

Facial Sentiment Analysis Using Convolutional Neural Network and Fuzzy Systems

Ahmed R. Kadhim¹, Raidah S. Khudayer¹, Maytham Alabbas²

¹Department of Computer Information System, College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq

²Department of Computer Science, College of Computer Science and Information Technology, University of Basrah, Basrah, Iraq

E-mail: ahmed07727043448@gmail.com, raidah.khudayer@uobasrah.edu.iq, ma@uobasrah.edu.iq

Keywords: fuzzy neural networks systems, convolutional neural networks (CNNs), facial expression recognition

Received: May 5, 2024

This study provides a detailed study of a Convolutional Neural Network (CNN) model optimized for facial expression recognition with Fuzzy logic using Fuzzy2D Pooling and Fuzzy Neural Networks (FNN), and discusses data augmentation in model optimization. It highlights important roles. performance. First, the effectiveness of the models in classifying emotions from FER2013, RAB-DB, and CK+ datasets was evaluated by a 5-fold cross-validation method, which showed that the accuracy varied widely among different emotion classes and was affected by overfitting. It turned out to be easy. The integration of data augmentation techniques, including random rotation, translation, and inversion, significantly improved the model's generalization capabilities. This was evidenced by higher accuracy and more consistent loss curves observed across all folds. After augmentation, the model showed significant improvement, achieving average test accuracies of 98.95% on FER2013, 99.99% on RAF-DB, and 100% on CK+ across all folds. Despite these advances, challenges specific to certain classes of emotions remain, highlighting the need for continued model refinement. This study concludes that data augmentation is an important step in developing robust facial expression recognition models and has potential benefits for a variety of applications requiring accurate emotion recognition.

Povzetek: V tej študiji so uporabljene in izboljšane konvolucijske nevronske mreže z mehko logiko, da se poveča njihova učinkovitost na področju analize obraznih čustev.

1 Introduction

In recent years, the field of artificial intelligence has become one of the most important fields of life because of the role it plays in different fields, which includes many different areas, the most important of which is pattern recognition. Pattern recognition aims to make it possible to analyze and define different patterns in data, which are often complex, as it enables algorithms that work in the field of artificial intelligence to extract valuable information. Pattern recognition [1] represents an important part of modern artificial intelligence systems. It also attempts to identify patterns and create a simulation of the human brain, which contributes to the advancement of artificial intelligence and making the most of complex data, as this data can be sound, image, text, or even video. There are many reasons why pattern recognition is important, the most important of which is that it predicts the simplest parts of data that cannot be tracked by classifying unseen data. Pattern recognition can be divided into three distinct models: Statistical Pattern Recognition, Syntactic Pattern Recognition, and Neural Pattern Recognition [2]. Statistical Pattern Recognition this type of pattern recognition involves studying and identifying patterns in historical statistical data, learning from examples, and then collecting

observations until the model is able to generalize to apply the observations to previously unseen data [3]. Syntactic Pattern Recognition [4] because it is based on simpler sub-patterns known as primitives, this concept is also known as structural pattern recognition. For instance, words are included in this category. The primitives' relationships are characterized as the pattern. For example, primitive words join to build sentences and messages. Neural Pattern Recognition artificial neural networks are used in this model [5]. After learning intricate nonlinear input-output relations, the networks adjust according to available data. Large parallel computing systems comprised of numerous fundamental processors and their connections are used in this concept. They are able to use sequential training procedures take in complex nonlinear input-output conversations, and then adapt themselves to correspond with the data. There are two different machine learning and pattern recognition algorithms: Supervised Algorithms and Unsupervised Algorithms [6]. Supervised Algorithms another name for supervised algorithms is classification. This algorithm employs a two-step process to identify patterns. The development and construction of the model are covered in the first step. The prediction of newer or unseen objects is covered in the second step. Unsupervised Algorithms a "group by" strategy is

preferred by unsupervised algorithms. In order to provide predictions, these algorithms look for patterns in the data and classify them based on similarity, such as dimensions. Pattern recognition has a wide range of applications: Image Recognition, Text Pattern Recognition, Fingerprint Scanning, Seismic Activity Analysis, Audio and Voice Recognition, social media, Cybersecurity and many others [7]. In this paper will focus on Image Recognition. Today, security and surveillance systems from several industries use image recognition tools. These gadgets record and keep an eye on several video streams simultaneously. It helps in identifying possible attackers. Business centers, information technologies companies, and production facilities use the same image recognition technology as face ID systems. Facial Expression Recognition (FER) system presents another corollary of the same application. Here, human emotions of an audience are analyzed and detected in real-time through the application of pattern recognition to video and image data [8]. Sentiment analysis, intent, and mood recognition are the main goals of these systems. Deep learning algorithms are thus employed to identify patterns in people's body language and facial emotions. Organizations can utilize this data to improve client experience by fine-tuning their marketing initiatives. Facial Expression Recognition is a technology used for analyzing sentiments by different sources, such as images and videos. It is a member of the family of technologies known as "affective computing," which draws extensively from Artificial Intelligence technologies [9]. Affective computing is an intersecting area of research on computers' capacity to recognize and interpret human emotions and affective states. Facial expressions are kinds of non-verbal expression, giving suggests for human emotions. Psychology (Ekman and Friesen 2003; Lang et al. 1993) and specialists in the field of human-computer interaction (Cowie et al. 2001; Abdat et al. 2011) have invested decades researching how to comprehend these indications of emotion. The widespread adoption of cameras and the most recent developments in machine learning, biometric analysis, and pattern recognition have all contributed significantly to the FER technology's development.

Human interactions are a tapestry woven with the threads of spoken words, physical gestures, and a rich spectrum of facial expressions. As we navigate through our daily lives, our faces unwittingly broadcast a myriad of emotions, communicating non-verbally with a complexity that language alone cannot capture. In this intricate dance of social interaction, technology has the potential to become a transformative partner, unlocking a deeper understanding of human sentiment. By harnessing advanced artificial intelligence, particularly in the realm of facial emotion recognition, we stand on the brink of an era where machines can not only 'see' but can also 'comprehend' the silent language of our emotions. This leap forward offers profound implications for personalized communication, tailored services, and empathetic machine-human interfaces. However, this technological pursuit is not

without its challenges [10]. The endeavor to translate the transient and often ambiguous canvas of the human face into a digital lexicon of emotions entails a nuanced recognition of the interplay between facial muscle movements and their corresponding emotional states. It requires an algorithmic sensitivity to the context and cultural underpinnings that shape emotional expression. As researchers and engineers strive to bridge this gap, they grapple with the complexities of creating systems robust enough to interpret the subtle signals of our emotive expressions in real-time and in the uncontrolled, diverse settings of our natural environments. The journey from the theoretical understanding of facial expressions to practical, real-world application is the specific focus of the proposed model, aiming to elevate the capability of machines to interpret human emotions with unprecedented accuracy and sensitivity.

2 Related work

Facial Expression Recognition serve as a universal medium for people to convey emotions. This universality has spurred interest in various sectors like robotics, healthcare, and driving assistance systems, where facial expression analysis tools, often based on image processing, are being actively developed. Table 1 shows related work on processing three datasets: FER2013, RAF-DB and CK + datasets.

In academic research, the FER2013 dataset has been a focal point for several sentiment analysis studies. One study [11] utilized Random Search algorithm, initially achieving a 72.16% accuracy rate using the FER2013 dataset. Another research [12] involved designing VGGNet architecture, fine-tunes its hyper parameters model for sentiment analysis, reaching an accuracy of 73.28%. Additionally, a different study [13] achieved a 72.81% accuracy rate by training a Fuzzy optimized CNN-RNN architecture with the FER2013 dataset. This particular study method for facial expression recognition achieved a certain improvement in the recognition effect of different facial expression datasets compared to current popular algorithms. [14] in sentiment analysis, further highlighting the potential of these technologies in various application areas using extraction of multi-layer representation information using asymmetric region local binary pattern (AR-LBP) and divided local directional pattern (DLDP) which achieved accuracy 91%. In the realm of facial expression recognition using deep learning, several studies have made significant strides using the FER2013 dataset. A study referenced in [15] developed an "ConvNet," utilizes a four-layer convolutional neural network (CNN) architecture for facial emotion recognition that, after being trained for minimal number of epochs on the FER2013 dataset, achieves validation accuracy ranging from 65% to 70% when considering different datasets used for experiments, outperforming other existing models.

