradually moving and covering the entire image. Then the ORB algorithm is used to locate the feature points at each mask position. The FAST corner detector and BRIEF descriptor are combined to detect and describe the feature points in the image quickly and accurately. A minimum denoising algorithm is used to identify and remove the clustered feature points by setting a threshold. Therefore, the quality of the feature points can be improved and the possibility of mismatching can be reduced. For retained feature points, the ORB algorithm generates descriptors that contain key information about the feature points, such as location, scale, and orientation. Then the whole image is traversed to obtain more feature points in a wider range and improve the accuracy of image registration. In the stage of feature point matching, hamming distance is used as the matching measure to measure the similarity between two feature descriptors. Meanwhile, the corresponding pair of feature points between two images can be found. The PROSAC algorithm is used to eliminate the wrong matching pairs and calculate the transformation matrix between images. The accuracy of matching is improved iteratively by adding matching pairs step by step and constructing a random sample consistency model. The Laplacian multi-resolution fusion algorithm is used for image fusion. When constructing the Laplacian pyramid, the top image and other levels do not use fusion rules, but directly use the original image data. The fusion process is shown in Figure 5.

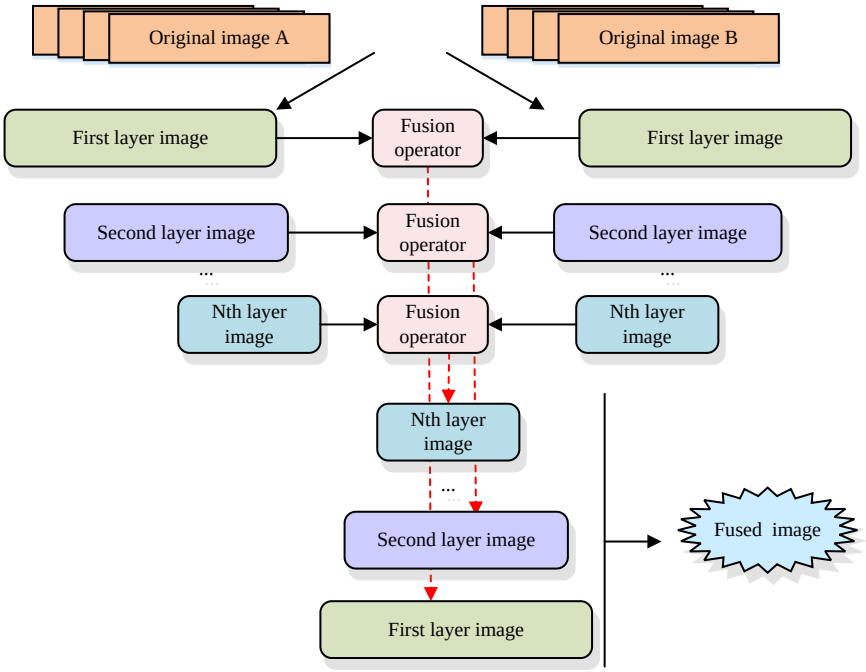


Figure 5: Multi-resolution fusion process based on Laplace

In Figure 5, the original images A and B are subjected to Laplace transform to obtain the first layer of images, respectively. The Nth layer of images is obtained sequentially by combing with the fusion operator. Pyramid reconstruction is performed on N-layer images to obtain fused images. In the improved image fusion algorithm, when establishing a panoramic view of the investigation area from police drone aerial images, it is necessary to ensure the clarity of the image and avoid "broken lines" and "ghosting". Traditional fusion methods cannot solve this problem well, but Laplace fusion can do this. Figure 6 shows an improved image fusion algorithm.

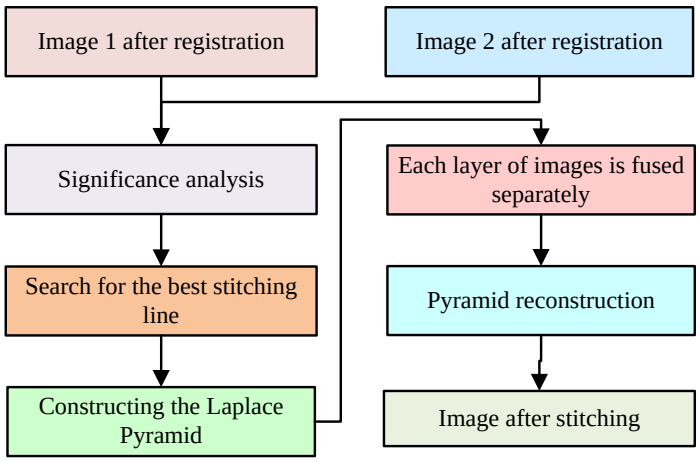


Figure 6: Improved image fusion algorithm flowchart

In Figure 6, the improved image fusion algorithm first performs image registration on these two images that need to be fused. Significant analyses are conducted on both. The optimal stitching line is searched by comparing the color and geometric differences between them. A Laplace pyramid is constructed to fuse images from each layer. The fused image is obtained after image stitching after pyramid reconstruction.

3 Comparative analysis of image registration and drone image stitching performance based on improved ORB

The study compared SIFT, SURF, ORB, and BRISK from multiple performance aspects through experiments to determine the optimal algorithm. This algorithm was optimized and improved, and its registration data were compared. Subsequently, the quality of the image obtained from stitching was evaluated.

3.1 Comparison of registration rates under different changing conditions among four registration algorithms

Four sets of image sequences with features such as scale and rotation, blur, perspective and lighting were selected. They were from the image library built in the Visual Geometry Laboratory at the University of Oxford. They were validated using 6 progressively increasing images as samples for the method. Four methods including SIFT, SURF, ORB, and BRISK were used to compare the feature matching rates and registration times. Faced with 6 images with scale, rotation, and blur changes, the first image was registered with the remaining 5 images based on SIFT, SURF, ORB, and BRISK in Figure 7.

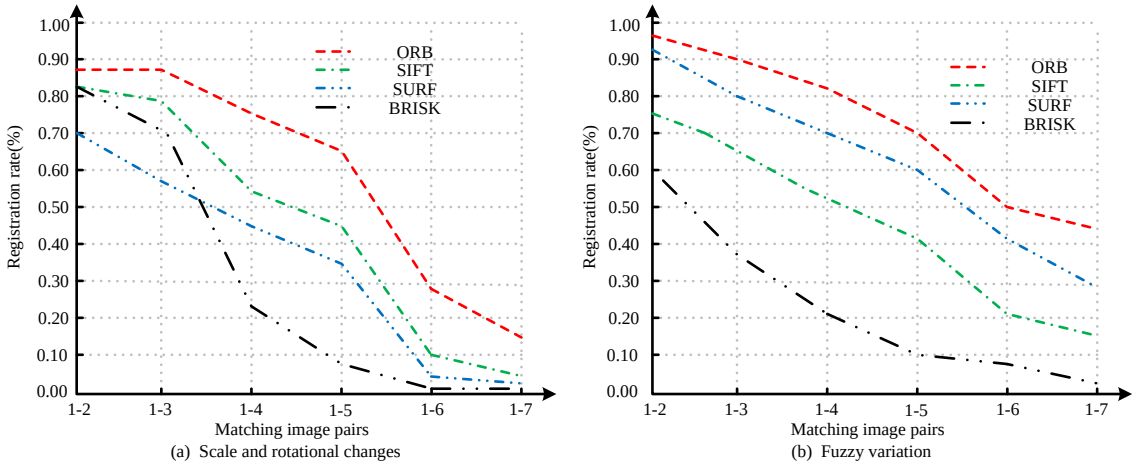


Figure 7: Comparison of registration rates of four algorithms under scale changes, rotation changes, and fuzzy changes

In Figure 7 (a), the accuracy of the SIFT, SURF, and ORB methods gradually decreased with increasing scale. The accuracy of BRISK decreased sharply within 1-4, and the robustness to scale and rotation was poor. In Figure 7 (b), the registration result of ORB was the best, followed by SURF and SIFT. The BRISK algorithm was the weakest. There were issues with changes in perspective and lighting in the image data captured by police drones. Therefore, four algorithms were used to register the images and obtain corresponding registration rates in Figure 8.

