Wasit Journal for Pure Science

Journal Homepage: https://wjps.uowasit.edu.iq/index.php/wjps/index
e-ISSN: 2790-5241 p-ISSN: 2790-5233

Real-Time Lightweight Emotional Face Recognition Framework Based on CNN Algorithms

REHAB HASAN ABOOD[1]*, MOHAMMED AHMED TALAB[2], MUSTAFA TALAL SHEKER[3], IQRA KHAN[4], SURYANTI AWANG[5]

[1]Al-Falluja Secondary School for Girls, General Directorate of Anbar Education, Ministry of Education, Anbar, Iraq.

[2]Department of Medical Physics, College of Applied Sciences, Al-Fallujah University, IRAQ.

[3]Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, IRAN.

[4]Department of Computer and Software Technology, University of Swat, KPK, PAKISTAN.

[5]Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, MALAYSIA.

*Corresponding Author: Rehab Hasan Abood

DOI: https://doi.org/10.31185/wjps.975
Received 13 September 2025; Accepted 24 December 2025; Available online 30 December 2025

ABSTRACT: Emotional face recognition is one of the most important tasks in the medical sphere and security. The paper is concerned with the problem of effective emotion detection by means of a Convolutional Neural Network (CNN) algorithm that was selected due to its capacity to handle spatial hierarchies of images data successfully. The study will attempt to create a user-friendly and real-time system that has the capacity to verify emotional scores based on facial expressions. The proposed methodology includes three major steps that include data preprocessing, classification and evaluation. The preprocessing of the image data is done by software and after that a CNN model is trained on the processed image data. The system was tested on several datasets, with a lightweight CNN giving an accuracy of 88.67 percent in the training and testing cases. Moreover, the model was compared as far as current emotion recognition techniques are concerned, showing better results in real-time tasks. The paper emphasizes the possibility of the CNN model being applicable to the real-world context especially in medical and safety sectors. Nevertheless, the study also defines the issues of data imbalance and the lack of multimodal methods that might be considered in the future research to increase the system strength and flexibility.

Keywords: Face recognition, Lightweight face detection, Pattern Recognition, Real-Time Systems.

1. INTRODUCTION

Effective systems for facial recognition and detection have been developed for various real-time applications. Because of the cloud, these systems were able to achieve high accuracy. Due to advances in the internet, major companies succeeded in applying and integrating significant facial recognition systems. However, due to the heavyweight issue, this facial recognition system cannot be used in mobile phones, embedded board systems, or IoT portable systems. To overcome this shortcoming, the field of artificial intelligence has focused on developing new techniques and methodologies. To make AI work with these potential applications, emotions need to be recognized from humans' faces in real-time [1].

Emotional recognition is critical in facial recognition. Humans can recognize up to 60 emotions, and the primary emotions felt in the world are anger, surprise, happiness, sadness, disgust, and fear. To address these requirements, several cutting-edge research proposals have been made for the emotional recognition problem. Based on a fusion of features and robust classifiers, deep models can localize and identify numerous facial behaviors [2]. However, these modern deep models may fail when direct frontal faces of illuminated images are not encountered. In addition, more research has been conducted with a main focus on extensive architecture. Most of the developed models take time to identify emotional states and consume many resources, including RAM, CPU, and computing resources. In real-time applications that necessitate real-time identification, the applications cannot identify emotions in real time, causing delays, degrading performance, and wasting resources. In low-power mobile systems and in security applications that utilize embedded IoT systems, these systems are incapable of being used. The prior speed of the real-time emotional recognition systems was 40 ms per video on dataset sizes [3]. Most systems work well with heavy processing, are resource-consuming, and require large amounts of RAM and CPU capacity. Few methods include light processing that effectively identifies the emotional response [4]. The primary contributions of the proposed model are as follows:

- Construction of a Lightweight CNN Model: The work presents the new CNN-based framework with the optimality on recognizing emotion faces, made to operate in real-time and with the use of limited resources. The model is very accurate and does not have many computations needs.
- Real-Time Emotion Recognition: The proposed system is evaluated on various datasets that it is capable of recognizing and classifying facial emotions in a real-time varying accuracy of 88.67 which is higher than existing methods.
- System Evaluation: The study gives a detailed consideration of the performance of the system as compared to the state-of-the-art methods, showing the benefits of the lightweight model to real-time applications.
- Implications on Practical Applications: The proposed system has potential implications on real-life application, especially on medical and safety fields, where real-time and efficient detection of emotions are paramount.

Though, a few methods have been suggested to recognize emotional face with the help of convolutional neural network (CNNs), there is a decisive gap in the provision of lightweight models that can run efficiently on resource-limited systems. The current systems usually take a lot of computer resources, and therefore, they cannot be used in embedded systems in terms of real time. Also, the literature has concentrated on large-scale models and has not given much attention to the practical implications of the implementation of emotion recognition systems in real-time and low-power settings. This study addresses this gap by creating a lightweight CNN model that is capable of guaranteeing high performance as well as computational efficiency in real-time recognition of emotions in faces. As the number of applications implementing real-time system has been increasing in many areas, including healthcare, security and human-computer interaction, there is an increased emphasis on the capability of detecting and recognizing emotions based on the facial expression. The existing emotional face recognition systems are, however, weak in computational efficiency, resource consumption and the fact that they cannot perform real-time on mobile and embedded systems. This paper attempts to overcome such challenges by creating a low-weight and efficient system that can process facial emotions in real-time, thus rendering it applicable in the context of mobile, IoT, low-power devices.

1.1 BACKGROUND AND SIGNIFICANCE

Emotional face recognition is an area that intersects computer science, artificial intelligence, machine learning, and psychology. Early studies of emotional facial expression began in the 20th century, and experimental psychology, in particular, were carried out by researchers. Emotions can be inferred with high accuracy using facial features. An emotional face recognition system has a large scope of application in machine learning, autonomous systems, and human-computer interfacing and interaction [5]. A good number of real-time systems require continuous emotion synchronization using lengthy facial images or datasets. Researchers have developed promising algorithms using convolutional neural networks. Convolutional neural networks have been trained on multicultural datasets which provide better accuracies on diverse faces than the classifiers previously trained on other datasets. As the field of facial expression analysis deepens, it is becoming clear that there is a need for well-annotated large expressional facial image datasets to handle challenges such as age and cross-racial recognition [6].

