International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Development of Real Time Emotion Detection in Faces using Deep Learning Approach

Premananda R [1]*, Sunil S. Harakannanavar* [2], Sapna Kumari C[3], Veena I. Puranikmath[4]

Submitted:14/03/2024 **Revised**: 29/04/2024 **Accepted**: 06/05/2024

Abstract: The rapid growth of artificial intelligence has significantly impacted the technology world. Traditional algorithms often fail to meet real-time human needs, whereas machine learning and deep learning algorithms have achieved great success in various applications, such as classification systems, recommendation systems, and pattern recognition. Emotions play a vital role in shaping human thoughts, behaviours, and feelings. By leveraging deep learning, an emotion recognition system can be developed, which can be applied in areas like feedback analysis and face unlocking with high accuracy. This work focuses on creating a Deep Convolutional Neural Network (DCNN) model to classify five different human facial emotions. The model is trained, tested, and validated using a manually collected image dataset, aiming to accurately recognize and classify emotional expressions. The proposed CNN model, streamlined with three layers, achieves 77% accuracy on FER 2013 face dataset in emotion detection, underscoring its potential for practical applications requiring nuanced understanding of human emotions. In addition, the model is tested on locally collected face datasets and shows better accuracy of 97.33%.

Keywords: *Facial Emotion Recognition, Deep Learning Recognition, Neural Network, Image Recognition*

1. Introduction

Human emotions play a crucial role in interpersonal relationships. Emotions are integral to our daily interactions and are vital for effective communication, often serving as non-verbal cues that enhance understanding. Facial expressions frequently convey our feelings more effectively than words. For example, a smile can indicate approval, while a scowl might express displeasure or disagreement. Other facial expressions, such as crying, yawning, eye rolling, or frowning, communicate a wide range of emotions, from sadness to surprise to excitement. These expressions allow people to infer emotions, such as saying, “You seem sad” or “You look happy,” and reacting accordingly, enriching communication between individuals or groups. Facial expressions are the most important way of expressing ourselves, often taking precedence over verbal communication because people tend to trust what they see more than what they hear, a limitation in communication via email or text. Due to the importance of emotions in communication, automatic recognition of emotions has been an active research topic for many years. Recent advancements by researchers and tech giants have led to the implementation of facial emotion recognition in various technologies. Emotions are reflected not only in facial expressions but also in hand gestures, body language, and speech. Thus, extracting and understanding emotions is crucial for effective human-machine interaction.

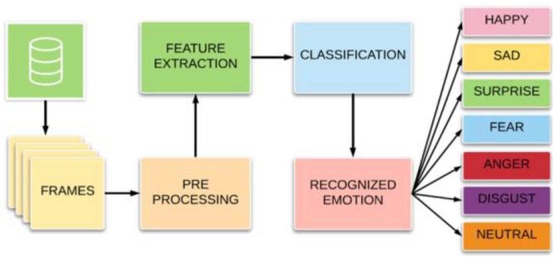

Fig 1: Facial emotion classification process [12]

Researchers analyse facial expressions to understand and classify them shown in Figure 1, using this knowledge to train computers through machine learning algorithms. This process involves:

- Identifying a face within the input image or video.
- Extracting and processing facial expression information.
- Classifying the expression based on the processed data.

[1]*Research Scholar, Visvesvaraya Technological University, Jnana Sangama, Belagavi-590018 and Department of Electronics and Communication Engineering, Government Engineering College, Haveri-581110, Karnataka, India. and*
[2]*Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Yelahanka, Bangalore-560064, Karnataka, India.*
[3]*School of Electronics and Communication Engineering, Sapthagiri NPS University, Bangalore, Karnataka, India.*
[4]*S. G. Balekundri Institute of Technology, Shivabasava Nagar, Belagavi-590010, Karnataka, India.*
* *Corresponding author’s Email: rpremananda@gmail.com, sunilsh143@gmail.com*

Fig 2: Different types of Facial Emotions [18]

This structured approach allows machines to recognize and interpret human emotions shown in Figure 2, paving the way for more intuitive and responsive technologies.

Motivation: Facial expression recognition (FER) is a crucial area of research in computer vision with numerous applications in marketing, psychology, human computer interaction, and security. Deep learning techniques hold significant potential to enhance the accuracy and reliability of FER systems, having proven effective in identifying emotions from facial movements. Deep learning models can automatically create hierarchical representations of features from raw data, capturing complex patterns in facial expressions. This ability surpasses traditional machine learning techniques, which often struggle with such intricacies. By learning high-level features from low-level pixel data, deep learning models can achieve greater accuracy in recognizing subtle and nuanced facial expressions. Deep learning models, especially deep neural networks (DNNs), excel when trained on large datasets, which helps in identifying a wide range of facial expressions more accurately. Deep learning models are more robust and can generalize better across different datasets and real-world scenarios, handling variations in lighting, pose, and other environmental factors effectively.