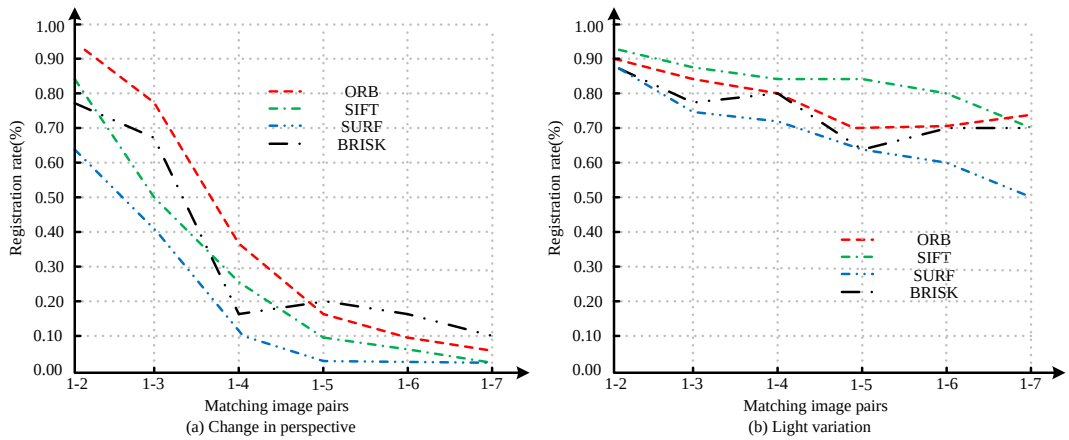


Figure 8: Comparison of registration rates of four algorithms under changes in perspective and lighting conditions

In Figure 8 (a), the registration results of the four methods were basically consistent. The registration performance of SIFT and ORB was slightly better. However, when there was a significant change in angle, the accuracy of all four methods decreased, resulting in weak robustness for large angle images. In Figure 8 (b), all four methods had good lighting adaptability, and there was a slight decrease in accuracy under different lighting conditions. The feature points, matching pairs, correct matching pairs, matching accuracy, and registration time of these two images under the influence of various objective factors were summarized in Table 2.

Table 2: Comparison of image feature point matching rate and matching time under objective changes

Change type/indicator	Algorithm type	Feature points in Figure 1	Feature points in Figure 2	Pairs	Correct matching	Matching accuracy	Matching accuracy standard deviation	Matching accuracy 95% confidence interval	Registration time(s)
Scale and rotational changes	SIFT	8679	8367	2567	2191	0.8535	0.020	[0.8335, 0.8735]	15.765
	SURF	3385	3405	1365	985	0.7216	0.015	[0.7011, 0.7421]	5.687
	ORB	502	502	275	243	0.8836	0.012	[0.8698, 0.8974]	0.962
	BRISK	2416	2424	312	265	0.8494	0.014	[0.8314, 0.8674]	2.105
Fuzzy variation	SIFT	3389	1803	756	539	0.7130	0.018	[0.6978, 0.7282]	15.355
	SURF	1446	986	724	615	0.8494	0.016	[0.8326, 0.8662]	5.757
	ORB	502	502	315	297	0.9429	0.013	[0.9291, 0.9566]	0.969
	BRISK	798	319	78	44	0.5641	0.022	[0.5458, 0.5824]	2.110
Change in perspective	SIFT	2486	3084	1065	865	0.8122	0.021	[0.7979, 0.8265]	15.769
	SURF	1093	765	505	321	0.6356	0.024	[0.6156, 0.6556]	5.705
	ORB	502	284	239	211	0.8828	0.020	[0.8682, 0.8974]	0.965
	BRISK	595	631	104	76	0.7308	0.023	[0.7146, 0.7470]	2.113
Light variation	SIFT	2436	2143	1136	996	0.8768	0.019	[0.8610, 0.8926]	15.739
	SURF	1533	1253	698	603	0.8639	0.017	[0.8499, 0.8789]	5.703
	ORB	502	502	240	214	0.8917	0.018	[0.8778, 0.9056]	0.967
	BRISK	873	687	203	166	0.8177	0.020	[0.8037, 0.8317]	2.113

In Table 2, under the type of scale rotation change, the ORB algorithm had the highest matching accuracy, reaching 0.8836 with a standard deviation of 0.012. The SURF algorithm had the lowest matching accuracy, 0.7216 with a standard deviation of 0.015. The registration time of the BRISK algorithm was the shortest, at 2.105 seconds. Under the fuzzy change type, the ORB algorithm had the highest matching accuracy, reaching 0.9429 with a standard deviation of 0.013. The BRISK algorithm had the lowest matching accuracy, 0.5641 with a standard deviation of 0.022. The registration time of SURF algorithm was the shortest, at 5.757 seconds. Under the type of perspective change, the SIFT algorithm had the highest matching accuracy, reaching 0.8122 with a standard deviation of 0.021. The SURF algorithm had the lowest matching accuracy, 0.6356 with a standard deviation of 0.024. The registration time of the BRISK algorithm was the shortest, at 2.113 seconds. Under different types of lighting changes, the SIFT algorithm had the highest matching accuracy, reaching 0.8768 with a standard deviation of 0.019. The SURF algorithm had the lowest matching accuracy, reaching 0.8639 with a standard deviation of 0.017. The registration time for ORB was the shortest, at 0.967 seconds. Overall, ORB showed high matching accuracy in most change types, while BRISK had the shortest registration time in most change types. When selecting a feature detection algorithm, it is necessary to choose a suitable algorithm based on specific application scenarios and requirements.

3.2 Application analysis of improved ORB fusion algorithm in drone image mosaic

The study selected two overlapping drone aerial images with a size of 3840*2160 to test the adaptability of this method to drone aerial images. By using SIFT, SURF, ORB, BRISK, and the proposed improved image registration method, these two images were registered and the registration results were verified. Figure 9 shows the registration rates and registration times for five different algorithms.

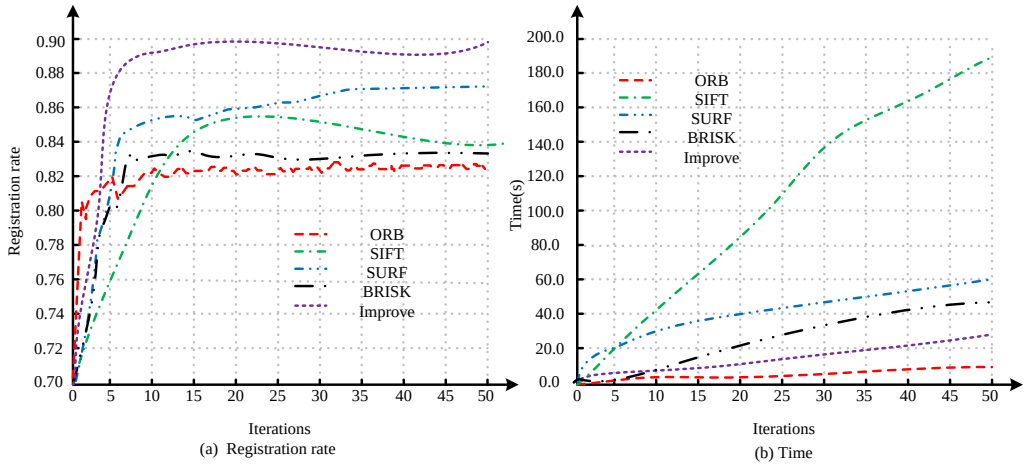


Figure 9: Registration data for ORB algorithm and improved algorithm

In Figure 9, this improved method was superior to the traditional ORB method in terms of registration accuracy and registration speed. The registration rate was increased from 82.89% to 90.29%. The standard deviation was 0.04%, an increase of 7.4%. The registration time is reduced from 189.33s to 35.36s. The standard deviation was 0.13s, which was reduced by 153.97s. Five different methods were used to extract features from drone aerial images. Preliminary feature matching was performed. On this basis, PROSAC purified specific matching pairs, removed incorrect pairs, and obtained the accuracy and time of registration for five algorithms.