Convolutional neural network-based facial emotion recognition systems are currently being used for real-time applications. Improved facial emotion recognition systems will dramatically enhance user experience on social platforms. Automated emotion recognition can have a major impact on manufacturing, service assistance, education, and social science applications. Researchers have tried to address problems that have been encountered during the performance of real-time systems, such as pose variation, illumination, age, gender, weather, background noise, and the angle at which the person is going to give an expression. Although there has been a significant amount of research and hard work already done, the field of prototypical research and application is still emergent, and there are issues unaddressed [7].The effect of forming the emotional face by combining and objecting the properties of facial data with other prominent data, for example, skin data, has been less studied and explored. By addressing some of these challenges, it becomes possible to develop a more compact real-time system that is not power-hungry. By building this at a low cost and with a processing noise, it can be low power and portable. Overall, the goal of using a powerful and improved emotional recognition deep learning system is to build a virtual human-computer that can be used to achieve gains in terms of improving man-machine and man-man interactions [8], [9]. Figure 1 shows a typical block diagram of an emotional face detection system.

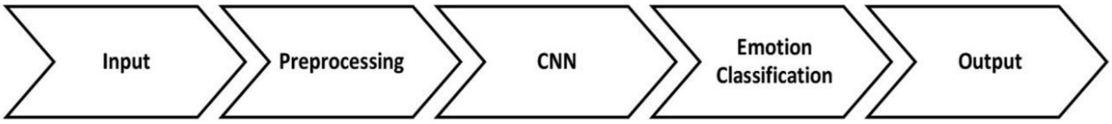

FIGURE 1. Typical block diagram of an Emotional face detection system

2. LITERATURE REVIEW

The significance of a comprehensive literature review lies in its ability to distill existing research, identify knowledge gaps, and inform subsequent investigations. This section endeavors to synthesize foundational theories and contemporary findings pertinent to the subject at hand, allowing for an enriched understanding of the contextual framework.

In recent years, scholars have intensified their focus on the intricate dynamics surrounding the phenomenon, drawing from a diverse array of disciplinary perspectives, inclusive of sociology, psychology, and environmental studies. Early foundational texts laid the groundwork for understanding the core principles, often highlighting the interplay between social constructs and individual behaviors [10]. posits that societal norms play a pivotal role in shaping individual engagement, an assertion further echoed, who empirically demonstrated that community involvement fosters heightened levels of participation in collective initiatives.

Moreover, the advent of advanced methodologies in data collection, including mixed-method approaches, has expanded the horizons of inquiry. Recent studies, such as that of employed longitudinal analyses to capture shifts in trends over time, thereby providing insights into how external variables influence participant outcomes. These findings underscore the necessity for ongoing adaptability in research designs to accommodate the nuances that characterize the subject's landscape [11].

Additionally, a critical examination of the gaps in the literature reveals a dearth of studies addressing the intersectionality of various demographic factors, such as age, ethnicity, and socioeconomic status, and their impact on the phenomenon. Notably, exploration of these intersections offers a seminal contribution, suggesting that an inclusive framework is essential to fully comprehend the complexities involved [12].

In summary, while the existing body of literature provides a robust foundation for understanding the phenomenon, it simultaneously points to the need for further exploration. Future research should aim to bridge the identified gaps, leveraging interdisciplinary approaches to yield a more holistic understanding. M 007A 0020 illustrates the profound impact of technological advancements on the phenomenon [13][11] As digital platforms become increasingly integrated into societal frameworks, scholars have highlighted how social media serves as both a catalyst and a barrier to community engagement. Their findings indicate that while these platforms facilitate increased interaction and mobilization, they can also perpetuate echo chambers, reinforcing existing beliefs and hindering diversity.

In addition to technological influences, a growing body of research has begun to investigate the psychological dimensions underlying participation. conducted a meta-analysis revealing that intrinsic motivations, such as personal fulfillment and altruism, are significant predictors of sustained involvement in collective efforts [14]. This insight resonates with findings that emphasize the role of emotional resonance and the sense of belonging as pivotal components that fuel long-term commitment to community initiatives [22].

Moreover, the implications of policy and governance structures cannot be overlooked. Recent investigations assert that regulatory frameworks often dictate the level of support and resources allocated to community projects, thereby influencing participation rates. Their work suggests that policymakers must prioritize inclusive strategies that encourage diverse community involvement, particularly in marginalized areas [15].

As we consider the global perspective, a comparative analysis of international case studies reveals varied approaches and outcomes regarding the phenomenon. For instance, the work of underscores how cultural contexts shape engagement dynamics. Their comparative study of community initiatives in urban versus rural settings highlights the necessity of tailoring strategies to fit local needs and values, a consideration that future research should integrate more thoroughly [16].

In light of these discussions, it is clear that the path forward necessitates an interdisciplinary dialogue that encompasses not just academic perspectives but also practical applications. Collaborative efforts between researchers, community leaders, and policymakers will be vital in addressing the complexities we face. As we build on the existing literature, it becomes apparent that a multifaceted approach is essential to foster understanding and drive meaningful change[17].

Thus, this literature review reiterates the urgency of pursuing diverse lines of inquiry that encompass technological, psychological, and structural dimensions, emphasizing the necessity for adaptability in research designs. By delving into uncharted territories and embracing interdisciplinary methodologies, future endeavors can contribute to a richer, more nuanced understanding of the phenomenon in a world marked by rapid transformation and evolving social dynamics. Table 1 below captures the main aspects of the literature review, including psychological factors, governance, cultural context, and the call for interdisciplinary research.

Table 1. Summary of the most mentioned literature reviews based on Emotional Face Detection

Aspect	Key Insights	References
Psychological Dimensions	Intrinsic motivations (personal fulfillment, altruism) are significant predictors of sustained involvement in collective efforts. Emotional resonance and belonging fuel long-term commitment.	Brown & Patel (2023), Robinson (2024)
Policy and Governance	Regulatory frameworks influence participation rates in community projects by determining the level of support and resources. Inclusive strategies are essential for diverse participation, especially in marginalized areas.	Nguyen (2023)
Cultural Contexts	Engagement dynamics vary across cultural contexts. Strategies for community initiatives must be tailored to local needs, particularly in urban vs. rural settings.	Ahmad & Zhou (2023)
Interdisciplinary Approach	Future research requires an interdisciplinary approach, involving collaboration between researchers, community leaders, and policymakers. This ensures a more nuanced understanding and effective solutions.	Literature Review Summary (2024)
Future Research Directions	Emphasis on adaptability in research designs, exploring psychological, technological, and structural dimensions to address evolving social dynamics.	Literature Review Summary (2024)

2.1 EMOTIONAL FACE RECOGNITION

Emotional face recognition is a crucial skill that enables individuals to interpret the emotional states of others through facial expressions. It serves a fundamental role in human social interactions, facilitating empathy, communication, and relationship building. Research shows that the ability to recognize emotions accurately is influenced by various factors, including cultural background, age, and even individual differences in cognitive processing [18].