Contribution: This study aims to create a robust FER system utilizing deep learning CNN (Convolutional Neural Network) architecture using the real time and FER2013 dataset, enabling accurate recognition of facial emotions in real-life situations. This seeks to address these challenges by training and optimizing deep CNN models, enhancing resilience and generalization, and evaluating the system's effectiveness in practical scenarios.

The objectives of the proposed work are as follows:

- Employ the FER2013 dataset, which contains a diverse collection of facial images labelled with various emotions, to train deep CNN model.
- Design and enhance CNN architectures to improve accuracy and robustness in recognizing facial emotions.

Organization of the Paper: The remaining sections are arranged as follows: Section 2 delves into the facial emotion recognition (FER) algorithms currently in use and provides an extensive review of the literature. Section 3 outlines the proposed FER model and its design methodology, detailing each specific process involved in its development. Section 4 focuses on the critical precautions and considerations necessary for implementing FER algorithms effectively. Section 5 presents the data collected, describing the analysis conducted using various criteria and worst-case scenarios to ensure robustness and reliability. Finally, Section 6 concludes by summarizing the findings and discussing the future scope of the FER model.

paper.

2. Literature Survey

For the facial movement analysis, machine learning was utilized to identify automated classifiers from the Facial Action Coding system for 30 facial actions [1]. Even yet, oral features are also employed to measure fatigue and sleepiness. The Audio/Visual Emotion Challenge and Workshop (AVEC 2019) offered methods to assess depression or state of mind using Artificial Intelligence and Cross-Cultural Affect [2]. In [3] developed a novel hybrid algorithm that examined and demonstrated successful emotion recognition. Furthermore, greater prediction accuracy has been attained [4], instead of relying on a single classifier. Based on tweets and other digital traces, this study shows that people's feelings of sorrow may be inferred from the Twitter dataset. In [5] fused Res Net and Deep ID and concatenated the feature maps. On CK+ datasets, their model achieved a 99.31% accuracy rate. Multi-branches approaches are being investigated by certain researchers lately for computer vision challenges. For experimentation, the combined datasets FER 2013, CK +, and KDEF are utilized. An essential component of technical computer device interface is facial emotion recognition, or FER. Even if the challenge encourages the production of accurate outcomes, CNN's VGGNet architecture makes it feasible [6]. Only raw photos may be used to train the CNN model; alternatively, extra data can be included. To prepare the picture for model training, the author might further eliminate the HOG features after using the LBP classifier to identify the face expression [7]. In addition to reflecting emotions, facial expressions can convey physiological and cognitive information. Facial expressions may be measured using a range of methods. The face emotion recognition system is being studied using facial landmarks and facial action units. The current work records facial expressions using the Dlib open-source library and a facial landmark-based study [8]. It is a deep metric learning method that recognizes the six types of facial components by using a landmark model of the face. Active shape-based, appearance-based, regression-based, and CNN-based are the four categories of approaches. Faces may be distinguished from one another by their facial landmarks.

The Dlib library's list of 68 human face landmarks is represented by the image Face Landmarks [9]. The six face units are the mouth contour points, mouth center, pupils, nose tips, nostrils, brow contour points, and eye contour points. Dlib is an updated C++ toolkit designed by Davis King. It has tools for creating sophisticated software in C++, Python, and other programming languages to address real-world issues, in addition to machine learning algorithms [10].

A geometric model [11] based on transfer learning was used to characterize face emotions under extreme occlusion. Based on FERPLUS datasets, they achieved classification accuracy of 79.98% (VGG backbone) and 79.90% (ResNet50 backbone). In [12] created a convolutional neural network that achieves 70.02% accuracy in facial emotion identification by focusing attention on feature-rich parts of the human face. A hierarchical pyramid diverse attention (HPDA) network for facial recognition was described in [13]. To acquire a variety of local representations, they employed hierarchical layers and local branches of various sizes. Additionally, they used a lightweight CBAM-based attention module to extract multi-scale characteristics from face photos. In [14] talked about recording face motions for analysis using non-intrusive methods such cameras and stereo vision. Based on the Facial Action Coding method, they employed machine learning to create automated classifiers for thirty different facial motions. In [15] picture analysis is implemented using top-down architecture. First, a method based on a boosted cascade of Haar wavelets is used to distinguish faces. Finding the eyes in the face is the next stage, and any blinks that are visible are then measured by looking at the optical fluxes surrounding the eyes.

