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.

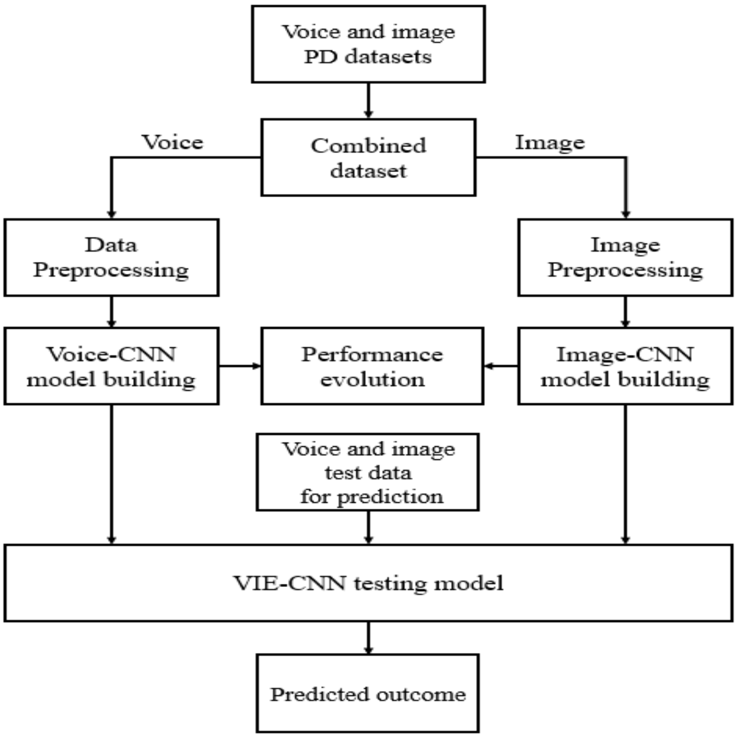


Figure 1: Proposed ParkinsonNet block diagram.

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.

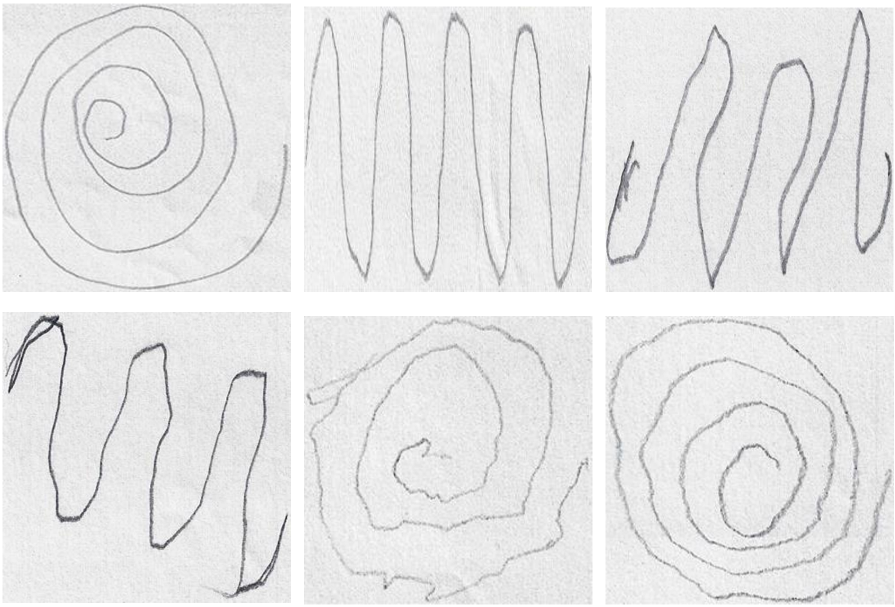


Figure 2: Sample images from the Image-PD dataset.

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.

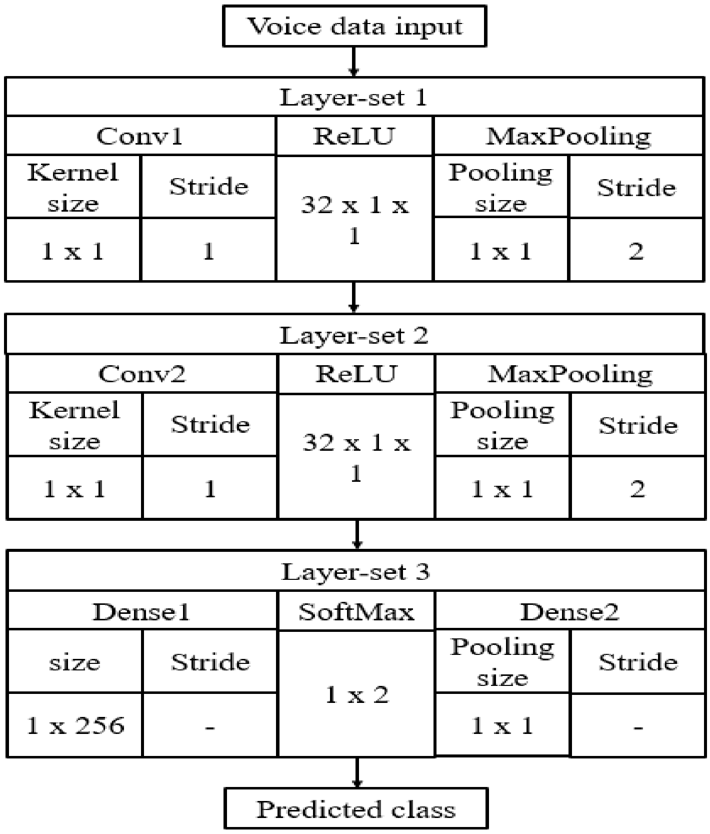


Figure 3: Voice-CNN model.