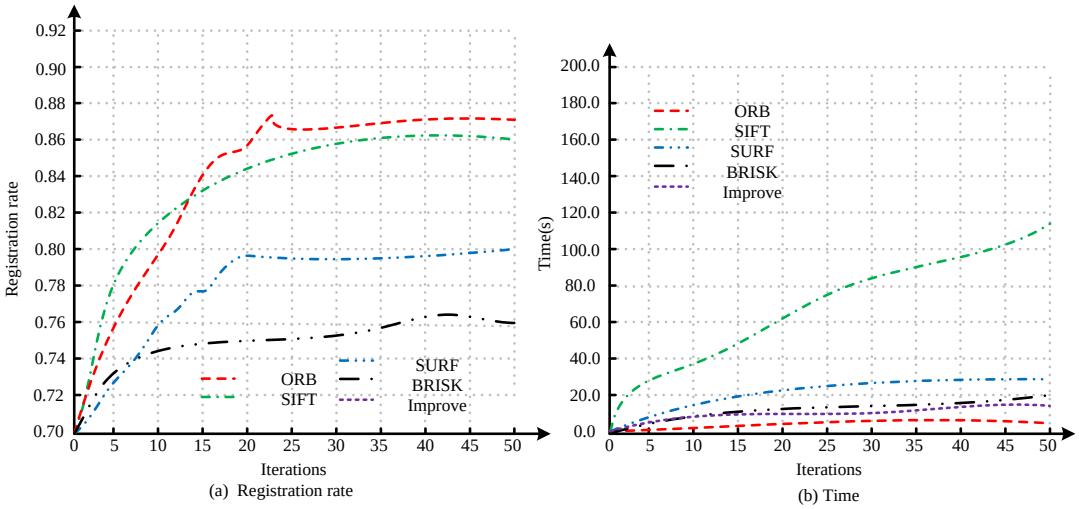


Figure 10: Registration data for ORB algorithm and improved algorithm

In Figure 10, this improved method had higher accuracy and higher registration efficiency than the traditional ORB method. The registration rate was increased from 87.89% to 93.49%. The standard deviation was 0.05%, an increase of 5.6%. The study tested the effectiveness of the improved fusion algorithm through three sets of experiments. Two images of 2300*1366 in size with translation and perspective changes were selected for experimentation. Firstly, an improved registration algorithm based on ORB was used for image registration. This algorithm was followed by direct averaging, gradual in and out, Laplace fusion algorithm, and improved fusion method for image fusion. For the evaluation of statistical features of a single image, commonly used objective evaluation indicators include information entropy, image mean and standard deviation, and average gradient. In this study, information entropy and average gradient were selected to evaluate the quality of concatenated images. Figure 11 shows the quality evaluation of the concatenated image obtained.

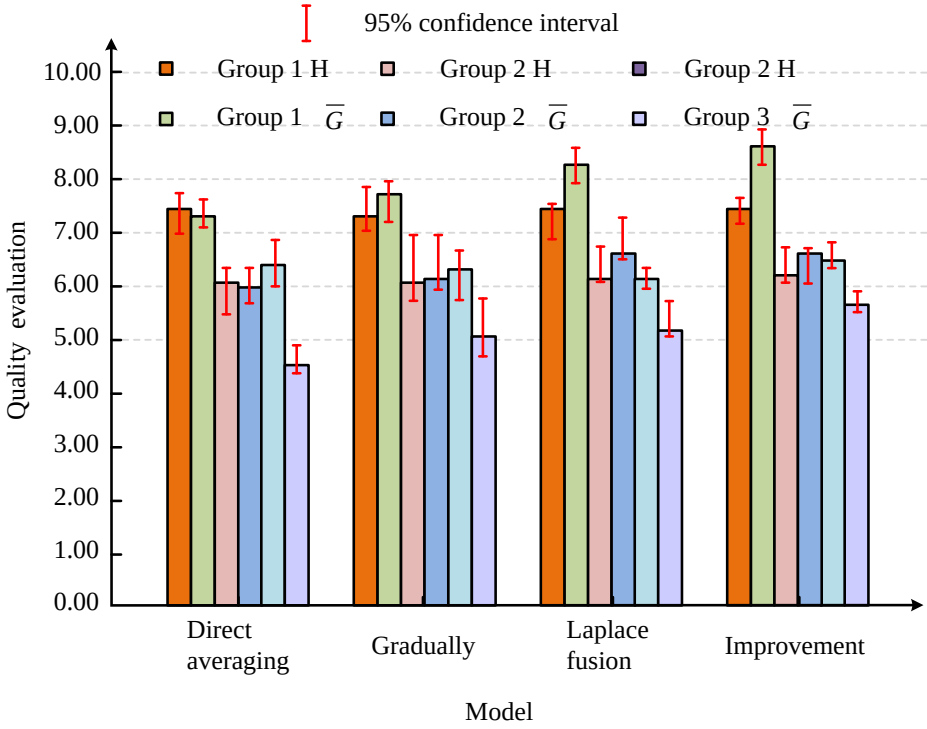


Figure 11: Quality evaluation of spliced images in three sets of experiments

In Figure 11, the improved algorithm resulted in improved stitching images in both information entropy and average gradient metrics compared with the other three methods. The improved algorithm improved the information entropy and average gradient of concatenated images. These images not only contained more information, but also had clearer details and textures. Especially in the third group of images with a large number of foreground targets, the significant improvement in average gradient further demonstrated. The improved algorithm had the advantage of preserving the details of the original image.

3.3 Discussion

The research method showed improved performance in multiple aspects compared with existing methods. Especially under the conditions of scale and rotation changes, the matching accuracy of the research method reached 0.8836. This method showed higher stability and accuracy compared to other algorithms. This improvement is mainly due to the denoising mechanism and Hamming distance matching strategy introduced in feature point detection and matching. This way effectively reduced the occurrence of mismatches. Under the condition of fuzzy changes, the performance of the research method was also outstanding, with a matching accuracy of up to 0.9429. The research method still maintained good registration performance even in dealing with degraded image quality. This is particularly important for image acquisition by drones in complex environments. Because the images captured by drones in practical applications are often subject to various interference. Compared with existing technologies, the research method had significantly improved in both information entropy and average gradient indicators. This result meant that the concatenated images not only contained more information, but also had clearer details and textures. Due to the optimization strategy in feature point extraction and image fusion stages, the research method effectively preserved the original details of the image while reducing visual distortion during the stitching. This algorithm-maintained efficiency and accuracy in multiple complex situations by introducing dynamic threshold adjustment and feature point quality evaluation mechanisms. This indicated that the research method had better environmental adaptability compared to existing methods.

4 Conclusion

Drone aerial photography technology can assist public security departments in conducting on-site investigations and evidence collection. Meanwhile, real-time command and dispatch can be conducted. The images captured by drones may be constrained by various factors such as shooting angle. Considering the characteristics of drone aerial images and the shortcomings of ORB, the detected feature points could be evenly distributed throughout the entire image based on the improved ORB registration algorithm. Through experimental verification, ORB was superior to other algorithms in terms of accuracy and registration speed. The improved ORB registration algorithm was also superior to traditional ORB in terms of registration accuracy and speed. The registration rate increased from 82.89% to 90.29%, an increase of 7.4%. The registration time decreased from 189.33 seconds to 35.36 seconds, a decrease of 153.97 seconds. The improved algorithm improved the information entropy and average gradient of the concatenated images. The details and textures of the images were also clearer. However, there are still shortcomings in the research. The current methods mainly optimize for specific drone models and environmental conditions. Different models of drones may have different camera specifications and flight characteristics. Different environmental conditions, such as lighting changes and weather conditions, may also affect image quality and registration results. Therefore, future work needs to consider the scalability of methods. Meanwhile, the study should explore how to adjust and optimize algorithms to adapt to different models of drones and diverse environmental conditions. In addition, extreme weather such as strong winds, rain and snow may affect the flight stability of drones, further affecting the quality of image acquisition. So further consideration is needed to combine and adapt with advanced flight control systems to enhance their applicability in various practical scenarios.