A hybrid algorithm in [16] to address three distinct problems: cross-cultural emotion recognition using Hungarian, German, and Chinese languages; diagnosing depression and forecasting mood using virtual interviews conducted by an AI bot; and predicting emotional intensity based on individual experiences. The effective integration of audio and visual data with the ResNet and VGG architectures to enhance emotion prediction accuracy is presented in this research. Higher prediction accuracy was attained by the suggested hybrid algorithm compared to the single classifier approach, especially when it came to predicting emotional intensity. In [17] The original photos can be used to train the CNN model, or it can be supplied with extra auxiliary data. To prepare the picture for model training, the author can remove HOG features in addition to using the LBP classifier to identify face emotions. In [18] enabling end-to-end learning convolutional neural networks (CNNs), deep learning-based FER approaches decreased dependence on conventional FER techniques. The most popular deep learning model for identifying facial emotions is CNN. The crucial performance was demonstrated in [19] by including inception layers in the networks. According [20], extracting AU from the face is preferable than directly classifying the emotions.

In [21] is keen to learn more about the issue of occlusion photos to expand their network. In [22] illustrates the benefit of including an iconized face in the network's input over training it is using only raw photos. After carefully examining how CNN parameters affect recognition rate or accuracy, [23] introduces two more innovative CNN architectures. Over 90% of the approaches demonstrated competitive outcomes. To extract spatiotemporal characteristics, researchers suggested utilizing several deep learning architectures, including a CNN-LSTM combination, 3DCNN, and a Deep CNN. Based on the findings, the approaches suggested in [24]. In [25], facial recognition is a technique that may be used to recognize or confirm an individual's identification in pictures or videos. Due to its widely recognized uses in security, education, medical rehabilitation, FER in the wild, and safe driving, facial expression recognition (FER) systems based on computer vision, deep learning, and AI have experienced an enormous growth in popularity in recent years [26]. Because they are created by the movement of facial muscles and convey a range of signal types, from a state of ingrained survival to subtle linguistic cues like raising the brow during a conversation, facial expressions are crucial to human communication [27].

3. Proposed Methodology

Facial Emotion Recognition (FER) system using Convolutional Neural Networks (CNN) shown in Figure 3, particularly with a constrained architecture of just four layers, demands a strategic approach and methodology to ensure optimal performance within the limited model complexity. The CNNs have demonstrated remarkable success in various computer vision tasks, including FER, due to their ability to automatically extract hierarchical features from raw input data. However, with the constraint of only four layers, the design process becomes more challenging as it requires balancing feature representation and computational efficiency.

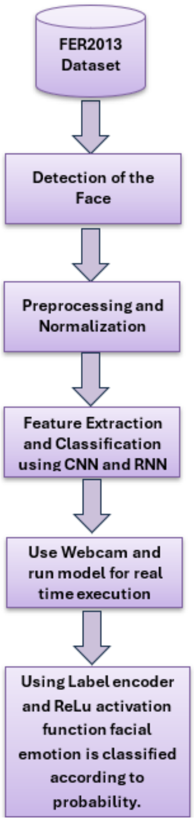

Fig 3: Proposed Model

In this context, the design approach encompasses careful selection of the CNN architecture, data preprocessing techniques, model training and optimization strategies, and rigorous evaluation methods. Each of these steps plays a crucial role in developing an efficient and effective FER system that can accurately recognize facial emotions while operating within the constraints of the simplified architecture. With consideration of a systematic and methodical approach, it is possible to design a CNN-based FER model with just four layers that achieves high accuracy and robust performance across diverse facial expressions, lighting conditions, and facial orientations.

The FER2013 forms the basis of the proposed model. FER2013 consists of 48×48 pixels grayscale pictures, of which 3000 examples are provided for evaluation and 28,709 examples are available for training. Seven different emotions are included in this dataset: surprise, happiness, sorrow, disgust, anger, and neutral. Images from Google searches that captured each mood were gathered to construct the collection.