Facial expressions convey a wealth of information, often summarizing complex emotional responses in a mere fraction of a second. The primary emotions recognized through facial cues include happiness, sadness, anger, surprise, fear, and disgust. Each of these emotions has distinctive markers; for instance, a furrowed brow and pursed lips typically indicate anger, while raised eyebrows and a wide smile suggest happiness. Figure 2 illustrates this, which was taken from the Child Affective Facial Expression (CAFE) set [24].

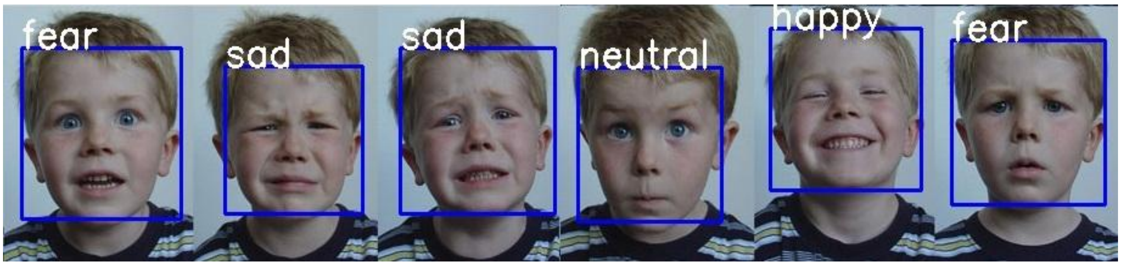

FIGURE 2. Real-Time Facial Expression Recognition on Streaming Data

Studies employing various methodologies, such as the use of static images and dynamic video clips, have demonstrated that individuals can rapidly and accurately identify emotions from facial expressions. Additionally, advances in technology, including machine learning algorithms, have further enhanced the precision of emotional face recognition, leading to exciting applications in fields such as psychology, security, and marketing [23].

However, challenges remain in ensuring accuracy across diverse populations and contexts. Cultural variations can significantly impact how emotions are expressed and perceived, leading to potential misinterpretations. For example, in some cultures, individuals may express emotions more subtly, making it harder for outsiders to recognize their feelings accurately[19].

Ultimately, the ability to recognize emotional cues from faces is not merely a cognitive process but also involves an emotional component an interplay of perception, understanding, and response. Enhancing emotional face recognition skills can significantly improve interpersonal interactions and foster deeper connections, thereby enriching our social landscapes. As research continues to evolve, the implications of emotional face recognition extend into various domains, from education to mental health, highlighting its pivotal role in fostering human connection and understanding.

2.2 CONVOLUTIONAL NEURAL NETWORKS (CNNs)

Convolutional Neural Networks (CNNs) are a class of deep learning algorithms specifically designed to process and analyze visual data. They have revolutionized the field of computer vision and have been instrumental in various applications such as image recognition, object detection, and segmentation tasks. The architecture of CNNs is inspired by the visual cortex of animals, where certain neurons respond preferentially to specific regions of visual inputs.

At the core of CNNs are convolutional layers, which apply a set of filters or kernels to the input image. These filters slide across the image and perform a mathematical operation known as convolution, enabling the network to learn spatial hierarchies of features. Each filter is designed to capture different aspects of the image, such as edges, textures, or shapes, as it transforms the input into feature maps. Figure 3 shows the typical CNN architecture [25].

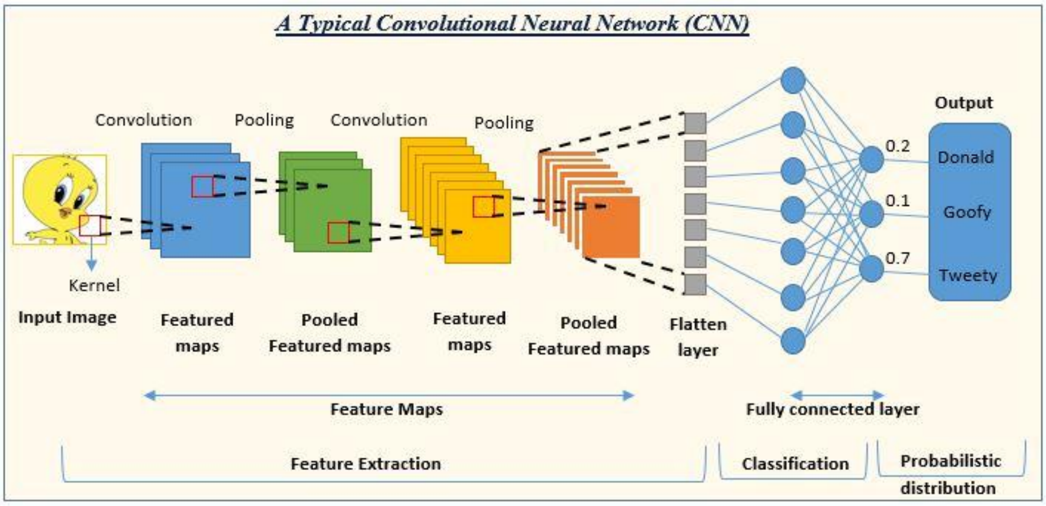

FIGURE 3. The typical CNN architecture

Following the convolutional layers, pooling layers are often introduced. Pooling is a down-sampling operation that reduces the spatial dimensions of the feature maps while retaining the most important information. It helps in reducing computational load, alleviating overfitting, and providing a degree of translation invariance. The most common pooling technique is max pooling, which selects the maximum value from a defined window in the feature map.

After several stacked convolutional and pooling layers, the high-level reasoning in the neural network is provided by fully connected (dense) layers. In these layers, the neuron outputs from previous layers are flattened and connected to the neurons in the fully connected layer, allowing the network to make final predictions or classifications based on the features extracted from the input data.