References

[1] E. D. Pusfitasari, J. Ruiz-Jimenez, I. Heiskanen, M. Jussila, K. Hartonen, and M. Riekkola, “Aerial drone furnished with miniaturized versatile air sampling systems for selective collection of nitrogen containing compounds in boreal forest,” Science of The Total Environment, vol. 808, no. 2, pp. 152011-152021, 2022. https://doi.org/10.1016/j.scitotenv.2021.152011

[2] C. Román, J. Llorens, A. Uribeetxebarria, R. Sanz, S. Planas, and J. Arno, “Spatially variable pesticide application in vineyards: Part II, field comparison of uniform and map-based variable dose treatments,” Biosystems Engineering, vol. 195, no. 7, pp. 42-53, 2020. https://doi.org/10.1016/j.biosystemseng.2020.04.013

[3] A. Georgiou, P. Masters, S. Johnson, and L. Feethan, “UAV-assisted real-time evidence detection in outdoor crime scene investigations,” Journal of forensic sciences. Vol. 67, no. 3, pp. 1221-1232, 2022. https://doi.org/10.1111/1556-4029.15009

[4] J. P. Kuehner, “Gravity-driven film flow inside an inclined corrugated pipe: An experimental investigation of corrugation shape and tip width,” Physics of Fluids, vol. 34, no. 12, pp. 122113-122132, 2022. https://doi.org/10.1063/5.0134555

[5] Z. Tuckey, “An integrated UAV photogrammetry-discrete element investigation of jointed Triassic sandstone near Sydney, Australia,” Engineering Geology, vol. 297, no. 2, pp. 106517-106532, 2022. https://doi.org/10.1016/j.enggeo.2022.106517

[6] F. Lodesani, M. C. Menziani, S. Urata, “Evidence of multiple crystallization pathways in lithium disilicate: a metadynamics investigation,” The journal of physical chemistry letters, vol. 14, no. 6, pp. 1411-1417, 2023. https://doi.org/10.1021/acs.jpclett.2c03563

[7] D. Pronoy, K. Sujit, P. Amalika, M. Abhisek, M. Golam, K. Chintak, K. Manoj, Q. Mohammed, and N. Uday, “Electric field guided fast and oriented assembly of mxene into scalable pristine hydrogels for customized energy storage and water evaporation applications,” Advanced Functional Materials, vol. 32, no. 47, pp. 2204622-2204635, 2022. https://doi.org/10.1002/adfm.202204622

[8] X. Tian, G. Zhou, and M. Xu, “Image copy-move forgery detection algorithm based on ORB and novel similarity metric,” IET Image Processing, vol. 14, no. 10, pp. 2092-2100, 2020. https://doi.org/10.1049/iet-ipr.2019.1145

[9] P. H. D. Batista, G. Almeida, and H. Pandorfi, “Thermal images to predict the thermal comfort index for Girolando heifers in the Brazilian semiarid region,” Livestock Science, vol. 251, no. 9, pp. 104667-104676, 2021. https://doi.org/10.1016/J.LIVSCI.2021.104667

[10] H. Wu, Y. Cao, H. An, Y. Li, H. Li, and C. Xu, “Fast and robust online three-dimensional measurement based on feature correspondence,” Optical Engineering, vol. 60, no. 7, pp. 74101-74117, 2021. https://doi.org/10.1117/1.oe.60.7.074101

[11] Z. Lin, W. Gao, J. Jia, and F. Huang, “CapsNet meets ORB: A deformation-tolerant baseline for recognizing distorted targets,” International Journal of Intelligent Systems, vol. 37, no. 6, pp. 3255-3296, 2022. https://doi.org/10.1002/int.22677

[12] C. Ma, X. Hu, J. Xiao, and G. Zhang, “Homogenized ORB algorithm using dynamic threshold and improved quadtree,” Mathematical Problems in Engineering, vol. 2021, no. 1, pp. 6693627-6693645, 2021. https://doi.org/10.1155/2021/6693627

[13] Z. Tang, G. Ma, J. Lu, Z. Wang, B. Fu, and Y. Wang, “Sonar image mosaic based on a new feature matching method,” IET Image Processing, vol. 14, no. 10, pp. 2149-2155, 2020. https://doi.org/10.1049/iet-ipr.2019.0695

[14] S. Laaroussi, A. Baataoui, A. Halli, and K. Satori, “Dynamic mosaicking: combining A* algorithm with fractional Brownian motion for an optimal seamline detection,” IET Image Processing, vol. 14, no. 13, pp. 3169-3180, 2020. https://doi.org/10.1049/iet-ipr.2019.1619

[15] G. Winkelmaier, R. Battulwar, M. Khoshdeli, J. Valencia, J. Sattarvand, and B. Parvin, “Topographically guided UAV for identifying tension cracks using image-based analytics in open-pit mines,” IEEE Transactions on Industrial Electronics, vol. 68, no. 6, pp. 5415-5424, 2021. https://doi.org/10.1109/TIE.2020.2992011

[16] R. Hahn, F. E. Hmmerling, T. Haist, K. Michel, and W. Osten, “Detailed characterization of a mosaic based hyperspectral snapshot imager,” Optical Engineering, vol. 61, no. 1, pp. 15106-15115, 2020. https://doi.org/10.1117/1.OE.59.12.125102

[17] C. J. Legleiter, and P. J. Kinzel, “Improving remotely sensed river bathymetry by image‐averaging,” Water Resources Research, vol. 57, no. 3, pp. 51-76, 2021. https://doi.org/10.1029/2020WR028795

[18] N. Thakur, N. U. Khan, and S. D. Sharma, "A Two-Phase Ultrasound Image De-Speckling Framework by Nonlocal Means on Anisotropic Diffused Image Data," Informatica, vol. 47, no. 2, 2023. https://doi.org/10.31449/inf.v47i2.4378

[19] D. Al-Karawi, S. Jassim, A. Sayasneh, C. Landolfo, D. Timmerman, T. Bourne, and S. Jassim, “An evaluation of the effectiveness of image-based texture features extracted from static b-mode ultrasound images in distinguishing between benign and malignant ovarian masses,” Ultrasonic Imaging, vol. 43, no. 3, pp. 124-138, 2021. https://doi.org/10.1177/0161734621998091

[20] T. Gherbi, A. Zeggari, Z. A. Seghir, and F. Hachouf, "Entropy-guided assessment of image retrieval systems: Advancing grouped precision as an evaluation measure for relevant retrievability," Informatica, vol. 47, no. 7, 2023. https://doi.org/10.31449/inf.v47i7.4661

https://doi.org/10.31449/inf.v48i2.5945 Informatica **48** (2024) 283–284 **283**

Concurrent Consideration of Technical and Human Aspects in Security Requirements Engineering

Damjan Fujs
University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia
E-mail: damjan.fujs@fri.uni-lj.si

Thesis summary

Keywords: information security, software feature, training

Received: March 18, 2024

This article is an extended abstract of the doctoral dissertation entitled "Tailoring security-related software and training requirements to users based on their categorization" [1]. Security has traditionally been ensured by technical solutions in the concluding stages of software development. The fact that security is considered an additional function means that a vulnerability is fixed with security patches as soon as it occurs. However, the importance of human factors is increasingly being recognized, as technical solutions alone are not enough to close security gaps. In order to address this shortcoming, we proposed an approach that simultaneously addresses technical as well as human aspects - already in the initial stages of software development.

Povzetek: Predstavljena je doktorska disertacija z naslovom »Prilagajanje z varnostjo povezanih zahtev za programsko opremo in usposabljanja uporabnikov na podlagi njihove kategorizacije«.

1 Introduction

The number of cyber security threats is increasing. Not only the number of threats, but also in terms of their impact and severity of consequences [4]. Despite the fact that established technical information security mechanisms can be effective, human factors remain one of the key challenges in ensuring information security because people are the weakest link in information security [3].

The main objective of the doctoral dissertation was to investigate how the technical and human aspects of information security can be addressed simultaneously to improve overall information security.

2 Methods

Our approach consists of two main phases and three steps. In the *first* phase, we developed a novel approach for tailoring information security training requirements (iSTR) based on end user categorization according to their different levels of information security performance. In the *second* phase, we developed an approach for balancing information security software requirements (iSSR) and iSTR (See Figure 1). The overall approach is based on existing studies in the field of software requirements engineering, human aspects of information security (user groups), information security standards and security-related software requirements and training.