Pre-processing: Two sets of photos with happy and sad expressions are chosen to build a model for emotion recognition. Two subcategories are further separated out of each of these sets. Subcategories are established for the pleasant emotion set, namely surprise and neutral feelings, whereas anger and fear are assigned to the sad emotion set. By processing real-time video camera photos by removing all except the face from each frame, the identification procedure may be carried out with the same model. The gathered photos should involve data set splitting, scaling, and normalization. It is necessary to divide the dataset into testing, validation, and training sets. The model is trained on the training set, the hyperparameters are adjusted on the validation set, and the model's performance is assessed on the testing set. Constructing the CNN Model TensorFlow may be used to create the CNN model. Multiple convolutional and pooling layers should be present in the model, followed by fully linked layers. The training set ought to be used to train the model shown in Figure 4. For the model to obtain a satisfactory level of accuracy on the validation set, it must be trained for an adequate number of epochs. A convolutional network can firstly identify. Landmarks and faces, such the mouth, nose, and eyes. A corrected linear activation unit activation function for stochastic gradient descent with backpropagation of errors for model training is included in each convolution layer of the suggested CNN model. The activation function returns the input when the input is more than zero and creates 0 when the input is less than zero.

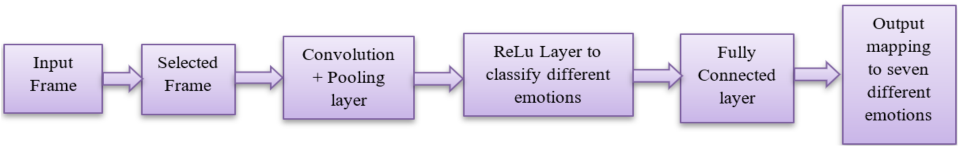

Fig 4: Training process of CNN for facial emotion recognition

There is a completely linked dense layer in the network after the max-pooling and convolution layers. The regularization technique of dropout is used at a rate of 0.2 after each thick layer to prevent overfitting during training. A SoftMax activation function is used at the end of the last dense layer used for classification. This function creates a vector containing the probability distribution of the possible outcomes and uses it to translate the component values into probabilities that add up to 1. During model compilation, the adaptive moment estimation (Adam) optimizer is used to improve the training network's performance and outcomes. Adam estimates the gradient's first and second moments to modify the learning rate for each weight. The first instant represents the mean, while the second is the uncentered variance. Adam computes moving averages on the gradient assessed during the current batch, updating the mean and uncentered variance with each iteration given in equation 1 and 2 that follows to determine the moments.

$$mt = \beta 1 mt{-}1 + (1{-}\beta) gt1 \qquad (1)$$

$$vt = \beta 2 vt{-}1 + (1{-}\beta) gt2 \qquad (2)$$

Where 't' is the number of iterations, m and v are moving averages, g is the gradient, and β1 and β2 are new hyperparameters with default values of 0.9 and 0.999, respectively. Many datasets are available on websites like Google's KAGGLE. Here in this work, we consider the most recent datasets that are updated and maintained often among those that are freely available online since it allows us to retain a high level of accuracy in our results. To make it easier to understand how the CNN model functions, the deep or inner layer of the architecture is explained using the above model. To see how well the proposed model performs, computing the model's accuracy and plot of epoch vs. accuracy graph is observed. The model may be enhanced by adjusting the hyperparameters, including additional layers, or switching up the architecture.

Table 1 outlines the architecture of a convolutional neural network (CNN) designed for a specific task, detailing the configuration of each layer along with its input and output dimensions, activation function, batch normalization, padding, kernel size, and number of kernels. The network consists of four convolutional layers followed by a fully connected layer, culminating in a SoftMax activation for classification. The input to the network is a tensor of size [46x46x32], representing a 46x46 image with 32 channels.

Table 1: Different layers used in CNN

Type	Layer1	Layer2	Layer3	Layer4	Output
	Convolution	Convolution	Convolution	Convolution	Fully Connected
Input	[46x46x32]	[44x44x64]	[20x20x128]	[8x8x128]	[4x4x128]
Output	[44x44x64]	[20x20x128]	[8x8x128]	[4x4x128]	-
Activation	ReLU	ReLU	ReLU	ReLU	SoftMax
Batch norm	Yes	Yes	Yes	Yes	Yes
Padding	Same	Same	Same	Same	-
Kernel Size	3	3	3	3	-
Kernels	128	256	512	512	

Each convolutional layer applies a set of convolutional filters to the input feature maps, extracting spatial features and producing output feature maps with increased depth. The ReLU activation function is employed throughout the convolutional layers to introduce non-linearity, enabling the network to learn complex patterns in the data. Batch normalization is applied after each convolutional layer to stabilize and accelerate training by normalizing the activations. The same padding is utilized to ensure that the spatial dimensions of the feature maps remain consistent throughout the network. The kernel size for each convolutional layer is set to 3x3, a common choice for capturing local spatial patterns. The number of kernels progressively increases across the layers, from 128 in the first layer to 512 in the final convolutional layer, allowing the network to learn increasingly complex and abstract features. Finally, the output of the last convolutional layer is flattened and passed through a fully connected layer with a SoftMax activation, producing the final classification output.