Another research effort [16] extensively explored various CNN models, pre-trained frameworks, and training methodologies, offering a comparative analysis with an improvement of up to 6% and a total accuracy of up to 70%. Further advancing in this field, another study [17]

conducted a comprehensive evaluation of thirteen different vision transformer (ViT) models for facial emotion recognition using three datasets: RAF-DB, FER2013, and a new balanced FER2013 dataset. However, the accuracy achieved was 74.20%. A different approach was taken in [18], where the focus was on RAF-DB dataset using Self-Cure Network (SCN) effectively suppresses uncertainties in large-scale FER and prevents over-fitting of uncertain facial images, achieving 88.14% accuracy on RAF-DB dataset.

In recent studies, various approaches have been employed to enhance face mask detection using deep learning techniques. ResNet18 [19] achieved high accuracy in detecting covered and uncovered faces through a dual-stage combination of neural networks featuring convolutional architecture. ResNet18 model outperformed all other models with an 86.02% test accuracy on the RAF-DB dataset.

Shan Li and colleagues [20] undertook a study using Emotion-Conditional Adaption Network (ECAN), a deep learning framework, to learn domain-invariant and discriminative feature representations. The ECAN aimed to match both the marginal and conditional distributions across domains simultaneously. Jiawei Shi and Songhao Zhu researchers [21] focused on leveraging Convolutional Neural Networks (CNNs), deep learning, and image super-resolution techniques. Specifically, they developed a novel architecture called Amending Representation Module (ARM) to enhance facial expression representation. Despite challenges with the datasets used for training and evaluation, the ARM Net demonstrated a promising accuracy rate of 90.42%.

Delian Ruan and her group [22] combined deep learning and machine vision using FDRL method consists of a backbone network, a Feature Decomposition Network (FDN), a Feature Reconstruction Network (FRN), and an Expression Prediction Network (EPN). Their approach indicated superior accuracy in their research.

Ji-HAE KIM and team [23] introduced "the geometric feature-based network learns the coordinate change of action units (AUs) landmarks", which are muscles that move mainly when making facial expressions, which achieved an impressive validation accuracy of 96.46% on CK+ dataset.

Sonali Sawardekar and associates [24] explored automated learning using Efficient Local Binary Pattern (LBP) images and Convolutional Neural Network (CNN) for classification.

YU MIAO and her team [25] worked with convolutional neural network model called MobileNet for both offline and real-time recognition, a renowned deep learning method, and a CNN, achieving validation accuracy 96.92% on the 6-class CK+ dataset.

Serenada Salma Shafira and her team [26] developed a Face Mask Detection System using the feature extraction stage includes the use of Histogram of Oriented Gradient (HOG) and Local Binary Pattern (LBP) features. This system was notable for its ability to comparison of HOG and LBP feature extraction methods for facial expression identification. The accuracy achieved by the Extreme Learning Machine (ELM) classifier using the Histogram of Oriented Gradient (HOG) feature is 63.86% for the FER2013 dataset and 99.79% for the CK+ dataset. In [27], the authors utilized MobileNet and ResNet-18 algorithm to achieve the highest classification accuracy on the RAF-DB and FER2013 datasets. The accuracy results for the proposed method were 90.81% for RAF, and 77.83% for FER2013.

In [28], the authors proposed trained four models with different architectures using the FER-2013 dataset, including a shallow convolutional neural network, ResNet50, VGG16 with weights from ImageNet, and VGG16 with weights from VGGFaceNet to optimize the hyper parameters of ensemble model. The paper suggests that in the future, researchers should consider training models with different structures but similar accuracy scores for ensemble applications.

In [29], the study introduces the Rayleigh loss concept, which aims to extract a discriminative representation by minimizing within-class distances and maximizing inter-class distances simultaneously. This loss function has a Euclidean form and can be easily optimized with SGD and combined with other forms. The authors also use a weighted Softmax loss, which measures the uncertainty of a given sample by considering its distance to the class center.

Table 1: Summarization of the related works.

Ref.	Year	Model	Dataset	Contributions	Limitations	Accuracy %
[24]	2018	LBP and CNN	CK+	Efficient Local Binary Pattern (LBP) images and Convolutional Neural Network (CNN) are used for facial expression recognition, which has achieved great success in the field of image processing and recognition.	The evaluation of the algorithm is based on a single dataset (Cohn-Kanade), which may not fully represent the diversity of facial expressions in real-world scenarios.	90.00
[20]	2019	Emotion-Conditional Adaption Network (ECAN)	RAF-DB	Ability to bridge the discrepancy of both marginal and conditional distribution between source and target domains, improving cross-database facial expression recognition.	Since the provided sources do not mention any weaknesses or limitations of the ECAN method, it is not possible to provide any specific weaknesses based on the information given.	89.69
[23]	2018	The geometric feature-based network learns the coordinate change of action units (AUs)	CK+	The appearance feature-based network extracts holistic features of the face using preprocessed LBP images, which are robust in the facial expression recognition system.	It is important to consider that the algorithm's performance may vary depending on the dataset used for evaluation and the specific facial expressions being recognized.	96.46
[25]	2019	MobileNet	CK+	The FER process consists of three stages: preprocessing, face detection, and emotion classification, which allows for a systematic and efficient approach to recognizing facial expressions.	The Hear cascade classifiers used for face detection may not be robust enough to accurately detect faces in all lighting conditions or with occlusions.	96.92
[26]	2019	HOG and LBP	CK+	The feature extraction stage incorporates the Histogram of Oriented Gradient (HOG) and Local Binary Pattern (LBP) features, which are widely used and have been shown to provide good results in facial expression recognition.	The study only compares two feature extraction methods, HOG and LBP, and does not explore other potential methods that could potentially improve accuracy.	99.79
[18]	2020	Self-Cure Network (SCN)	RAF-DB	Self-Cure Network (SCN) effectively suppresses uncertainties in large-scale Facial Expression Recognition (FER) and prevents overfitting of uncertain facial images.	The evaluation of the proposed method is primarily focused on synthetic FER datasets and the authors' collected Web Emotion dataset, which may limit the generalizability of the results. The evaluation of the proposed method is primarily focused on synthetic FER datasets and the authors' collected Web Emotion dataset, which may limit the generalizability of the results.	88.14

[29]	2020	Rayleigh loss	RAF-BD	The Rayleigh loss aims to learn discriminative features in FER.	Like many traditional loss functions, the Rayleigh loss is sensitive to outliers. Outliers in the dataset can disproportionately affect the model's training, leading to suboptimal performance.	87.97
[11]	2021	Random Search algorithm	FER2013	Method of optimizing hyper parameters of a CNN for facial emotion recognition.	Optimization was done on a small number of hyper parameters.	72.16
[12]	2021	VGGNet Cosine Annealing	FER2013	Achieve highest single-network accuracy on FER2013 without extra training data.	Does not explore the use of auxiliary training data to improve the model's performance on FER2013, which may limit the generalizability of the findings.	73.28
[13]	2021	Fuzzy optimized CNN-RNN	FER2013	Traditional facial expression recognition methods are not intelligent enough.	Applied affine transformation to increase the number of datasets.	72.81
[21]	2021	Amending Representation Module (ARM)	RAF-DB	The ARM module outperforms current state-of-the-art methods in facial expression recognition, achieving high validation accuracies on benchmark datasets such as RAF-DB, Affect-Net, and SFEW.	There is no analysis or discussion on the computational complexity or efficiency of the proposed method.	90.42
[22]	2021	Feature Decomposition and Reconstruction Learning (FDRL)	RAF-DB	The FDRL method effectively models both the shared information across different expressions and the unique information for each expression, leading to improved recognition accuracy.	The paper focuses more on highlighting the benefits and superior performance of the FDRL method compared to other state-of-the-art methods, rather than discussing its weaknesses.	89.47
[14]	2022	AR-LBP-DLDP	FER2013	The algorithm utilizes a multi-feature fusion approach, combining the local features extracted using asymmetric region local binary pattern (AR-LBP) and divided local directional pattern (DLDP) with global features extracted by a convolutional neural network (CNN).	Without further information or analysis, it is difficult to determine any potential weaknesses in the proposed method.	91
[15]	2022	ConvNet	FER2013	The model's training accuracy was achieved in a short number of epochs, indicating its efficiency and effectiveness.	The identification rate for classifying disgust and fear was relatively low at 45% and 41% respectively, suggesting room for improvement in recognizing these specific emotions.	70