To test the proposed approach, we conducted an experiment (our main research method) among experienced information system professionals from the wider field of software development (N=128). For the needs of the experiment, we prepared supporting artefacts in which we introduced the basic concepts, the research process and additional explanations to the participants. Participants were randomly assigned to an experimental (N=66) or control group (N=62).

3 Results

The main result shows a clear difference between the two groups in favour of the experimental group. The difference between the two groups is statistically significant (p-value < 0.001). Assessments of evaluators (i.e., experienced experts in the field of information security) show that the requirements created by the experimental group were better than the requirements created by the control group.

In addition, we conducted a post-hoc survey in which we asked the participants of the experiment about three indicators: complexity, compatibility and usefulness. The results are statistically significant (p-value < 0.001) for the usefulness. These results indicate that the participants in the experimental group did not perceive our approach as significantly more complex or less consistent with their knowledge. In addition, participants in the experimental group found our approach significantly more useful than the approach they used in the control group.

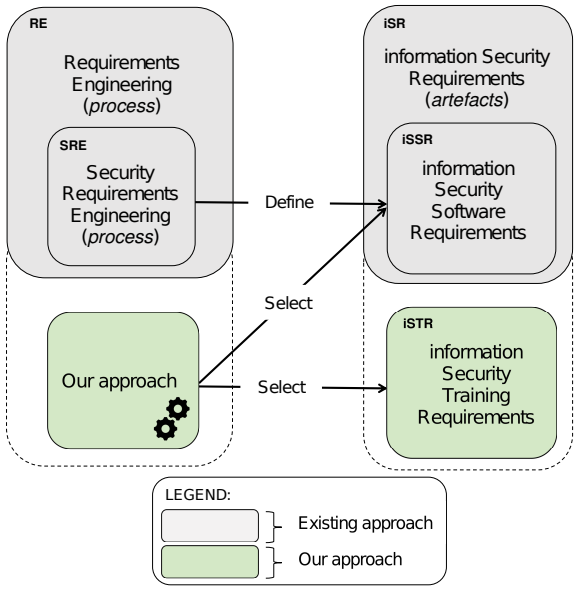


Figure 1: Main research concepts and their relationship. The green color shows our contribution (i.e. our approach that enables balancing iSSR and iSTR), while the grey color represents elements that are already established in the literature. Figure modified from Fujs [1].

4 Conclusion

The following contributions to science in the field of computer and information science are presented in the doctoral dissertation [1]: 1. A novel approach for identifying information security-related training requirements based on end-user categorization of their information security performance (security-related knowledge, attitudes and behaviors). 2. A novel approach that allows iSSR and iSTR to be considered simultaneously based on end user categorization.

Acknowledgements

The author would like to thank the mentors (Damjan Vavpotič and Simon Vrhovec) and committee (Denis Trček, Igor Bernik, Steffen Wendzel and Tomaž Curk). Many thanks to President of the World Federation of Scientists, Prof Antonino Zichichi, for granting a scholarship to support the completion of the author's doctoral studies.

References

[1] Fujs, D. (2024). *Tailoring security-related software and training requirements to users based on their categorization* [Doctoral dissertation]. Repository of the University of Ljubljana.

[2] Fujs, D., Vrhovec, S., & Vavpotič, D. (2023). Balancing software and training requirements for information security. *Computers & security*, 134, 103467. https://doi.org/10.1016/j.cose.2023.103467

[3] Wiley, A., McCormac, A., & Calic, D. (2020). More than the individual: Examining the relationship between culture and Information Security Awareness. Computers & Security, 88, 101640. https://doi.org/10.1016/j.cose.2019.101640

[4] European Union Agency for Cybersecurity, Svetozarov Naydenov, R., Malatras, A., Lella, I., Theocharidou, M., Ciobanu, C., Tsekmezoglou, E. (2022). ENISA threat landscape 2022. https://doi.org/10.2824/764318

Seepage Detection Using Passive Temperature Measurements by Fiber Optic DTS

Yaser Ghafoori
Slovenian National Building and Civil Engineering Institute
E-mail: yaser.ghafoori@zag.si

Thesis summary

Keywords: optical fiber DTS, soil, temperature, seepage

Received: February 22, 2024

This article is an extended abstract of the doctoral dissertation "Optimization of early seepage detection in embankments using a distributed temperature system based on fiber optic sensing" [1]. Distributed Temperature Sensors (DTS) measure temperature precisely with high spatial resolution, enabling seepage flow detection through the interpretation of temperature measurement. Active DTS measurement which introduces heat into the soil to monitor its dissipation, has been extensively studied for seepage detection. However, the potential of passive measurement, which measures natural soil temperature, for seepage detection demands further exploration. The doctoral research investigated the capability of passive DTS measurements for seepage detection. The findings show that passive DTS detects very small temperature variations as a result of the seepage flow. Additionally, the results offer valuable insights into the coupled hydro-thermal behavior and the influence of seepage on heat transfer. The thesis classified and compared different techniques for temperature data calibration and interpretation. Furthermore, an interpretation approach based on numerical simulation and a fully dynamic calibration technique was developed.

Povzetek: Predstavljena je doktorska disertacija z naslovom »Optimizacija postopka zgodnjega zaznavanja pojava precejanja vode v nasipih na podlagi spremljanja temperaturnega polja z optičnimi kabli«.

1 Introduction

The last decade has witnessed rapid advancements in the field of optical fiber Distributed Temperature Sensors (DTS), facilitating continuous temperature monitoring of various infrastructures. Optical fiber DTS utilizes Raman backscattering to quantify temperature along the entire length of the optical fiber. Despite the active method's capability for seepage detection by introducing heat into the soil and monitoring its dissipation, its functionality for large-scale monitoring is limited. This limitation is due to the high heating power required, which not only increases costs but also limits the measurement length. Unlike the active method, the passive method which measures the soil's natural temperature, is more suitable for long-life, large-scale coverage, and cost-effective monitoring systems.

The primary objective of the doctoral thesis was to explore the potential of passive DTS measurements for accurate and early seepage flow detection. The efficiency of DTS employment for seepage detection relies on application methods, measurement reliability, and data interpretation. Therefore, the other aim of this research was to enhance seepage detection through the optimization of data calibration and interpretation.

2 Method and results

2.1 Passive DTS performance evaluation

A series of laboratory tests were conducted on different soils subjected to transient seepage flow [1], [2], [3]. The soil temperature was measured by passive DTS optical fiber using a developed distribution configuration composed of optic loops allowing precise point-based measurements. The system's spatial resolution and response time, measured in laboratory experiments, were 64 cm and 60 seconds respectively. The accuracy of DTS temperature measurements was confirmed by the embedded thermometers. The hydraulic properties of the materials were determined and the corresponding coupled hydrothermal numerical simulations were conducted.

The findings demonstrated that passive DTS can detect the initiation of seepage and trace its flow by precise temperature measurements, even when there is a minor difference (about 3 degrees) between the temperature of the soil and the water [1].

Experimental observations and numerical modeling revealed a correlation between temperature changes and increases in degree of saturation, with the latter preceding heat flow.

As the degree of saturation increased, the temperature within the sand changed gradually due to the increased thermal conductivity caused by the presence of water in the soil pores. However, once the soil reached full saturation, temperature changes occurred dramatically due to the convection heat transfer caused by seepage flow [3]. The results indicated that within soil subjected to seepage flow, convection is the dominant heat transfer process, even at a small Darcy velocity on the order of 10^{-5} m/s. The temperature distribution within the soil was used to estimate the phreatic line, which showed reliable agreement with experimental and numerical calculations.

Experimental measurements indicated that the thermal velocity is greater within sand having a higher hydraulic conductivity, resulting in faster Darcy velocity. Minor variations in the soil's hydraulic conductivity impacted heat transfer, while the thermal conductivity of the grain (solid) had an almost negligible effect on heat transfer within the soil subjected to seepage flow [3], [4]. Long-term temperature measurements were taken after the conclusion of the seepage experiment. Once the water had drained from the soil, the temperature within the sand increased [1].