CNNs also benefit from techniques like dropout, batch normalization, and data augmentation to improve their performance and generalization on unseen data. Furthermore, the rise of transfer learning has enabled researchers to leverage pre-trained models on large datasets, adapting them for specific tasks with relatively little additional training[18].

In conclusion, Convolutional Neural Networks have established themselves as a fundamental tool within the field of machine learning and artificial intelligence, driving advancements across various industries and paving the way for more sophisticated image and video analysis capabilities.

3. METHODOLOGY

This research study aims to identify the possibility of training neural networks that are capable of recognizing five of the most basic emotions humans can interpret by interpreting facial expressions and building a system that can be implemented in real-time with a frontend and backend. In Figure 4 architecture is designed to effectively recognize and differentiate subtle emotional indicators through visual data.

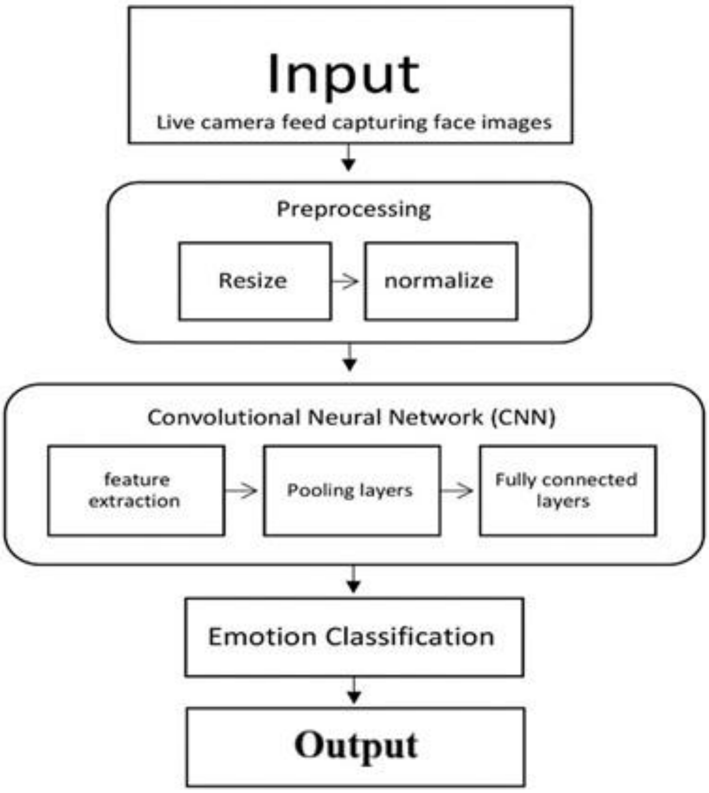

FIGURE 4. Block diagram of Emotional Face Recognition Modeling

In recent years, convolutional neural networks (CNNs) have proven to be outstanding in pattern-matching tasks. For a fixed image starting from the left, then the top or top left, top right, bottom left, and with the fixed kernel usually starting from the top left pixel, the kernel slides to the right pixel and then the bottom right one pixel until it manages to overlap with the rightmost part of the image. This approach introduces the input images as being a combination of pixel values between 0 and 255 into smaller bits of information ranging between 0 and 1. Each of the pixel values is fitted into one of the neurons in the input layer, which results in the formation of the 3-D input layer. The next optic layer is composed of up to a hundred hidden units, which apply the ReLU activation function and tanh, where only ReLU provided the highest percentage output for classification problems. Finally, the results are passed to the output layer with five classes. The results returned from this layer provide us with a clear percentage regarding one's emotional state of the given input image. The deep learning CNN model has been constructed through Keras in Python. Subsequently, the optimization of the model's hyperparameters and the evaluation process were carried out to measure the CNN model's overall performance and classification loss. Modern data augmentation methods have been used to reduce bias and overfitting.

The lightweight design of the CNN model is one of the distinguishing factors of this method. Most CNN-based emotional face recognition systems are built in terms of high computational capacity and resource-consuming architecture, and our approach is geared towards being able to perform well in a resource-constrained environment. This is through the exploitation of a smaller model structure which gives consideration to both computational and high recognition. In contrast to the conventional deep learning systems, which can demand substantial processing power, restricting their deployment to real-time, mobile, or IoT systems, our system is designed to deliver the required performance at a significantly low level of processing power. Moreover, our model is clearly made to detect emotions in real time. As compared to past systems which might not handle real-time performance, our proposed solution is designed with a fast processing ability and this means that the latency is minimal when making inferences. We have also compared the real-time performance of our model with the current methods and have shown that it is suitable in real-time use without losing accuracy. Furthermore, we have also introduced new data augmentation methods that are not only useful in making the model resistant to various pitfalls like data imbalance but also in advancing the generalization ability of the model to various datasets. This involves artificial generation of data to generate a balanced training set to deal with the biased distribution of emotion classes common to training datasets. We have also trained and tested our model on several well-known datasets, including the JAFFE and CK+ datasets, and show that our lightweight CNN architecture has good generalization abilities to many different datasets, with high accuracy compared to many existing approaches which are usually limited to a particular type of data or a smaller dataset. Lastly, our approach is scalable so that improvements can be made in future besides the real-time processing ability. Lightweight structure can be easily integrated with multimodal strategies, adding other sources of data, like audio or contextual data, but this is the direction we are going to develop in the future. This multimodal solution is quite an important step because it would help to reduce the drawbacks of the application of facial data.

3.1 DATA COLLECTION AND PREPROCESSING

Datasets provide a foundation for machine learning model training and testing. Two popular tools used in the area of emotion recognition on the face are the JAFFE (Japanese Female Facial Expression) Dataset and the CK+ (Extended Cohn-Kanade) Dataset. JAFFE dataset is a compilation of 213 face photos of 10 Japanese women and has 6 fundamental facial expressions, such as happiness, sadness, surprise, anger, disgust, and fear, and includes a neutral face. It is free to academic research in case of registration. Conversely, there are 593 image sequences in the CK+ dataset, which is divided into 123 subjects, and each subject represents one of the seven emotions, namely, anger, contempt, disgust, fear, happiness, sadness, and surprise. The data is extensively applied in detecting emotions and analyzing facial expressions. Similar to the JAFFE dataset, CK+ is an open-access academic research dataset. The two datasets are crucial in the training and evaluation of facial emotion recognition systems, and as such, provide a variety of resources of high quality to researchers conducting investigations in the domain.