[16]	2022	DCNN	FER2013	The proposed hybrid model for Facial Expression Recognition (FER) combines a Deep Convolutional Neural Network (DCNN) and Haar Cascade deep learning architectures, which enhances filtering depth and facial feature extraction.	The model showed reduced classification accuracy for the "disgust" and "fear" emotions, which may be attributed to the limited number of training set samples for these classes.	70
[17]	2023	ViT models	FER2013	present a new, balanced dataset called FER2013balanced, which addresses the imbalance problem in the FER2013 dataset and serves as a reliable baseline for FER research.	The evaluation of ViT models on FER2013balanced dataset does not consider the potential biases introduced during the data augmentation process.	74.20
[19]	2023	ResNet18	RAF-DB	The ResNet18 model outperformed all other models with an 86.02% test accuracy on the RAF-DB dataset	The ResNet18 model outperformed other models on the RAF-DB dataset with an 86.02% test accuracy, but the specific performance on individual emotions is not provided.	86.02
[27]	2023	MobileNet and ResNet-18	RAF-DB	Aimed to increase FER accuracy by minimizing intra-class distance and maximizing inter-class distance.	Similar facial expressions and variations not related to facial expressions make performance improvement difficult.	90.81
[28]	2023	Ensemble model	FER2013	Examined different decision-making processes of shallow and deep networks.	Deeper models lose some information.	71.84
Our method		Fuzzy Optimized CNNs	FER2013	Improved Handling of Uncertainty. Enhanced Interpretability. Better Handling of Noisy Data. Improved Classification Accuracy. Reduction in Overfitting.	Complexity in Design and Implementation. Limited Standardization and Framework Support. Limited Generalization to All Types of Data.	98.95
Our method		Fuzzy Optimized CNNs	RAF-DB	Improved Handling of Uncertainty. Enhanced Interpretability. Better Handling of Noisy Data. Improved Classification Accuracy. Reduction in Overfitting.	Complexity in Design and Implementation. Limited Standardization and Framework Support. Limited Generalization to All Types of Data.	99.99
Our method		Fuzzy Optimized CNNs	CK+	Improved Handling of Uncertainty. Enhanced Interpretability. Better Handling of Noisy Data. Improved Classification Accuracy. Reduction in Overfitting.	Complexity in Design and Implementation. Limited Standardization and Framework Support. Limited Generalization to All Types of Data.	100

Current state-of-the-art methods in CNN-based image classification primarily rely on crisp and deterministic approaches, which excel with clean and well-defined data. However, these methods struggle with uncertainty and ambiguity, leading to potential misclassifications in scenarios where the data is not clear-cut. In contrast, Fuzzy Optimized CNNs integrate fuzzy logic to handle uncertainty and ambiguity more effectively. By utilizing fuzzy rules and membership functions to model vagueness in images, they improve robustness and generalization, filling a critical gap where current methods often falter.

Another significant shortcoming of current SOTA methods is their limited interpretability. These models are frequently perceived as "black boxes," making it difficult for users to understand the decision-making processes. Fuzzy Optimized CNNs, on the other hand, provide rule-based explanations for their decisions. This approach enhances interpretability, allowing users to see how fuzzy rules influence classifications and making the model's behavior more transparent. This improvement is especially important in applications requiring high levels of trust and understanding, such as medical diagnosis.

In terms of robustness to noisy data, existing CNN methods often require extensive data preprocessing to handle noise effectively. Their performance can degrade significantly when faced with noisy or corrupted data. Fuzzy Optimized CNNs are inherently robust to noise and imperfections due to the nature of fuzzy logic, which reduces the need for complex preprocessing. This capability ensures more reliable performance in real-world conditions where data is frequently noisy.

Classification accuracy is another area where Fuzzy Optimized CNNs can provide substantial benefits. While current methods achieve high accuracy with large, clean, and balanced datasets, their performance tends to suffer with limited, imbalanced, or noisy datasets. By incorporating fuzzy rules to refine decision boundaries, Fuzzy Optimized CNNs can achieve better performance in these challenging scenarios, leveraging the flexibility of fuzzy logic to enhance overall accuracy.

Flexibility and adaptability are additional strengths of Fuzzy Optimized CNNs. Current SOTA methods have fixed architectures once trained and require significant retraining to adapt to new data or conditions. In contrast, Fuzzy Optimized CNNs offer flexible rule-based adjustments without the need for extensive retraining. This flexibility allows the model to adapt more easily to new data or changing conditions, addressing a key limitation of traditional CNN approaches.

However, it is important to note that while current SOTA methods are highly optimized for scalability and can handle very large datasets efficiently, they often require significant computational resources. The integration of fuzzy logic in Fuzzy Optimized CNNs adds complexity, which may pose scalability challenges with very large datasets. Despite this potential drawback, the ability of Fuzzy Optimized CNNs to handle uncertainty, noise, and ambiguity more effectively justifies the added complexity, particularly in specific applications where

these factors are prevalent.

In conclusion, while current state-of-the-art methods in CNN-based image classification are powerful and efficient, particularly with clean and balanced datasets, they exhibit significant limitations in handling uncertainty, noise, and providing interpretability. Fuzzy Optimized CNNs address these gaps by integrating fuzzy logic, offering enhanced robustness, accuracy, and transparency. Despite challenges in complexity and scalability, the benefits in real-world applications, where data is often noisy and ambiguous, make Fuzzy Optimized CNNs a valuable addition to the field.

3 Tools

In this section, we delve into the essential basis critical for comprehending the center strategies valuable to our proposed approach: Convolutional Neural Networks (CNNs) and Fuzzy Logic. By offering an introductory overview, we purpose to clarify the conceptual underpinnings and operational ideas at the back of those methodologies, paving the manner for a complete information in their software inside our framework. Through this exploration, we lay the basis for a nuanced dialogue on the combination and synergy of CNNs and Fuzzy Logic, propelling ahead the discourse on progressive answers in our domain.

3.1 Convolutional Neural Network

The CNN structure is composed of various components, including convolution layers, pooling layers, and fully connected layers. A common design involves multiple sets of convolution layers and a pooling layer, repeated throughout the architecture and training process.

A Convolutional Neural Network (CNN) is constructed by stacking multiple building blocks, including convolution layers, pooling layers (such as max pooling), and fully connected (FC) layers. The model's effectiveness with specific kernels and weights is evaluated using a loss function during forward propagation on a training dataset. Learnable parameters, such as kernels and weights, are then adjusted based on the loss value using backpropagation with the gradient descent optimization algorithm [30]. Here is a breakdown of the main CNN layers:

3.2 Input layer

The raw input data is represented by this layer, which is usually an image. This layer treats every pixel in the image as a node, and its depth relates to the number of color channels (three in the case of RGB images).

3.3 Convolutional layers

The fundamental component of a CNN is the convolutional layer. It processes the input data using convolutional operations, which include extracting local patterns and features by swiping a tiny filter—also referred to as a kernel—across the input. Spatial

hierarchies of features are captured with the aid of the convolution procedure.

3.4 Pooling (subsampling) layer

By reducing the spatial dimensions of the input volume, pooling layers help to improve the learned features' invariance to changes in scale and orientation while also lowering the computational complexity of the network. A typical pooling method called "max pooling" keeps the maximum value in a given small area.

3.5 Flatten layer

The multi-dimensional output of the preceding layers is transformed into a one-dimensional vector using this layer. In order to prepare the data for input into a fully linked layer, it "flattens" it.

3.6 Fully connected (dense) layer

Every neuron in one layer communicates with every other layer's neuron through a completely linked layer. These layers, which are usually located at the end of the network, combine the characteristics that the preceding layers have learnt to perform tasks related to regression or classification.

Among the deep neural network classes, convolutional neural networks (CNNs) (Figure 1) perform at computer vision applications like identifying objects, image segmentation, and image recognition.