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).

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

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

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.

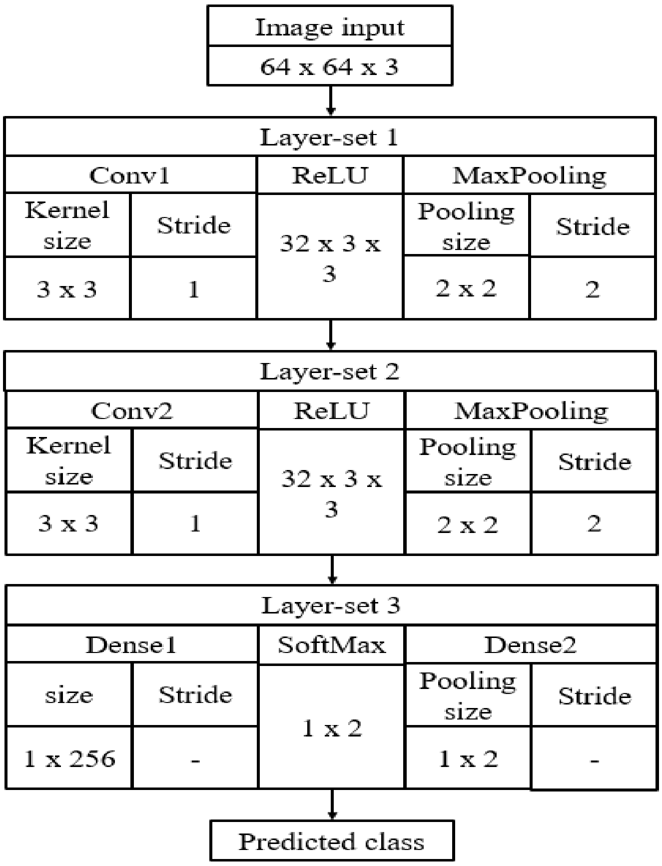


Figure 4: Image-CNN model.

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.

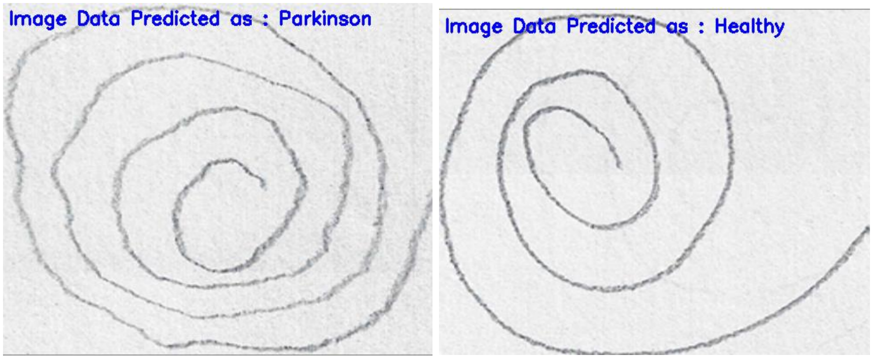
SoftMax classifier: SoftMax Classifier is also an activation function for activating hyperbolic tangents. It's a form of logistic sigmoid activation function, and it has a significant meaning in terms of the biological neurons. One of the most distinguishing features of a SoftMax tangent function is 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].

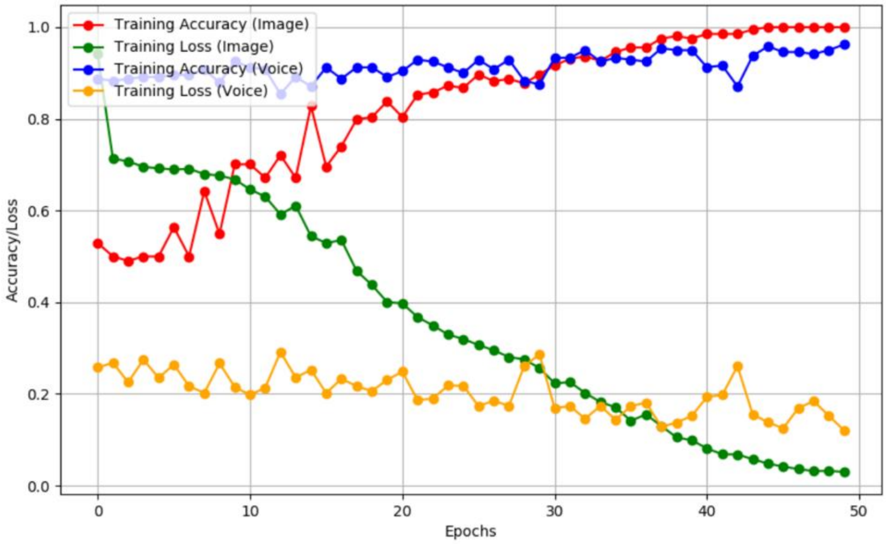


Figure 6: Accuracy and Loss outcomes using ParkinsonNet.

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.