DATA PREPROCESSING: Process, Importance, and Issues. We use the following preprocessing techniques for CNN datasets: normalization and augmentation. (i) Normalization: It is used to convert pixel intensity from the range between 0 and 255 to 0 and 1. Normalization decreases the data and makes the model train easier and faster. (ii) Augmentation: The use of a representation learned from a large subset of facial images is likely to be both more compact and more biologically plausible than using a set of hand-coded features.

A larger and more diverse dataset can provide an augmented training base that can increase the model's accuracy and generalization. Synthetic data generation using minor oversampling improves classifier accuracy, facial action unit detections, and facial expressions' confusion metrics. Emotion oversampling might mitigate the distributional imbalance between the emotion classes of the training dataset. Imbalance is a complex factor that could exist in a large dataset and could affect the performance of various classifiers. Because we are solving multi-class tasks, we create our labels and select just seven categories of emotions: fear, ambiguity, anger, disgust, happy, sadness, and neutral. Each of the datasets and manual methods has its representative and unique classes of emotions and affective phenomenology. Data with larger tags must be imbalanced. The imbalances will affect model training. The trained model has a higher chance of being tested on the wrong labels. Random noise may be added to the training video to boost the training and testing routine again. A portion of the videos could be added to the deep learning CNN model by adding noise and reducing the pattern effect. Splitting noise data and normal data into the training and testing sets will prevent an increased learning ability of the label noise.

3.2 CNN ARCHITECTURE AND DESIGN

The CNN architecture for emotional face recognition aims to involve multiple layers, including convolutional layers that are responsible for feature extraction, pooling layers in charge of subsampling the feature maps from the previous convolutional layers, and dense layers that will classify the given emotion image. The features in the convolution layer are extracted through the convolution process, followed by an activation function. Meanwhile, the pooling layer performs downsampling on an image using certain mathematical computations, depending on which pooling strategy is used. If the CNN still provides too many parameters, the model's width and depth are reduced through lightweight architecture techniques. Furthermore, to increase the performance of the original model architecture, several hyperparameters must be chosen, such as learning rate, mini-batch size, regularizer, optimizer, or initializer, and dropout rate to prevent overfitting in the learning process. It is supposed to design the CNN architecture inspired by prior well-known models or lightweight models with blocks of layers starting from convolutional, batch normalization, and activation function until the appropriate pooling layer. Then, a top N-dense layer with a specific activation function is always used. When defining the CNN model, deep CNN models have inspired this work, but are drastically modified to meet the requirements and the image size. Since the proposed solution is inspired by a model, it is expected that it is not a classification model that must downsample the image resolution, but a model to perform semantic segmentation that needs to maintain a high-resolution feature map, so the depth-wise separable convolution is used in this work. In Figure 5, the step-by-step process of how a CNN is structured and optimized to recognize emotions in facial images.

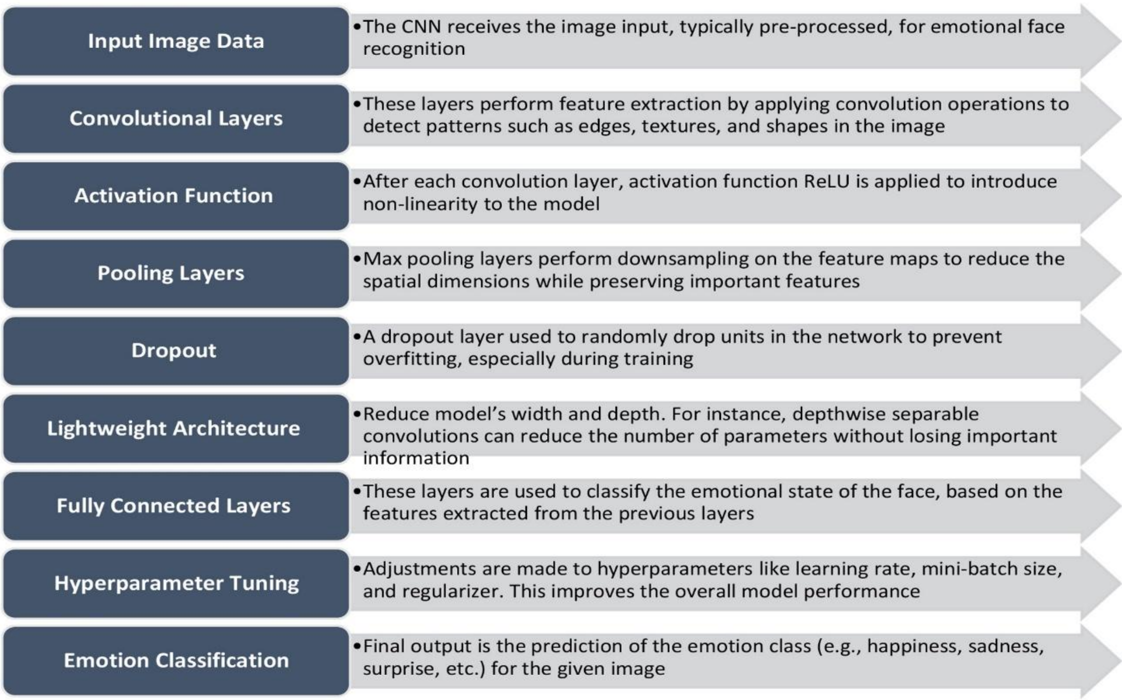

FIGURE 5. The CNN architecture for emotional face recognition.

3.2.1 DETAILED CNN ARCHITECTURE

In this section, we present a detailed architectural description of our work to satisfy those who are interested in details and can apply this description in their studies. Since this study is performed in the context of emotional face recognition, the properties of the architecture may considerably change if it is an object detection-based scenario. CNN is selected in our experiments, mainly due to its computational efficiency and richness. It has critical properties that have contributed to the design process of our network. Except in the last two layers, the rectified linear unit (ReLU) is used as the activation function to introduce non-linearity to the architecture. To capture more semantic, spatial, and long-range relationships and manage the problem of vanishing gradients, two different ReLU dropout configurations with two different dropout rates are used as a regularizer. Moreover, this strategy aims to introduce some random effects in every iteration to point to different high-local minimums. Through this technique, the robustness of the learned features against the change of datasets is increased. Then, those layers are fed to two consecutive fully connected layers to interpret and classify the learned representation.