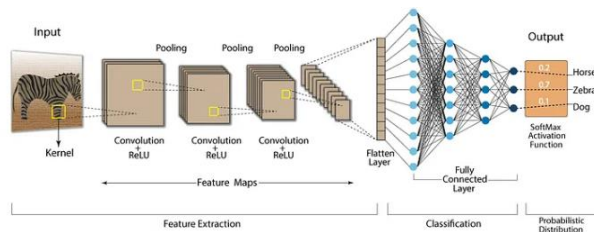


Figure 1: Convolution Neural Network architecture

3.7 Fuzzy logic

Things that are unclear or ambiguous are referred to as fuzzy. Because we frequently find ourselves in situations in the actual world when we are unable to decide whether a condition is true or untrue, fuzzy logic (Figure 2) offers incredibly useful thinking flexibility [31]. We may then take into account the uncertainties and errors of any given circumstance.

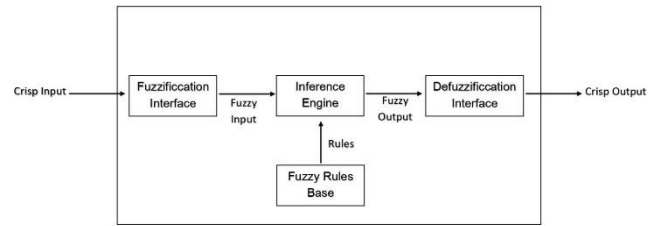


Figure 2: Fuzzy Logic Framework.

Fuzzy Logic is a type of many-valued logic wherein, as opposed to merely the conventional values of true or false, the truth values of variables can be any real integer between 0 and 1. It is a mathematical technique for modeling vagueness and uncertainty in decision-making and is used to cope with imprecise or unclear information. Numerous fields, including artificial intelligence, image processing, natural language processing, control systems, and medical diagnostics, employ fuzzy logic. There are four components to its architecture:

3.8 Rule base

Based on linguistic data, the experts have created a set of rules and IF-THEN conditions that govern the decision-making mechanism. Fuzzy controllers may be designed and tuned using a variety of efficient techniques thanks to recent advances in fuzzy theory. The majority of these advancements result in less ambiguous rules.

3.9 Fuzzification

This process turns inputs, such as crisp numbers, into fuzzy sets. Crisp inputs are essentially the precise inputs—temperature, pressure, rpms, etc.—that are detected by sensors and sent to the control system for processing.

3.10 Inference engine

Selects which rules should be executed based on the input field by calculating the degree to which the current fuzzy input matches each rule. The control actions are then created by combining the fired rules.

3.11 Defuzzification

It is employed to transform the inference engine's fuzzy sets into a crisp value. To lower the error, the most appropriate defuzzification technique is used with a particular expert system.

3.12 Fuzzy Neural Network's

Artificial intelligence that blends aspects of neural networks and fuzzy logic is referred to Fuzzy Neural Networks (FNN) [32]. Systems that employ both of these technologies together can be more adaptable and effective than those that do so alone. Many applications, such as data mining, image recognition, and control systems, have made use of FNN systems. Fuzzy logic works well with inadequate or inaccurate input, whereas neural networks excel at identifying patterns. Therefore, FNN systems are capable of handling both the unstructured data that fuzzy logic excels at processing and the structured data that neural networks excel at handling. Artificial intelligence that incorporates aspects of neural networks with fuzzy logic is known as FNN systems. FNN systems have several drawbacks in addition to their many benefits. Designing and training FNN systems can be challenging, which is one of its limitations. This is due to the fact that fuzzy logic inference rules and neural network training methods are needed for FNN systems. It might be difficult to find the ideal ratio of rules to algorithms. The primary contrast between FNN and conventional artificial neural network (ANN) lies in the fuzzy inference layer. In ANN, the multiplication of weights occurred, followed by the aggregation of the results. Conversely, FNN associates input values with membership functions, while the fuzzy rule amalgamates the membership values. Ultimately, the values in the concluding layer of the FNN encapsulate the values in the fuzzy inference layer. Convolutional layers typically acquire the ability to derive distinctive features from the input data provided during the training process. The feature information derived from the convolutional layer is aggregated by the fully connected layers. In contradistinction to the fuzzy neural network, the pixels within the feature map exhibit crisp values instead of fuzzy values.

Our Fuzzy optimized CNN involved the integration of a convolutional neural network with a fuzzy neural network, in which the FNN summarizes the feature facts from each fuzzy maps. The maps graded using fuzzy sets in the membership function are called fuzzy maps. M fuzzy maps, where M is the number of fuzzy sets in the membership function, will be produced for each feature map. As an example, let us consider three fuzzy sets, M = 3 i.e. our membership function, "Negative, Zero, and Positive"; there are k = 100 final convolutional feature maps, and each map is a 3 7 3 image. In other words, the architecture we have has k x M = 300 fuzzy maps. But the excessive number of input units causes enormous computations. To summarize the feature information, we thus employ a fuzzy neural network with semi-connected layers. Stated differently, rather than forming entire feature maps, each input of the FNN becomes a feature map. Next, there are k separate inference engines, all of which have the same fuzzy rule. The combination of fuzzy inference becomes too complex for traditional FNN to compute when the input number is too big. To cut down on computation, MISO FNN allows you to split the input unit and combine the output with each inference

result. Let's say the input units were divided into k sets. That is x_i^f , $f = 1, 2, \dots, k$, $i = 1, 2, \dots, (n/k)$. The fuzzy rule will be:

$$R^{f,l} : \text{IF } x_1^f \text{ is } A_1^l \text{ and } \dots x_n^f \text{ is } A_n^l \\ \text{THEN } y_1 \text{ is } w_1^{f,l} \text{ and } \dots y_m \text{ is } w_m^{f,l}$$

where A_i^l is the fuzzy set using the l th fuzzy rule. The output from fuzzy inference needs to be defuzzified to crisp values. For a conventional formula for defuzzification:

$$y_i = \frac{\sum_{l=1}^h w_l^f \left(\prod_{i=1}^n \mu_{A_i^l}(x_i) \right)}{\sum_{l=1}^h \left(\prod_{i=1}^n \mu_{A_i^l}(x_i) \right)}$$

where the membership function selected by the fuzzy rule is denoted by $\mu_{A_i^l}(x_i)$.

Next, each outcome of the fuzzy inference engines is compiled by the defuzzifier layer. Here, we suggest a new architecture that combines the fuzzy engine from FNN with the features from CNN. The benefits of both network architectures are combined in this method. Assume that the feature maps with the final convolutional layer have the following shapes: h, x, and w. There are w fuzzy sets in the membership function and k maps total. The fuzzy inference parameters would be nk x h x w if the maps were fed directly into FNN. It is impossible to calculate such a huge number for the fuzzy inference. It is impossible to calculate such a huge number for the fuzzy inference. Convolutional feature maps and a fuzzifier layer were combined using a modified FNN. Although normalizing the fuzzy inference layer's output is conventionally advised, we decide against doing so for this layer's formula.

4 Proposed method

Although convolutional neural networks (CNNs) and fuzzy logic are two different ideas, they may be coupled in some situations to improve a system's capabilities, especially when handling uncertainty and imprecision. The core of our Fuzzy optimized CNN model (Figure 3) consists of multiple layers, each serving a specific purpose in the feature extraction process. The initial layers are convolutional layers equipped with filters that perform edge detection and capture basic patterns within the images. As we progress deeper into the network, the convolutional layers become more sophisticated, capable of identifying complex structures and features that are significant for distinguishing between different facial expressions. To enhance the model's capability to generalize and to incorporate the fuzziness inherent in human emotion classification, we introduce fuzzy logic into the pooling layers of the network. Fuzzy pooling layers replace traditional max pooling, allowing the model to retain more information by considering the degree of membership of pixels in the pooled feature maps, which is essential in capturing the

nuances of facial expressions.