2.2 Optimization of DTS application

To extract seepage-related information from the temperature measurements, different data interpretation techniques were compared and classified. This comprehensive review can guide the selection of an appropriate method for tracing seepage at the early stages. The numerical simulation in the studies accurately predicted the seepage and thermal distribution. Having this, an interpretation technique was proposed to utilize the simulated numerical results for temperature data interpretation. A well-defined coupled hydrothermal analysis provides essential information to identify and localize potential unexpected seepage behavior and its impact on temperature distribution [1], [5]. Additionally, the monitoring system is optimized for cost and efficiency through the installation of fiber based on the seepage risk magnitude determined by numerical calculations.

The efficiency of the DTS application is strongly dependent on the precision of temperature measurements that can be enhanced through various calibration techniques. A comparative review of calibration methods was provided based on the accuracy, installation approaches, and complexity of the application [6]. Using this review, the proper calibration technique can be selected before the cable installation based on the measurement requirements and the site's availability.

The doctoral research on DTS calibration was expanded to develop a fully dynamic calibration technique, aimed to overcome the limitations of conventional methods [7]. The embedded temperature loggers within the soil were used as references for DTS measurements and a calibration method was developed by the model-independent Parameter ESTimation (PEST) tool. The developed calibrated technique improved the accuracy of DTS measurements by reducing the absolute bias of temperature measurements from 0.27 °C to 0.08 °C. A significant advantage of this method is its easy installation process, which eliminates the need for a constant temperature reference. Furthermore, the number of reference points can be easily expanded by using small-sized thermometers to improve the calibration.

3 Conclusion

The doctoral dissertation findings showed that passive optical fiber DTS provides precise temperature measurements, facilitating early seepage flow detection, even at low-temperature differences between soil and water. The study demonstrated the significant effect of seepage flow in heat transfer through the convection process and utilized temperature measurements to estimate the soil's phreatic line. The doctoral thesis offered a comprehensive review of calibration and interpretation methods, aiding in the selection of suitable techniques based on measurement requirements and site conditions. The thesis introduced a method for interpreting temperature data using a coupled hydro-thermal simulation. This approach allows the prediction and assessment of seepage and thermal behavior. The thesis results led to the development of a dynamic calibration method utilizing external thermometers to establish reference points. This method improved the measurement accuracy and simplified optical fiber installation.

References

[1] Ghafoori Y (2023), Optimization of early seepage detection in embankments using a distributed temperature system based on fiber optic sensing: Ph.D. Thesis. *University of Ljubljana*.

[2] Ghafoori Y, Maček M, Vidmar A, Říha J, Kryžanowski A (2020). Analysis of seepage in a laboratory scaled model using passive optical fiber distributed temperature sensor. *Water*, 12, 367. https://doi.org/10.3390/w12020367.

[3] Ghafoori Y, Maček M, Vidmar A, Říha J, Kryžanowski A (2021). Heat transfer by seepage in sand: Influence of saturated hydraulic conductivity and porosity. *Acta Hydrotechnica*, 34, pp. 61–75. https://doi.org/10.15292/acta.hydro.2021.05.

[4] Ghafoori Y, Maček M, Vidmar A, Říha J, Kryžanowski A (2020). Monitoring of temperature distribution by optical fiber DTS within soil subjected to seepage flow. *Second SLOCOLD -MACOLD Symposium on topic Water reservoirs*, Skopje, Republic of N. Macedonia, p. 77–84.

[5] Ghafoori Y, Logar J, Říha J, Kryžanowski A (2021). Numerical modeling for identification of seepage by thermal monitoring in the embankment of Brežice HPP. *Macedonian Committee on Large Dams (5th Congress on Dams)*, Struga, Republic of N. Macedonia: MACOLD, p. 99–105.

[6] Ghafoori Y, Vidmar A, Říha J, Kryžanowski A (2020). A review of measurement calibration and interpretation for seepage monitoring by optical fiber distributed temperature sensors. *Sensors*, 20(19):5696. https://doi.org/10.3390/s20195696.

[7] Ghafoori Y, Vidmar A, Kryžanowski A (2022). A Dynamic Calibration of Optical Fiber DTS Measurements Using PEST and Reference Thermometers. *Sensors*; 22(10):3890. https://doi.org/10.3390/s22103890.

An Efficient Iterative Algorithm to Explainable Feature Learning

Dino Vlahek
Faculty of Electrical Engineering and Computer Science at the University of Maribor (UM FERI)
E-mail1: dino.vlahek1@um.si

Thesis Summary

Keywords: data classification, explainable artificial intelligence, feature learning

Received: April 23, 2024

This paper summarizes a doctoral thesis introducing the new iterative approach to explainable feature learning. Features are learned in three steps during each iteration: feature construction, evaluation, and selection. We demonstrated superior performances compared to the state of the art on 13 of 15 test cases and the explainability of the learned feature representation for knowledge discovery.

Povzetek: To delo povzema vsebino doktorske disertacije, v kateri predstavimo iterativni pristop k učenju razložljivih značilnic. Med vsako iteracijo se značilnice naučijo čez naslednje korake: gradnja, ocenjevanje in izbira značilnic. Na 13 od 15 testnih primerov smo demonstrirali vrhunsko zmogljivost v primerjavi s stanjem tehnike in razložljivost predstavitve naučenih značilnic za odkrivanje znanja.

1 Introduction

Supervised feature learning describes a set of techniques that enable defining augmented data representation for improved utilization of classification or regression models [1]. These methods replace traditional feature engineering tasks in a wide range of machine learning applications. Supervised feature learning methods can be divided into feature selection, dimensionality reduction, supervised dictionary learning, and deep learning. Feature selection methods select a subset of relevant features from the original feature space [3]. Such methods are limited in their accuracy as they cannot recombine features. In contrast, supervised dimensionality reduction recombines input features by mapping input samples on linear or non-linear manifolds [4]. However, significant distortions may be introduced to the data by this process as a consequence of changing distances between learning samples. Thus, resulting classification models are challenging to interpret. In addition, these approaches can only reduce the feature space's dimensionality [4]. On the other hand, supervised dictionary learning learns new feature space from the input set by recombining an arbitrary number of basic elements, called atoms, that compose a dictionary [2]. It is considered an optimization problem, where the sparsity of representation is maximized and the reconstruction errors minimized. Latter is defined as the difference between learning data and sparse representation. Dictionaries can be shared and class-specific depending on the mechanism for processing discriminatory information. Shared ones are learned from the entire data set, regardless of class labels. Using such dictionaries requires an additional classifier that significantly increases computational complexity due to the non-convex optimization problem [2]. On the other hand, class-specific dictionaries are learned for each class separately [6], enabling straightforward classification of unknown samples based on the reconstruction error introduced by such dictionaries. However, this can become computationally demanding with the increasing number of classes, while it is challenging to extract useful knowledge when the dictionary contains a large number of atoms [2]. Similar drawbacks are also noted when considering deep learning approaches. These are based on various architectures of artificial neural networks with multiple hidden layers of neurons that allow for extracting higher-level features progressively from the raw input [5]. Both linear and nonlinear functions can model neurons' activation functions, thus optimizing feature representation within the decision function. By increasing the hidden layers, artificial neural networks can approximate increasingly complex decision functions and achieve high classification accuracies. However, the presence of multiple local optima and many hyperparameters [5] also increases the training procedure's complexity, while we consider these methods as black-box function approximators [1]. In order to address above-mentioned challenges, a new method is proposed in [7] that learned interpretable features from input ones that achieved improved accuracy in comparison to the current state-of-the-art.

2 Methodology

The proposed method [7] allows for exploiting non-linear codependencies between features in order to improve an arbitrary classifier's classification performance while providing a meaningful feature representation for knowledge discovery. Each iteration consists of the following three steps: Feature construction that generates the new feature space,

feature evaluation that assesses the quality of the individual feature by a new metric that defines the feature's suitability for classification tasks, and feature selection that selects the high-quality dissimilar features using a new method based on vertex-cut. Here, we introduce two input parameters used to define the graph. The first represents the necessary level of features' quality to be included in the output feature space, and the second determines the minimal level of dissimilarity between them.