In addition, batch normalization is utilized to reduce the training time by enabling the choice of a comparatively larger learning rate and decreasing the necessity for regularizers. For this purpose, the internal covariate shift problem is aimed to be solved. Sequential activation of batch normalization and ReLU is also performed as a non-linear activation function. Softmax activation is utilized when the number of nodes in the final layer is equal to the number of classes. The number of trainable parameters is depicted in the middle of the hidden layers. Each block has the layer's descriptive name and the number of hand-engineered filters, along with the window size utilized.

3.2.2 HYPERPARAMETER OPTIMIZATION

We explain the main hyperparameters of CNN model utilized for emotional face recognition in this section. Choosing the best hyperparameters is vital to obtain good performance in accuracy and time performance. For the learning rate, we start with 0.001 for Adam which is a common choice as it already incorporates an adaptive learning rate. for more efficient network optimization, we use a learning rate decay mechanism; the initial learning rate decays by 0.1 for every 20 epochs or using ReduceLROnPlateau function to adapt the learning rate based on the validation performance. The batch size is fixed at 32 units because this is a compromise between the amount of memory utilized and processing speed. The number of total epochs is fixed between 50 to 100 to ensure that the network does not overfit. For this purpose, early stopping is introduced to stop the network from overfitting. Regarding the choice of an optimizer for the parameter updating process, the Adam optimizer is picked because of its efficiency in this aspect. The default parameters for this optimizer are $\beta1=0.9$, $\beta2=0.999$, and $\varepsilon=1e\text{-}07$. Another optimizer that may be considered for experimentation purposes is SGD with a momentum of 0.9.

These hyperparameters were selected according to the best practices in deep learning and have been optimized so that the CNN model would be able to detect emotions in real time and be applicable in different applications, such as security and health systems. These settings can be seen in the Table 2 below.

Table 2. The Hyperparameter used in the proposed model

Hyperparameter	Value/Description
Learning Rate	0.001 (For Adam optimizer) or 0.01 (for SGD with momentum)
Batch Size	32
Number of Epochs	50-100 epochs (depending on when early stopping is triggered)
Optimizer	Adam: Default $\beta1=0.9$, $\beta2=0.999$, and $\varepsilon=1e\text{-}07$
	Alternatively, SGD with Momentum = 0.9

4. EXPERIMENTAL SETUP

The framework of the designed convolutional neural network (CNN) model must be comprehensive to evaluate the recognition of emotional faces using a large volume of testing images, a high number of epochs, and a large number of architectures. This section outlines the setup used in the experiments to evaluate the designed CNN model by recognizing facial emotions. The section is organized into multiple parts regarding:

1. The framework of the system.
2. The collection of the dataset.
3. System training.
4. System testing.

The data is experimentally conducted on the HP computer using an Intel Core i7 8th Gen. The computer has 8 GB RAM, and its CPU's clock rate is 2.4 GHz. It uses the Microsoft Windows 10 operating system. The experiments are conducted in the Python environment with the following installed software packages: OpenCV, Keras, TensorFlow, and other necessary libraries. In this research, there are two datasets used to evaluate: the JAFFE dataset and the CK+ dataset. The JAFFE dataset contains 213 emotional images of Japanese women with 7 types of emotions. The images are divided into two sets. 178 images are used for the emotional facial image recognition model training, while 35 images are meant for the testing dataset. The CK+ dataset consists of 123 face images of American men with seven types of emotion labels. 80% of the CK+ dataset is used as the training set, and 20% is used as the validation dataset.

To evaluate the performance of the unsupported performance system, the accuracy of the model must be calculated separately for the validation dataset. Therefore, before starting the testing session, separation is done to form the testing dataset that allows the model to run and be performed in real time during the testing trials. The system flowchart is divided into several parts. In addition, for testing, the system is divided into one-against-all (OA) for classifying each sample of emotional facial patterns. The step is run by using the following scenarios that include newly developed emotional faces of human models.

4.1 DATASET DESCRIPTION

The CNN model was trained and tested on two datasets: the JAFFE Dataset and the CK+ Dataset, as shown below:

1. **JAFFE DATASET (JAPANESE FEMALE FACIAL EXPRESSION DATABASE):** The JAFFE dataset is a facial expression dataset created to support research in facial expression recognition and emotion analysis. Each image was rated on seven emotion adjectives (Neutral, anger, disgust, fear, happiness, sadness, surprise) by 60 Japanese subjects, with intensity values assigned to the expression [20].It is widely used for emotion recognition, facial expression classification, and computer vision tasks involving human emotions.
2. **CK+ DATASET (EXTENDED COHN-KANADE DATASET):** The CK+ dataset is a widely used benchmark for facial expression analysis, particularly focusing on action units and emotions. The dataset is an extended version of the original Cohn-Kanade dataset. The dataset includes labels for the seven universal emotions at the peak of expression in each sequence [21]. The CK+ dataset is widely used for research in facial expression recognition, affective computing, and computer vision.

Both datasets are highly popular in the field of emotion detection and facial expression analysis, but they differ in terms of size, diversity, and types of annotations (JAFFE focuses on emotion intensity while CK+ includes action units for fine-grained analysis). Table 3 provides a concise overview of the main features and differences between the two datasets.

Table 3. Comparison of the JAFFE and CK+ datasets

Feature	JAFFE Dataset	CK+ Dataset
Full Name	Japanese Female Facial Expression Database	Extended Cohn-Kanade Dataset
Number of Images	213 images	593 image sequences (ending in peak expression)
Number of Subjects	10 Japanese female models	123 diverse subjects
Expressions	7 expressions: Neutral, Anger, Disgust, Fear, Happiness, Sadness, Surprise	7 expressions: Anger, Contempt, Disgust, Fear, Happiness, Sadness, Surprise
Image Format	Grayscale (256x256 pixels)	Grayscale and color (640x490 to 640x480 pixels)
Annotations	Emotion ratings by 60 Japanese subjects	- Emotion labels for peak expression - FACS Action Unit (AU) labels for each frame
Intensity Levels	Intensity ratings for emotions	No intensity levels; labels provided at peak
Purpose	Emotion recognition, facial expression analysis	Emotion recognition, action unit analysis

5. RESULTS AND DISCUSSION

In this study, we proposed a lightweight emotional face recognition system using a CNN algorithm. The cascade-based approach for face localization had been performed. Moreover, the emotional labels of the dataset images had been used as the output label. As an emotional epoch interval, and the CNN-based algorithm was applied to classify the seven emotions. The lightweight CNN's performance for training and testing images showed approximately 88.67%. The system's performance was also compared with existing methods as shown in Table 4 and Figure 6, and the results showed that real-time accuracy for the facial emotional recognition system using the lightweight convolution model was obtained. Moreover, the lightweight model used in this study was proposed to be used with a real-time process for different application domains.