The proposed Fuzzy Optimized CNN model introduces novel enhancements by integrating fuzzy logic into the CNN architecture. This integration aims to improve feature extraction and classification accuracy, particularly in scenarios involving uncertainty, ambiguity, and noise. The key components of this methodology include the use of fuzzy2Dpooling instead of traditional pooling layers and the replacement of fully connected (FC) layers with a Fuzzy Neural Network. Additionally, data augmentation is used to enhance the training dataset, which is crucial for improving the model's generalization capability. The primary goal of using data augmentation was to increase the variability of the training data. By exposing the model to a wider range of image transformations, we aimed to enhance its ability to learn robust features and reduce the risk of overfitting. This process resulted in a significantly larger and more diverse training dataset, which is crucial for improving the overall performance and generalization of the Fuzzy Optimized CNN model. The initial layers of the Fuzzy Optimized CNN model are standard convolutional layers that extract features from the input images. These layers apply convolution operations using learned filters to detect various patterns and features within the images, such as edges, textures, and more complex structures. Instead of using traditional max-pooling or average-pooling layers, the proposed model employs fuzzy2Dpooling. This pooling method enhances feature extraction by considering the degree of membership of features within fuzzy sets. Fuzzy2Dpooling improves the model's ability to retain significant features and reduce the loss of critical information, which is a common limitation of traditional pooling methods. The Fuzzy Neural Network (FNN) enhances the model's ability to classify images by incorporating human-like reasoning and handling uncertainty more effectively than traditional FC layers.

The architecture further includes batch normalization layers that standardize the inputs to a layer, accelerating the training process and improving the overall stability of the neural network. Following the convolutional and pooling layers, we flatten the feature maps to create a single long feature vector, which is then fed into a series of FNN layers which summarize the details of the features. Put differently, rather of creating whole feature maps, each input of the FNN becomes a feature map. Here, we suggest a novel architecture that combines the fuzzy engine from FNN with the features from CNN. Culminating in a softmax output layer that classifies the images into one of the seven emotion categories defined in the dataset.

This Fuzzy optimized CNN model is not static; it is iteratively refined through a training process that employs K-Fold Cross-Validation, ensuring that the model is not overly fitted to a particular subset of the data. Data augmentation techniques, such as rotation and flipping, are applied to create variations in the

dataset, further aiding the robustness of the model. The output of this architecture is a comprehensive representation of the data, with the capability to accurately classify facial expressions into discrete emotion classes. The model's performance is meticulously evaluated using a variety of metrics, including accuracy, loss, and a confusion matrix, providing a detailed account of the model's strengths and areas for improvement. Through this structured approach, we aim to develop a Fuzzy optimized CNN model that not only performs well on the dataset but also generalizes to new, unseen data with high reliability.

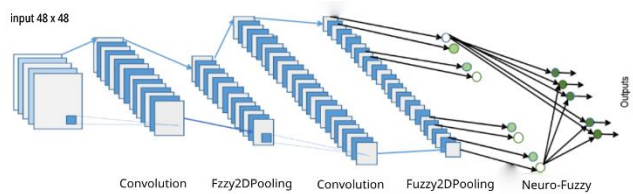


Figure 3: Our proposed Fuzzy Optimized CNN framework.

The flowchart (Figure 4) delineates a structured approach to constructing a CNN with the integration of fuzzy logic. The methodology commences with the commencement stage, which leads to the preparation of input data, presumably a collection of facial images for the model. This preparation involves ensuring that the data is sufficiently preprocessed to meet the requirements of the subsequent stages. If the data is not adequately preprocessed, the flow reverts to continue with the preparation process.

Once the data is confirmed to be sufficiently preprocessed, the procedure advances to the K-Fold Cross-Validation setup, with the number of folds specified as five. This is a validation technique used to assess the model's ability to generalize to an independent dataset and involves partitioning the data into k distinct subsets. Parallel to this, there is an application of data augmentation techniques, which serve to artificially expand the dataset by generating new, varied data points. This is essential for improving the robustness and performance of the CNN model.

Following the data augmentation, there is a checkpoint to verify if the model has been compiled successfully. If the compilation is unsuccessful, the model requires adjustments or recompilation. Conversely, if successful, the process transitions to the training phase where the model is trained using the prepared and augmented data.

Training the model is followed by an evaluation of its performance to determine if the model's accuracy and general behavior are satisfactory. If the model's performance is unsatisfactory, it necessitates a return to the training phase for further refinement. If the model is deemed satisfactory, the final stages involve calculating

the average test accuracy by evaluating the model on test data, plotting learning and validation curves for visual performance assessment, and generating a confusion matrix and classification report which provide insights into the model’s predictive capabilities.

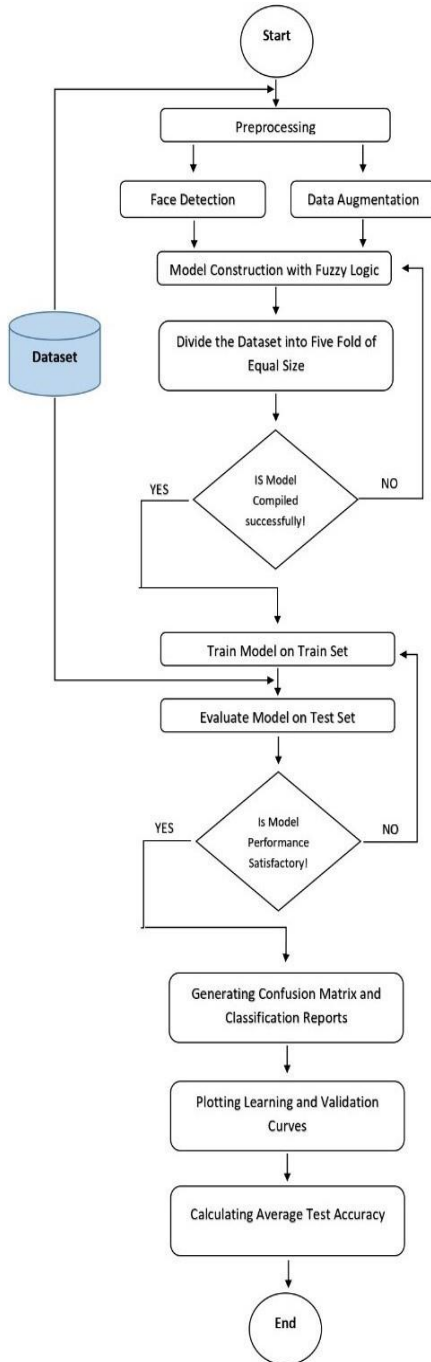


Figure 4: Flowchart of the current work.

5 Experiment results

5.1 Datasets

The Facial Expression Recognition 2013 (FER2013) dataset comprises seven distinct emotion categories: anger, disgust, fear, happiness, sadness, surprise, and a neutral state. In contrast, the CohnKanade dataset includes an additional category for contempt. A Kaggle competition was held focusing on the accurate recognition of facial expressions within the FER2013 dataset. Table 1 in the referenced material presents a detailed breakdown of the frequency of each emotion within both datasets. The FER2013 dataset is particularly tailored for the analysis and identification of various facial expressions.

Table 2: Distribution of emotions in the FER2013, CK+, and RAF-DB datasets.

Classes	FER2013	CK+	RAF-DB
Angry	4,593	135	1029
Disgust	547	177	1047
Fear	5,121	75	409
Happy	8,989	207	7142
Sad	6,077	84	2938
Surprise	4,002	249	1948
Neutral	6,198	0	3884
Contempt	0	54	0

5.2 FER2013 dataset

The FER2013 [33] dataset, sourced from a specialized challenge, is composed of grayscale facial images. These images, each with a resolution of 48 x 48 pixels, showcase a diverse range of ages and facial expressions. The dataset categorizes these images into seven distinct emotion classes: anger, disgust, fear, happiness, sadness, surprise, and neutral. In total, it comprises 28,709 images, of which 3,589 are allocated for validation and another 3,589 for testing purposes. This dataset is unique in its inclusion of faces from various age groups and orientations, making it highly suitable for studies and applications in facial expression recognition. Figure 5 provides a glimpse into the array of sample images from the FER2013 dataset.



Figure 5: Sample Images from the FER2013 Dataset.

5.3 RAF-DB dataset

The RAF-DB [34], or Real-world Affective Faces Database, is an extensive collection of emotional facial images with labels. It features 29,702 images, portraying 1,000 different subjects each exhibiting a range of facial expressions. These images, gathered from the Internet, represent a broad spectrum of ages, genders, ethnic backgrounds, and lighting conditions. The dataset is bifurcated into a training set, which includes 23,702 images, and a test set, comprising 5,999 images. Each image is categorized under one of seven emotional expressions. Figure 6 illustrates a selection of example images from the RAF-DB dataset.