3 Results and discussion

The proposed method is extensively tested on fifteen benchmark datasets. During the sensitivity analysis, optimal values of two input parameters were identified, and the performance of five traditional classifiers was estimated on learned features. The study showed that the learned features statistically significantly improved the classification accuracy of all tested classifiers, while the random forest classifier achieved the best results. As demonstrated by experiments, the proposed method achieved or exceeded the classification accuracy of six state-of-the-art in all test cases. The correctness of learned features interpretation on a well-studied dataset was also demonstrated.

The proposed method is used in many research applications, ranging from pure research to industrial and scientific projects. We plan to extend the proposed method application to regression with a new feature evaluation metric for the suitability of features for regression tasks.

References

[1] Y. Bengio, A. Courville, and P. Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, aug 2013. doi: https://doi.org/10.1109/TPAMI.2013.50.

[2] Mehrdad J. Gangeh, Ahmed K. Farahat, Ali Ghodsi, and Mohamed S. Kamel. Supervised dictionary learning and sparse representation-a review. *ArXiv*, abs/1502.05928, 2015. doi: https://doi.org/10.48550/arXiv.1502.05928.

[3] Huan Liu and Hiroshi Motoda. *Feature Extraction, Construction and Selection: A Data Mining Perspective*. Kluwer Academic Publishers, Norwell, MA, USA, 1998.

[4] Yunqian Ma and Yun Fu. *Manifold Learning Theory and Applications*. CRC Press, Inc., USA, 1st edition, 2011.

[5] Michael A. Nielsen. *Neural Networks and Deep Learning*, page 216. Determination Press, 2018.

[6] W. Tang, A. Panahi, H. Krim, and L. Dai. Analysis dictionary learning based classification: Structure for robustness. *IEEE Transactions on Image Processing*, 28(12):6035–6046, Dec 2019. doi: http://doi.org//10.1109/TIP.2019.2919409.

[7] Dino Vlahek. Učinkovit iterativni algoritem učenja razložljivih značilnic za izboljšano klasifikacijo. *PhD thesis, UM FERI*, 2024.

Advancing Cadastral Mapping with UAVs and Automated Boundary Delineation

Bujar Fetai
University of Ljubljana, Faculty of civil and geodetic engineering, Slovenia
E-mail: bujar.fetai@fgg.uni-lj.si

Thesis summary

Keywords: cadastre, land, boundary, detection, UAV, deep learning

Received: July 26, 2024

Visible land boundaries allow for automatic detection using remote sensing data with optical sensors. The dissertation aimed to improve cadastral mapping using unmanned aerial vehicle (UAV) photogrammetry. The aim was to evaluate the accuracy of cadastral data concerning land boundaries and to develop an automated approach for delineating these boundaries. The data captured by UAVs was analyzed to identify discrepancies between physical (visible) and formal (cadastral) land boundaries. The process includes boundary detection, geo-referencing, evaluation of up-to-dateness, and vectorization of the predicted boundary maps. Initially, image processing methods were tested for automatic detection. Subsequently, deep learning methods were used to improve the detection process using UAV data. Manual delineations were also carried out to validate and assess the accuracy of the automated detections. Different approaches and methods were tested in case studies, especially in rural areas where visible land boundaries are more common. Although primarily tested with drones, it can also be adapted to satellite or aerial imagery and provides a cost-effective way to detect and revise cadastral maps. Automatic detection identifies areas needing cadastral updating and is supported by manual verification to ensure accuracy.

Povzetek: Doktorska disertacija preučuje izboljšanje katastrskih načrtov z uporabo UAV fotogrametrije. To je razširjeni povzetek disertacije, katere cilj je bil raziskati neskladja med katastrskimi in dejanskimi mejami ter razviti pristop za posodobitev obstoječih katastrskih načrtov na podlagi podatkov iz UAV fotogrametrije.

1 Introduction

Efforts to establish cadastral systems have primarily targeted countries with low cadastral coverage. However, maintaining and updating land data in countries with complete cadastral systems has received comparatively less attention. In countries with complete cadastres, scanning and vectorizing analog maps created the digital ownership layer. In such cases, discrepancies between physical and formal land boundaries may exist due to the dynamic nature of cadastres and changes in land boundaries over time [1,2].

Detection of physical or visible boundaries using remote sensing, particularly UAV imagery due to its accuracy and flexibility, opens new possibilities. While many case studies report manual delineation of land boundaries, it is argued that a significant number of cadastral boundaries are visible and coincide with natural or man-made physical features. Automatic detection approaches, such as image processing and deep learning, can be applied to detect these land boundaries from images [3,4].

The main objective of this doctoral dissertation [1] was to design a UAV-based cadastral mapping approach to evaluate the currency of land boundary data and update cadastral maps.

2 Methods and results

2.1 Inconsistencies in land boundaries

The study [2] presents an approach to identify inconsistencies in the cadastral boundary data and evaluate the digitization process.

The focus was on the importance of survey-based cadastres and the digitization process. This is illustrated with examples from Slovenia and North Macedonia, both of which have complete cadastral coverage. Challenges in digitization and boundary definition were identified, with significant differences in horizontal accuracy (about 0.5 meters) and area deviations (about 25 square meters).

A systematic workflow was designed for the digitization of analog cadastral data and the resolution of discrepancies in land data, classified into the following categories: i) inconsistencies between cadastral and land registry data, ii) within cadastral data, and iii) between cadastral data and the situation on the ground. Furthermore, we present a model for cadastral data maintenance to identify inconsistencies between physical and cadastral boundaries using UAV data.

2.2 Delineation approaches of visible land boundaries

Land boundaries can be delineated manually or automatically from UAV data, with automation ranging from semi-automatic to fully automatic. This research developed a general workflow to investigate UAVs in cadastral mapping. Manually vectorized land boundaries served as a benchmark to evaluate the quality and reliability of automatic delineation. We quantitatively assessed the discrepancies by overlaying the automatic and existing cadastral boundaries [3-5].

We tested several classic edge-detection and segmentation algorithms, with Envi FX outperforming the others by providing vector-format parcel boundaries without additional image processing. However, further simplification was necessary for cadastral applications, and fully automating the process remains challenging [3]. Semi-automatic approaches are required, particularly in complex urban areas.

The study also compared UAV-based mapping to traditional ground-based methods. When boundaries were delineated manually, UAV orthophotos demonstrated high planar accuracy, with an RMSE of 1 cm, compared to total station measurements [1].

2.3 Deep learning for land boundary delineation

Upscaling deep learning solutions, such as convolutional neural networks (CNNs), for visible land boundary detection is becoming increasingly important, especially for UAV data [4].

CNNs generally offer higher accuracy in delineating land boundaries than traditional image processing methods. However, collecting and processing thousands of UAV images for boundary delineation is challenging, even with data augmentation [4]. To address this, we used publicly available datasets like the Berkeley Segmentation Dataset and Benchmark (BSDS500), a standard for image segmentation and edge detection tasks. We applied this dataset to detect land boundaries in UAV imagery [5].

Model performance during training was monitored using the F1 score. High overall accuracy was not suitable due to the imbalance in class pixels—boundary pixels are much fewer than background pixels. The U-Net architecture was the most promising for our domain's complexity. Additionally, we customized U-Net to reduce the number of trainable parameters, which proved sufficient for our needs. The model was trained in Google Collaboratory, and implemented in Keras. The process was written in Python using the TensorFlow library.

The trained model was tested using UAV images sized 256 × 256 pixels. Determining a perfect F1 score for cadastral mapping was challenging since not all cadastral boundaries are visible, and testing scenes differ, making reliable comparisons difficult. An F1 score of 60 is considered sufficient for rapid analysis, especially in rural areas.

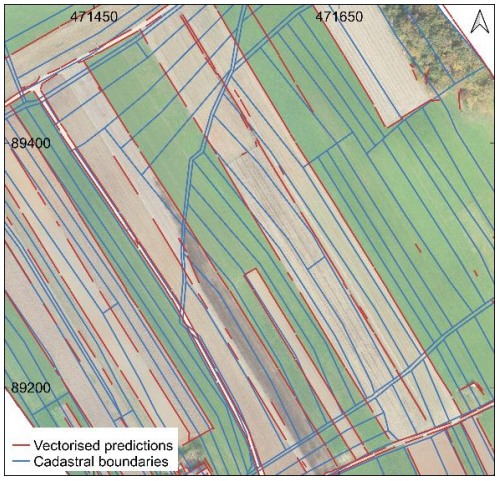


Figure 1. Example of map revision: Automated boundary delineation (red) vs. cadastral boundaries (blue).