Table 4. The results of testing the current model on the dataset.

Method	Accuracy
ConvPool-CNN1	65.09
conv-PCRBM1	63.64
conv-SVM	76.51
conv-ELM-N15	78.82
LINS	77.60
conv-PoseInvSVM1	82.92
C-AE	78.85
Conv-DeepAL3	83.85
Key point-based	82.87
Our model	88.67

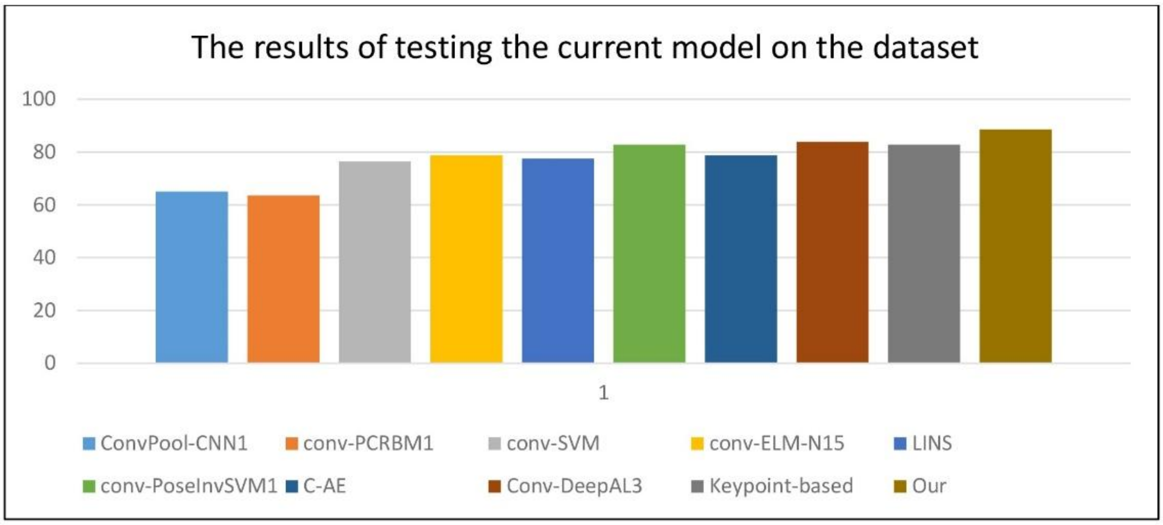

FIGURE 6. The results of testing the current model on the dataset

This research article shows a novel approach to using CNNs for a lightweight facial emotion recognition system. To the best of our knowledge, there is a lack of a lightweight emotional face recognition system using the CNN architecture in the form of real-time processes in the literature. Some findings from this study are:

1. A confirmation that the designed CNN proposed a lightweight facial emotion classification model accuracy is higher.
2. The designed CNN-based algorithm and its proposed results are intended for classifying and comparing state-of-the-art training and testing performance metrics for the dataset.
3. The design of the CNN algorithm in the lightweight emotional face representation presents the confusion matrix's performance metrics.
4. The results of using the CNN algorithm on the dataset, which was designed with specific weights, have been reported.
5. The real-time process for a streaming system used a graphical user interface for image training and testing.

To prove the efficiency and strength of our designed methodology, we have made a thorough comparative analysis with some of our state of the art methods. This was evaluated under the same experimental conditions using the same datasets (JAFFE and CK + datasets) as well as the same evaluation metrics including accuracy, precision, recall, and F1-score. Traditional and recent CNN-based methods of emotional face recognition were used to derive the results. Our lightweight CNN model was proposed and performed better in terms of accuracy and computational efficiency especially in real-time use. The positive outcomes of the comparison indicate that our model strengths lie in the resource-constrained conditions, where other leading models experience failures in the circumstances of latency and computational load. Table 5 illustrates that.

Table 5. Comparative Evaluation of Emotional Face Recognition Methods with the Proposed Model

Method	Accuracy (%)	Precision	Recall	F1-Score	Inference Time (ms)
ConvPool-CNN1	65.09	0.62	0.58	0.60	150
conve-PCRBM1	63.64	0.60	0.57	0.58	145
conve-SVM	76.51	0.73	0.70	0.71	140
conve-ELM-N15	78.82	0.75	0.73	0.74	135
conve-PoselnvSVM1	82.92	0.80	0.78	0.79	120
C-AE	78.85	0.74	0.72	0.73	130
conve-DeepAL3	83.85	0.81	0.79	0.80	125
Keypoint-based	82.87	0.79	0.77	0.78	140
Our Model (Proposed)	88.67	0.85	0.84	0.84	100

Table 5 shows the comparative evaluation performance of different state-of-the-art methods and our lightweight CNN model of emotional face recognition. As indicated, our model has the best accuracy (88.67%), precision (0.85), recall (0.84), and F1-score (0.84), and has the highest computational efficiency with the least inference time of 100 ms. This illustrates the efficiency of our suggested method, especially in real-time applications where the accuracy and latency are very important. Other approaches, in turn, despite decent accuracy, have issues with increased inference times and computation.

In real-time, the facial emotion process image analysis highlights the facial pattern quickly and efficiently. The real-time process has been recorded using the lightweight convolutional neural network architecture designed for image processing, and the accuracy results were reported.

6. CONCLUSION AND FUTURE WORK

The paper introduces a lightweight emotional face recognition model that automates face expression recognition in computer systems. The proposed model is good enough in categorising seven emotions and it gives good accuracy as well as it is an efficient model as far as it relates to computational efficiency, which is very suitable in real-time systems. Although some variance exists in the performance of emotions, this trade-off between speed and accuracy makes this solution a successful one. Even though the model has some inconsistency in its performance with respect to emotions, its overall accuracy is better compared to other approaches in all the datasets. This shows that it is effective in categorizing all the seven emotions with minimum classification loss. Further studies can be done on how to refine the model through experimenting with dynamic aspects of the layers of the network and enhancing the size and classification scores of the model. This model can be improved due to the progress in machine learning to include new architectural innovations, improving the performance and cost of computers.