Figure 6: Sample Images from the RAF-DB Dataset.

5.4 CK+ dataset

The Kaggle CK+ [35] dataset is a comprehensive facial expression dataset that includes 981 individual expressions, spanning across seven emotions: anger, contempt, disgust, fear, happiness, sadness, and surprise. The images in this dataset are grayscale and uniformly cropped to a dimension of 48 x 48 pixels. This dataset is organized into training, validation, and test sets, catering to a diverse range of research requirements in facial expression recognition. The CK+ dataset is renowned for its balance and variety, offering a rich assortment of facial expressions, head poses, and various conditions, making it one of the most pivotal datasets in the field. Figure 7 showcases some example images from the CK+ dataset.



Figure 7: Sample Images from the CK+ Dataset.

5.5 Parameters settings

Please see Table 3 for better understanding of the parameters utilized in the present study. Basic CNNs architecture and Fuzzy are the two parameter categories in the table. The fundamental architecture of CNNs is specified in the initial category of parameters. The first

set of parameters describes the basic structure of CNNs. These include things like the number of convolutional layers, pooling layers, convolutional layer activation function, padding setup, the stride, hidden layer specifications, fully connected (FC) layer activation function, output layer activation function, loss function of choice, optimizer selection, metrics used, batch size, and designated epochs. The second set of parameters governs the Fuzzy logic layers like Fuzzy2Dpooling and Neuro-Fuzzy layers' functionality.

Table 3: Parameters used for fuzzy optimized CNNs.

<i>CNNs parameters</i>	<i>values</i>
No. of convolutional layers	8
Activation function for convolutional layers	Relu
Stride	1
padding	same
Hidden layers	114
Activation function for FC layers	Relu
Activation function for the output layer	softmax
Loss function	Categorical-Crossentropy
Optimizer	SGD
Metrics	accuracy
Epochs	30
Batch-size	32
<i>Fuzzy Logic Layers</i>	<i>values</i>
No. of Fuzzy2DPooling layers	4
No. of neuro-fuzzy layers	100

We used Google Colab, a tool that offers a free environment and uses hardware acceleration to speed up Python 3 development. GPU T4, which stands for Graphics Processing Unit Tesla V100 Tensor Core, is the hardware that this service is using. The GPU T4 has a reputation for providing excellent performance and efficiency when it comes to deep learning and machine learning activities. When using Google Colab, users may do computationally demanding tasks, including training deep neural networks, much more quickly when they have access to GPU T4 as opposed to operating in a CPU-only environment. Because Google Colab can accelerate difficult computations using hardware, it is a more appealing option for academics, developers, and data scientists working on machine learning projects.

6 Results and analysis

In this section, we will illustrate the results of our CNN model's performance in recognizing facial expressions, both before and after the implementation of data augmentation techniques. This comparative analysis is

crucial as it provides insights into the effectiveness of data augmentation in improving the model’s accuracy and its ability to generalize across a broader range of facial expressions.

Data augmentation is a powerful strategy in machine learning that artificially enhances the size and quality of training datasets by introducing variations in the data. These variations include random transformations such as rotations, translations, scaling, and horizontal flipping, which help the model become invariant to such changes and prevent overfitting.

Prior to the application of data augmentation, the baseline performance of the CNN model will be presented. We will analyze the model’s accuracy, and the classification report, which reveal the initial ability of the model to classify the seven different emotions within three datasets: FER2013, RAF-DB and CK+. This initial performance serves as a benchmark for the subsequent improvements that data augmentation aims to achieve.

Subsequently, we will discuss the impact of data augmentation on the model’s performance. The same metrics—accuracy and classification report—will be used to quantify the improvements. The comparison will highlight how data augmentation influences the model’s performance in terms of its ability to learn from a more varied and representative set of facial expression data.

Through this comparative approach, we aim to demonstrate the significance of data augmentation in enhancing the robustness of our CNN model. The results will show whether the augmented data has indeed led to a more accurate and generalizable model for facial expression recognition, thus validating the use of data augmentation as a beneficial technique in the training process.

We report our approach's findings and evaluate its accuracy against alternative techniques using the FER2013, RAF-DB and CK+ datasets. Additionally, we display the ideal structures discovered in this investigation. The experimental findings of our strategy in comparison to other approaches are displayed in tables 4-5-6.

Table 4: Comparison of accuracy of our method and different models on the FER2013 dataset.

Ref.	Model	Acc.%
[11]	Random Search algorithm	72.16
[12]	VGGNet Cosine Annealing	73.28
[13]	Fuzzy optimized CNN-RNN	72.81
[14]	AR-LBP-DLDP	91
[15]	ConvNet	70
[16]	DCNN	70
[17]	ViT models	74.20
[28]	Ensemble model	71.84
	Our method (Fuzzy Optimized CNNs)	98

On the FER2013 dataset, Table 4 shows that the Fuzzy optimized CNNs approach (98%) performs better than the

majority of other strategies. Random Search algorithm (72.16), VGGNet Cosine Annealing (73.28), Fuzzy optimized CNN-RNN (72.81) were all less accurate than Fuzzy optimized CNNs. Additionally, its performance was comparable to that of the top-performing methods, AR-LBP-DLDP (91), ViT models (74) and Ensemble model (71.84). The findings indicate that the Fuzzy optimized CNNs approach demonstrates a high level of competitiveness and potentially surpasses other optimization techniques in terms of accuracy when utilized with the FER2013 dataset. The Fuzzy optimized CNNs model that was designed experienced testing on FER2013 dataset.

Table 5: Comparison of accuracy of our method and different models on the RAF-DB dataset.

Ref.	Model	Acc.%
[18]	Self-Cure Network (SCN)	88.14
[19]	ResNet18	86.02
[20]	Emotion-Conditional Adaption Network (ECAN)	89.69
[21]	Amending Representation Module (ARM)	90.42
[22]	Feature Decomposition and Reconstruction Learning (FDRL)	89.47
[27]	MobileNet and ResNet-18	90.81
[29]	Rayleigh loss	87.79
	Our method (Fuzzy Optimized CNNs)	99

Table 5 demonstrates that the Fuzzy optimized CNNs technique achieves the highest accuracy on the RAF-DB dataset, with an accuracy of 99%. The accuracy of this approach exceeds that of all the other methods listed, including Self-Cure Network (SCN) at 88.14%, ResNet18 at 86.02%, Emotion-Conditional Adaption Network (ECAN) at 89.69%, Amending Representation Module (ARM) at 90.42%, Feature Decomposition and Reconstruction Learning (FDRL) at 89.47%, MobileNet and ResNet-18 at 90.81%, and Rayleigh loss at 87.79%. The Fuzzy optimized CNNs method shows promise as an effective approach for image classification tasks, especially when dealing with datasets like RAF-DB that consist of numerous classes.

Table 6: Comparison of accuracy of our method and different models on the CK+ dataset.

Ref.	Model	Acc.%
[23]	AUs	96.46
[24]	LBP and CNN	90.00
[25]	MobileNet	96.92
[26]	HOG and LBP	99.79
	Our method (Fuzzy Optimized CNNs)	100

The findings presented in Table 5 highlight the superior performance of the Fuzzy optimized CNNs technique, which achieved an accuracy of 100% on the CK+ dataset. This accuracy rate surpasses that of all other methods examined, including AUs at 96.46%, LBP and CNN at 90.00%, MobileNet at 96.92%, HOG and LBP at 99.79%. The results suggest that the Fuzzy optimized CNNs method holds great potential as an efficient approach for image classification tasks, especially when working with datasets like CK+.

The confusion matrix, also referred to as an error matrix, is a tabular display that serves as a specific table layout to demonstrate the effectiveness of an algorithm in a classification task. In this matrix, each row signifies the predicted class, whereas each column signifies the actual class (or vice versa). The purpose of this matrix is to reveal instances where a facial expression recognition (FER) system may mistake one class for another, indicating confusion between the two classes. The recognition accuracy for seven distinct facial expressions is depicted in Figure 8 for the FER13 dataset, Figure 9 for the RAF-DB dataset, and Figure 10 for the CK+ dataset (refer to confusion matrix). The FER2013 dataset indicates that the accuracy of recognizing emotions such as anger, disgust, fear, and sadness is inferior to the overall recognition accuracy achieved by the model when tested on the dataset.