The predicted boundary maps were then georeferenced, merged, and vectorized to comply with cadastral boundary formats. With this approach, we were also able to revise existing cadastral maps (Figure 1).

3 Conclusions

The research addressed the challenge of maintaining existing cadastral maps by proposing an automated approach. The workflow consisted of training a CNN to detect land boundaries from UAV imagery, georeferencing the predictions and vectorizing them to revise cadastral maps. This method, which has been tested with UAV imagery, could also be applied to satellite or aerial imagery to efficiently identify areas that need to be updated. It is important to note that automatically delineated boundaries are not final boundaries, but can be considered provisional, which increases the efficiency of cadastral mapping. Nevertheless, manual delineations hardly differ from ground-based methods in terms of positional accuracy and can still be used for verification and refinement.

References

[1] Fetai, B. “Izboljšava katastrskih načrtov z uporabo daljinsko vodenih zrakoplovov”, doktorska disertacija, Univerza v Ljubljani, 2023.

[2] Fetai, B.; Tekavec, J.; Fras, M.K.; Lisec, A. Inconsistencies in Cadastral Boundary Data—Digitisation and Maintenance. *Land* 2022, *11*, 2318. https://doi.org/10.3390/land11122318

[3] Fetai, B.; Oštir, K.; Kosmatin Fras, M.; Lisec, A. Extraction of Visible Boundaries for Cadastral Mapping Based on UAV Imagery. *Remote Sens.* 2019, 11, 1510.

[4] Fetai, B.; Račič, M.; Lisec, A. Deep Learning for Detection of Visible Land Boundaries from UAV Imagery. *Remote Sens.* 2021, *13*, 2077.

[5] Fetai, B.; Grigillo, D.; Lisec, A. Revising Cadastral Data on Land Boundaries Using Deep Learning in Image-Based Mapping. *ISPRS Int. J. Geo-Inf.* 2022, *11*, 298.

JOŽEF STEFAN INSTITUTE

Jožef Stefan (1835-1893) was one of the most prominent physicists of the 19th century. Born to Slovene parents, he obtained his Ph.D. at Vienna University, where he was later Director of the Physics Institute, Vice-President of the Vienna Academy of Sciences and a member of several sci- entific institutions in Europe. Stefan explored many areas in hydrodynamics, optics, acoustics, electricity, magnetism and the kinetic theory of gases. Among other things, he originated the law that the total radiation from a black body is proportional to the 4th power of its absolute tem- perature, known as the Stefan–Boltzmann law.

The Jožef Stefan Institute (JSI) is the leading independent scientific research institution in Slovenia, covering a broad spectrum of fundamental and applied research in the fields of physics, chemistry and biochemistry, electronics and information science, nuclear science technology, en- ergy research and environmental science.

The Jožef Stefan Institute (JSI) is a research organisation for pure and applied research in the natural sciences and technology. Both are closely interconnected in research de- partments composed of different task teams. Emphasis in basic research is given to the development and education of young scientists, while applied research and development serve for the transfer of advanced knowledge, contributing to the development of the national economy and society in general.

At present the Institute, with a total of about 900 staff, has 700 researchers, about 250 of whom are postgraduates, around 500 of whom have doctorates (Ph.D.), and around 200 of whom have permanent professorships or temporary teaching assignments at the Universities.

In view of its activities and status, the JSI plays the role of a national institute, complementing the role of the uni- versities and bridging the gap between basic science and applications.

Research at the JSI includes the following major fields: physics; chemistry; electronics, informatics and computer sciences; biochemistry; ecology; reactor technology; ap- plied mathematics. Most of the activities are more or less closely connected to information sciences, in particu- lar computer sciences, artificial intelligence, language and speech technologies, computer-aided design, computer architectures, biocybernetics and robotics, computer automa- tion and control, professional electronics, digital communi- cations and networks, and applied mathematics.

The Institute is located in Ljubljana, the capital of the in dependent state of S**love**nia (or S♡nia). The capital today isconsidered a crossroad bet between East, West and Mediter-ranean Europe, offering excellent productive capabilities and solid business opportunities, with strong international connections. Ljubljana is connected to important centers such as Prague, Budapest, Vienna, Zagreb, Milan, Rome, Monaco, Nice, Bern and Munich, all within a radius of 600 km.

From the Jožef Stefan Institute, the Technology park "Ljubljana" has been proposed as part of the national strat- egy for technological development to foster synergies be- tween research and industry, to promote joint ventures be- tween university bodies, research institutes and innovative industry, to act as an incubator for high-tech initiatives and to accelerate the development cycle of innovative products.

Part of the Institute was reorganized into several high-tech units supported by and connected within the Technology park at the Jožef Stefan Institute, established as the beginning of a regional Technology park "Ljubljana". The project was developed at a particularly historical moment, characterized by the process of state reorganisation, privatisation and private initiative. The national Technology Park is a shareholding company hosting an independent venture-capital institution.

The promoters and operational entities of the project are the Republic of Slovenia, Ministry of Higher Education, Science and Technology and the Jožef Stefan Institute. The framework of the operation also includes the University of Ljubljana, the National Institute of Chemistry, the Institute for Electronics and Vacuum Technology and the Institute for Materials and Construction Research among others. In addition, the project is supported by the Ministry of the Economy, the National Chamber of Economy and the City of Ljubljana.

Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel.:+386 1 4773 900, Fax.:+386 1 251 93 85
WWW: http://www.ijs.si
E-mail: matjaz.gams@ijs.si
Public relations: Polona Strnad

Informatica

An International Journal of Computing and Informatics

Web edition of Informatica may be accessed at: http://www.informatica.si.

Subscription Information Informatica (ISSN 0350-5596) is published four times a year in Spring, Summer, Autumn, and Winter (4 issues per year) by the Slovene Society Informatika, Litostrojska cesta 54, 1000 Ljubljana, Slovenia.
The subscription rate for 2024 (Volume 48) is
– 60 EUR for institutions,
– 30 EUR for individuals, and
– 15 EUR for students
Claims for missing issues will be honored free of charge within six months after the publication date of the issue.

Typesetting: Blaž Mahnič, Gašper Slapničar; gasper.slapnicar@ijs.si
Printing: ABO grafika d.o.o., Ob železnici 16, 1000 Ljubljana.

Orders may be placed by email (drago.torkar@ijs.si), telephone (+386 1 477 3900) or fax (+386 1 251 93 85). The payment should be made to our bank account no.: 02083-0013014662 at NLB d.d., 1520 Ljubljana, Trg republike 2, Slovenija, IBAN no.: SI56020830013014662, SWIFT Code: LJBASI2X.

Informatica is published by Slovene Society Informatika (president Niko Schlamberger) in cooperation with the following societies (and contact persons):
Slovene Society for Pattern Recognition (Vitomir Štruc)
Slovenian Artificial Intelligence Society (Sašo Džeroski)
Cognitive Science Society (Olga Markič)
Slovenian Society of Mathematicians, Physicists and Astronomers (Dragan Mihailović)
Automatic Control Society of Slovenia (Giovanni Godena)
Slovenian Association of Technical and Natural Sciences / Engineering Academy of Slovenia (Mark Pleško)
ACM Slovenia (Ljupčo Todorovski)

Informatica is financially supported by the Slovenian research agency from the Call for co-financing of scientific periodical publications.

Informatica is surveyed by: ACM Digital Library, Citeseer, COBISS, Compendex, Computer & Information Systems Abstracts, Computer Database, Computer Science Index, Current Mathematical Publications, DBLP Computer Science Bibliography, Directory of Open Access Journals, InfoTrac OneFile, Inspec, Linguistic and Language Behaviour Abstracts, Mathematical Reviews, MatSciNet, MatSci on SilverPlatter, Scopus, Zentralblatt Math

Volume 48 Number 2 August 2024 **ISSN 0350-5596**

Informatica

An International Journal of Computing and Informatics

Informatica **48** (2024) Number 2, pp. 145–294