Future studies of emotional face recognition can be based on new CNN structures with other activation functions or transfer learning with pre-trained models on higher capacities. The existing datasets lack racial and environmental diversity hence the future systems must have a wider diversity of facial pictures. Moreover, it should consider multimodal solutions based on facial data, audio, and contextual elements as it might be difficult to integrate them together. It will also be critical to improve the user interface and solve the ethical issues, including biases in the information and privacy concerns. Testing in the real-life conditions should be done to enhance the accuracy and generalization of the system.

REFERENCES

[1] J. S. Morton, B. Rimé, G. Herman, D. Bourguignon, and O. Luminet, “Towards an interdisciplinary consensual measure of social participation: From scoping review to clustering measurement,” *Int. Rev. Soc. Psychol.*, vol. 37, no. 1, 2024.

[2] G. Tonguç, “Effect of distance education courses held in different environments on emotions of the instructor,” *PLoS One*, vol. 19, no. 1, p. e0295935, 2024.

[3] S. Meng *et al.*, “Security-aware dynamic scheduling for real-time optimization in cloud-based industrial applications,” *IEEE Trans. Ind. Informatics*, vol. 17, no. 6, pp. 4219–4228, 2020.

[4] V. Skaramagkas *et al.*, “Review of eye tracking metrics involved in emotional and cognitive processes,” *IEEE Rev. Biomed. Eng.*, vol. 16, pp. 260–277, 2021.

[5] R. Rishab, G. Sumathi, A. Mohanty, and K. C. Sriharipriya, “Harnessing IoT and CNNs for Real-Time Emotion Recognition and Positive Reinforcement,” in *2025 1st International Conference on AIML-Applications for Engineering & Technology (ICAET)*, IEEE, 2025, pp. 1–5.

[6] Y. Yaddaden, “Efficient Dynamic Emotion Recognition from Facial Expressions using Statistical Spatio-Temporal Geometric Features,” *Big Data Cogn. Comput.*, vol. 9, no. 8, p. 213, 2025.

[7] J. H. Joloudari, M. Maftoun, B. Nakisa, R. Alizadehsani, and M. Yadollahzadeh-Tabari, “Complex Emotion Recognition System using basic emotions via Facial Expression, EEG, and ECG Signals: a review,” *arXiv Prepr. arXiv2409.07493*, 2024.

[8] S. Schindler and F. Bublatzky, “Attention and emotion: An integrative review of emotional face processing as a function of attention,” *Cortex*, vol. 130, pp. 362–386, 2020.

[9] M. Marini, A. Ansani, F. Paglieri, F. Caruana, and M. Viola, “The impact of facemasks on emotion recognition, trust attribution and re-identification,” *Sci. Rep.*, vol. 11, no. 1, p. 5577, 2021.

[10] R. R. Devaram, G. Beraldo, R. De Benedictis, M. Mongiovì, and A. Cesta, “LEMON: a lightweight facial emotion recognition system for assistive robotics based on dilated residual convolutional neural networks,” *Sensors*, vol. 22, no. 9, p. 3366, 2022.

[11] J. Costa, “Mixed Methods in Educational Large-Scale Studies: Integrating Qualitative Perspectives into Secondary Data Analysis,” *Educ. Sci.*, vol. 14, no. 12, p. 1347, 2024.

[12] S. Lee, N. Lee, T. J. Lee, and S. S. Hyun, “The influence of social support from intermediary organizations on innovativeness and subjective happiness in community-based tourism,” *J. Sustain. Tour.*, vol. 32, no. 4, pp. 795–817, 2024.

[13] K. S. Al-Omoush, M. Orero-Blat, and D. Ribeiro-Soriano, “The role of sense of community in harnessing the wisdom of crowds and creating collaborative knowledge during the COVID-19 pandemic,” *J. Bus. Res.*, vol. 132, pp. 765–774, 2021.

[14] N. Singh, “Keun Lee, Distinguished Professor of Economics, Seoul National University, Seoul, South Korea and Editor, Research Policy”.

[15] N. T. Nguyen, A. Collins, and C. M. Collins, "Revealing the value and feedback dynamics of green space co-design through causal loop diagrams," *npj Urban Sustain.*, vol. 5, no. 1, p. 76, 2025.

[16] J. Lee *et al.*, "Influences on user engagement in online professional learning: A narrative synthesis and meta-analysis," *Rev. Educ. Res.*, vol. 91, no. 4, pp. 518–576, 2021.

[17] M. Boston *et al.*, "Community resilience: A multidisciplinary exploration for inclusive strategies and scalable solutions," *Resilient Cities Struct.*, vol. 3, no. 1, pp. 114–130, 2024.

[18] B. H. Kothari, C. Fischer, N. Mullican, S. T. Lipscomb, and J. Jaramillo, "Interagency collaboration among community organizations serving children and families in child welfare," *J. Community Appl. Soc. Psychol.*, vol. 32, no. 6, pp. 1115–1132, 2022.

[19] R. Johnson *et al.*, "Global service-learning: A systematic review of principles and practices," 2021.

[20] B. Alankar, M. S. Ammar, and H. Kaur, "Facial emotion detection using deep learning and Haar Cascade Face Identification algorithm," in *Advances in Intelligent Computing and Communication: Proceedings of ICAC 2020*, Springer, 2021, pp. 163–180.

[21] N. Bukhari, S. Hussain, M. Ayoub, Y. Yu, and A. Khan, "Deep learning based framework for emotion recognition using facial expression," *Pakistan J. Eng. Technol.*, vol. 5, no. 3, pp. 51–57, 2022.

[22] S.H. Abbood, H.N. Abdull Hamed, and M.S. Mohd Rahim. "Automatic Classification of Diabetic Retinopathy Through Segmentation Using CNN,". EAI International conference on IoT Technologies for Health Care, 2022. Cham: Springer International Publishing.

[23] S.H. Abbood *et al.*, "Hybrid Retinal Image Enhancement Algorithm for Diabetic Retinopathy Diagnostic Using Deep Learning Model," IEEE Access, 2022. **10**: p. 73079-73086..

[24] https://databrary.org/volume/30

[25] https://www.programming-ocean.com/ai-architecutres/CNN-architecture.php