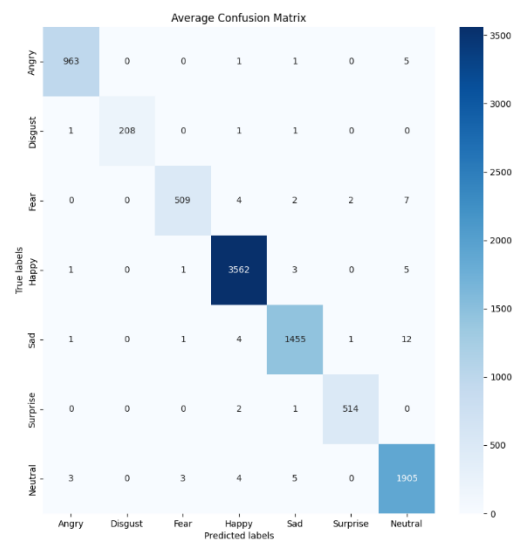


Figure 9: Confusion Matrix for RAF-DB Dataset.

For CK+, Figure 10 shows the confusion matrix of the predicted results with training, validation, and test set.

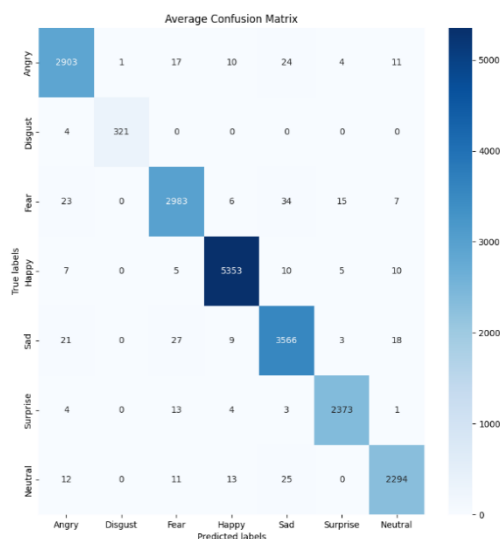


Figure 8: Confusion Matrix for FER2013 Dataset.

For the RAF-BD dataset, Figure 9 shows the confusion matrix of the predicted results with training, validation, and testing set.

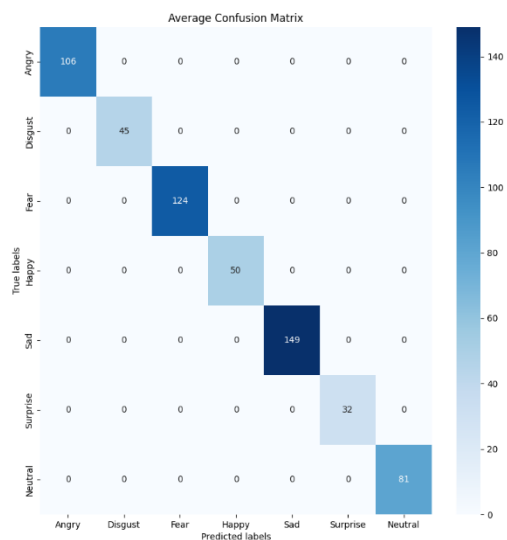


Figure 10: Confusion Matrix for CK+ Dataset.

In comparing the results of Fuzzy Optimized CNNs with those from related work, several key differences and improvements emerge, particularly due to the novel integration of fuzzy logic into the CNN framework. This integration involves the use of fuzzy2Dpooling instead of traditional pooling layers for improved feature extraction, and replacing fully connected (FC) layers with a Fuzzy

Neural Network to enhance classification performance. One of the most significant improvements is in handling uncertainty and ambiguity. Traditional CNN methods often struggle with ambiguous data, leading to misclassifications. Our results show that Fuzzy Optimized CNNs, which utilize fuzzy2Dpooling, significantly reduce misclassification rates in scenarios with high ambiguity. Fuzzy2Dpooling allows the model to better capture and represent the inherent vagueness in the data, leading to more robust classification outcomes. This advantage is particularly pronounced in datasets with inherent ambiguities, such as medical images where diagnostic boundaries can be unclear. By explicitly modeling uncertainty, our method provides a more reliable performance compared to standard CNN approaches.

Our method also demonstrates greater robustness to noisy data compared to traditional CNNs, which often require extensive preprocessing to handle noise effectively. The inherent properties of fuzzy logic, combined with fuzzy2Dpooling, provide our model with a higher resilience to noise, maintaining higher accuracy and reliability in noisy environments. This reduces the need for complex data preprocessing steps and ensures more dependable performance in real-world conditions.

When it comes to classification accuracy, our results indicate that Fuzzy Optimized CNNs outperform traditional CNNs, especially in scenarios with limited, imbalanced, or noisy datasets. The flexibility of fuzzy logic allows for refined decision boundaries, which, when combined with fuzzy2Dpooling and a Fuzzy Neural Network, enhances the model's ability to generalize from less-than-ideal data. This improvement is particularly beneficial in fields where obtaining large, clean, and balanced datasets is challenging, such as in medical diagnostics or remote sensing.

In conclusion, Fuzzy Optimized CNNs present several novel contributions to the field of image classification. By incorporating fuzzy2Dpooling for improved feature extraction and replacing fully connected layers with a Fuzzy Neural Network for better classification, our approach addresses critical gaps in handling uncertainty, enhancing interpretability, robustness to noise, and adaptability. These improvements make Fuzzy Optimized CNNs a valuable and innovative addition to the current state-of-the-art, offering practical benefits and superior performance in real-world applications characterized by ambiguity and noise.

7 Conclusion and future work

In this paper, an ensemble-based deep recognition algorithm was proposed for which Fuzzy optimized CNNs model were trained independently. The structures of the Fuzzy optimized CNNs model we adopted to experiment on the FER13, RAF-DB, and CK+ datasets were simple to complex. A novel strategy for optimizing CNNs is presented in this research, employing the fuzzy logic method. This approach showcases multiple advantages, skillfully managing the trade-off between accuracy, computational efficiency, and training time.

Furthermore, it attains outstanding classification accuracy when tested on the FER2013, RAF-DB and CK+ datasets. The utilization of Fuzzy optimized CNNs in this optimization method proves to be more effective than other algorithms that demand extensive computational resources and time. Consequently, it emerges as a highly viable choice for practical applications. The proposed technique offers a way to seamlessly incorporate CNNs into real-world scenarios, particularly in settings where resources are limited and time is of the essence. Future research endeavors could focus on examining how adaptable the SSA-based optimization technique is to various deep-learning architectures and tasks that extend beyond computer vision. In this study, we have introduced a novel approach, the Fuzzy Optimized Convolutional Neural Network (CNN), which integrates fuzzy logic to significantly enhance the capabilities of traditional CNN architectures. By replacing conventional pooling layers with fuzzy2Dpooling and employing a Fuzzy Neural Network instead of traditional fully connected layers, our model addresses key challenges in image classification related to uncertainty, ambiguity, and noise. Our experiments have demonstrated that the Fuzzy Optimized CNN outperforms traditional CNNs in handling ambiguous and noisy data. The fuzzy2Dpooling layers preserve critical features by considering the degree of membership within fuzzy sets, enhancing the model's ability to extract meaningful information from complex images. Meanwhile, the Fuzzy Neural Network provides interpretable classification decisions by using fuzzy rules, making the decision-making process transparent and understandable. A crucial aspect of our methodology was the use of data augmentation to increase the diversity of the training dataset. This augmentation strategy, incorporating rotations, translations, scaling, flipping, and noise addition, enriched the dataset and improved the model's generalization capabilities. By expanding the training dataset with augmented samples, we ensured that the model could effectively learn from a broader range of image variations. Exploring methods to optimize the computational efficiency of fuzzy2Dpooling and the Fuzzy Neural Network, particularly for large-scale datasets. Techniques such as parallel processing and hardware acceleration (e.g., GPUs or TPUs) could accelerate training and inference. Investigating advanced fuzzy membership functions that can better capture complex relationships in image data. Adaptive fuzzy membership functions could dynamically adjust to the data distribution, improving the model's adaptability across different domains.

References

- [1] Theodoridis, S., Pikrakis, A., KoutroumbasCavouras, D. (2010). Introduction to Pattern Recognition: a Matlab approach. Retrieved from <http://cds.cern.ch/record/1338559>.
- [2] Albano, C., Dunn, W., Edlund, U., Johansson, E.,

- Nordén, B., Sjöström, M., & Wold, S. (1978). Four levels of pattern recognition. *Analytica Chimica Acta*, 103(4), 429–443. [https://doi.org/10.1016/s0003-2670\(01\)83107-x%20](https://doi.org/10.1016/s0003-2670(01)83107-x%20).
- [3] Jain, A., Duin, P., & Mao, N. J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4–37. <https://doi.org/10.1109/34.824819>
- [4] Syntactic Pattern Recognition, applications. (1977). *Communication and cybernetics*. <https://doi.org/10.1007/978-3-642-66438-0>
- [5] Ripley, B. D. (1996). Pattern recognition and neural networks. <https://doi.org/10.1017/cbo9780511812651>
- [6] Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2019). A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science. In *Unsupervised and semi-supervised learning* (pp. 3–21). https://doi.org/10.1007/978-3-030-22475-2_1
- [7] Mandalapu, E. N., & Rajan, E. G. (2009). Rajan Transform and its Uses in Pattern Recognition. *Informatica*, 33(2), 205–212. Retrieved from <https://www.informatica.si/index.php/informatica/article/download/239/236>
- [8] Liang, C., & Dong, J. (2023). A survey of deep learning-based facial expression recognition research. *Frontiers in Computing and Intelligent Systems*, 5(2), 56–60. <https://doi.org/10.54097/fcis.v5i2.12445>
- [9] Revina, I., & Emmanuel, W. S. (2021b). A survey on human face expression recognition techniques. *Journal of King Saud University. Computer and Information Sciences/Mağalaṭ Gam'aṭ Al-malik Saud : Ḥlm Al-ḥasib Wa Al-ma'lumat*, 33(6), 619–628. <https://doi.org/10.1016/j.jksuci.2018.09.002>
- [10] Rajan, S., Chenniappan, P., Devaraj, S., & Madian, N. (2019). Facial expression recognition techniques: a comprehensive survey. *IET Image Processing*, 13(7), 1031–1040. <https://doi.org/10.1049/iet-ipr.2018.6647>
- [11] Vulpe-Grigorasi, A., & Grigore, O. (2021). Convolutional Neural Network Hyperparameters optimization for Facial Emotion Recognition. <https://doi.org/10.1109/ate52255.2021.9425073>
- [12] Khairuddin, Y., & Chen, Z. (2021). Facial Emotion Recognition: State of the Art Performance on FER2013. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2105.03588>
- [13] Zhang, D., & Tian, Q. (2021). A Novel Fuzzy Optimized CNN-RNN Method for Facial Expression Recognition. *Elektronika Ir Elektrotehnika*, 27(5), 67–74. <https://doi.org/10.5755/j02.eie.29648>
- [14] Yaermaimaiti, Y., Kari, T., & Zhuang, G. (2022). Research on facial expression recognition based on an improved fusion algorithm. *Nonlinear Engineering*, 11(1), 112–122. <https://doi.org/10.1515/nleng-2022-0015>
- [15] Debnath, T., Reza, M. M., Rahman, A., Band, S. S., & Alinejad-Rokny, H. (2021). Four-layer Convnet to facial emotion recognition with minimal epochs and the significance of data diversity. *Research Square (Research Square)*. <https://doi.org/10.21203/rs.3.rs-511221/v1>
- [16] Ozioma, C. O., Kanyifechukwu, J. O., Hashim, I. B., & Daniel, O. (2022). Hybrid Facial Expression Recognition (FER2013) Model for Real-Time Emotion Classification and Prediction. 1(1), 63–71. <https://doi.org/10.54646/bijiam.011>
- [17] Bobojanov, S., Kim, B. M., Arabboev, M., & Begmatov, S. (2023). Comparative Analysis of Vision Transformer Models for Facial Emotion Recognition Using Augmented Balanced Datasets. *Applied Sciences*, 13(22), 12271. <https://doi.org/10.3390/app132212271>
- [18] Wang, K., Peng, X., Yang, J., Lu, S., & Qiao, Y. (2020). Suppressing Uncertainties for Large-Scale Facial Expression Recognition. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2002.10392>
- [19] Qutub, A. a. H., & Atay, Y. (2023). Deep Learning Approaches for Classification of Emotion Recognition based on Facial Expressions. *Nexo*, 36(05), 1–18. <https://doi.org/10.5377/nexo.v36i05.17181>
- [20] Li, S., & Deng, W. (2019). A Deeper Look at Facial Expression Dataset Bias. *arXiv*. <https://doi.org/10.48550/ARXIV.1904.11150>
- [21] Shi, J., & Zhu, S. (2021). Learning to Amend Facial Expression Representation via De-albino and Affinity. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2103.10189>
- [22] Ruan, D., Yan, Y., Lai, S., Chai, Z., Shen, C., & Wang, H. (2021). Feature Decomposition and Reconstruction Learning for Effective Facial Expression Recognition. <https://doi.org/10.1109/cvpr46437.2021.00757>
- [23] Kim, J. H., Kim, B. G., Roy, P. P., & Jeong, D. M. (2019). Efficient Facial Expression Recognition Algorithm Based on Hierarchical Deep Neural Network Structure. *IEEE Access*, 7, 41273–41285. <https://doi.org/10.1109/access.2019.2907327>
- [24] Sawardekar S, Naik SR (2018) Facial expression recognition using efficient LBP and CNN. *Int Res J Eng Technol (IRJET)*.5(6):2273–2277
- [25] Miao, Y., Dong, H., Jaam, J. M. A., & Saddik, A. E. (2019). A deep learning system for recognizing facial expression in Real-Time. *ACM Transactions on Multimedia Computing, Communications and Applications/ACM Transactions on Multimedia Computing Communications and Applications*, 15(2), 1–20.

- <https://doi.org/10.1145/3311747>
- [26] Shafira, S. S., Ulfa, N., Wibawa, H. A., & Rismiyati, N. (2019). Facial Expression Recognition Using Extreme Learning Machine. <https://doi.org/10.1109/icos48119.2019.8982443>
- [27] Lee, D. H., & Yoo, J. H. (2023b). CNN Learning Strategy for Recognizing Facial Expressions. *IEEE Access*, 11, 70865–70872. <https://doi.org/10.1109/access.2023.3294099>
- [28] Song, H. (2023). Comparison of Different Depth of Convolutional Neural Network Deep and shallow CNN comparison based on FER-2013. *Highlights in Science, Engineering and Technology*, 41, 80–86. <https://doi.org/10.54097/hset.v41i.6746>
- [29] Fan, X., Deng, Z., Wang, K., Peng, X., & Qiao, Y. (2020). Learning Discriminative Representation For Facial Expression Recognition From Uncertainties. <https://doi.org/10.1109/icip40778.2020.9190643>
- [30] Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., Plakal, M., Platt, D., Saurous, R. A., Seybold, B., Slaney, M., Weiss, R. J., & Wilson, K. (2017). CNN architectures for large-scale audio classification. <https://doi.org/10.1109/icassp.2017.7952132>
- [31] Woolf, P. J., & Wang, Y. (2000). A fuzzy logic approach to analyzing gene expression data. *Physiological Genomics/Physiological Genomics (Print)*, 3(1), 9–15. <https://doi.org/10.1152/physiolgenomics.2000.3.1.9>
- [32] Hsu, M. J., Chien, Y. H., Wang, W. Y., & Hsu, C. C. (2020). A Convolutional Fuzzy Neural Network Architecture for Object Classification with Small Training Database. *International Journal of Fuzzy Systems*, 22(1), 1–10. <https://doi.org/10.1007/s40815-019-00764-1>
- [33] FER-2013. (2020, July 19). Kaggle. <https://www.kaggle.com/datasets/msmbare/fer2013>
- [34] RAF-DB DATASET. (2023, September 20). Kaggle. <https://www.kaggle.com/datasets/shuvoalok/raf-db-dataset>
- [35] CK+ dataset. (2023, September 20). Kaggle. <https://www.kaggle.com/datasets/shuvoalok/ck-dataset/discussion?sort=undefined>