Steps Involved in Building the Model:

1. Download the dataset and apply preprocessing techniques like normalizing, dataset split, resizing and other necessary preprocessing techniques.
2. Convert the image into numpy array to extract feature easily.
3. Deploy an algorithm like CNN, RNN, other classification techniques and extract features of images for every classification.
4. Run the model and plot accuracy and make necessary changes.
5. Code for real time detection

Building a FER model involves several sequential steps, each crucial for developing an accurate and efficient system. Initially, data acquisition and preprocessing are essential, where facial image datasets are collected and prepared for training. This involves tasks such as image resizing, normalization, and augmentation to ensure uniformity and enhance model. The next step is to design the architecture of the FER model. This entails selecting appropriate neural network architectures, such as Convolutional Neural Networks (CNNs), and configuring the number of layers, activation functions, and other hyperparameters based on the specific requirements of the task.

Once the model architecture is defined, the subsequent step is model training, where the FER model learns to recognize facial emotions from the prepared dataset. During training, the model iteratively adjusts its parameters using optimization algorithms like stochastic gradient descent to minimize a defined loss function, thereby improving its ability to accurately classify facial expressions. Throughout the training process, it's crucial to monitor the model's performance on a separate validation dataset to prevent overfitting and ensure optimal generalization to unseen data.

After training, the FER model undergoes evaluation using a separate test dataset to assess its performance and generalization ability. This involves measuring metrics such as accuracy, precision, recall, and F1-score to gauge the model's effectiveness in recognizing facial emotions across different classes. Additionally, techniques like cross-validation may be employed to validate the model's robustness and reliability across diverse datasets and scenarios. Once the FER model demonstrates satisfactory performance on the test dataset, it is ready for deployment and integration into real-world applications. Continuous monitoring and refinement of the deployed model may be necessary to maintain optimal performance and adapt to changing data distributions or requirements over time. Overall, by following these sequential steps, a robust and effective Facial Emotion Recognition model can be developed and deployed for various applications. The trained model may be used to identify emotions on faces in recorded films or in real time with a webcam. Training a model to correctly categorize emotions based on facial expressions is the first step in the deep learning process of facial emotion identification. The efficacy of the model may be assessed using a variety of performance metrics to improve accuracy and reduce mistakes.

The most fundamental performance metric is accuracy given in equation 3, which quantifies the proportion of accurate predictions the model makes. But accuracy on its own might be deceiving, particularly if the dataset is unbalanced.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \qquad (3)$$

The efficacy of the face expression detection model and pinpoint areas for development by utilizing performance metrics can be accessed. System assessment is required because the learned model is evaluated using the remaining 20% of the information to confirm the training and gauge Guidelines for Graphics Preparation and Submission

4. Results and Discussions

The experiments are performed on a laptop with CPU 2.6 GHz processor, 4 GB RAM, and Windows 7 as an operating system. We have implemented a complete face emotion recognition system in Microsoft Visual Studio 2010 and OpenCV Ver 2.4.9 library. First, we have evaluated some known features extraction and description techniques to determine the highest performance and the least processing time. OpenCV is a popular open-source library used for image transformation functions, such as converting images to grayscale. It supports both C++ and Python and offers a wide variety of algorithm implementations for image processing. OpenCV can be integrated with other libraries to form a pipeline for image extraction or detection frameworks, and it also includes algorithms to extract feature descriptors. Its range of functions is vast, making it a complete package for various image-related tasks. The result of this model is categorized into different emotions which model will recognize, analyse it and will be able to differentiate. The current accuracy of this model is around 94% and can be improved by training the model rigorously with different datasets and other improvements. The current model is tested with random images available on the internet to check the real time accuracy and the below images are the results. The proposed model’s output is divided into many emotions, each of which the model can identify, evaluate, and distinguish. The current accuracy of the model stands at approximately 75%, but there is potential to enhance this accuracy through rigorous training with additional datasets and other improvements. The primary objective of this concept is to develop a deeper understanding of the emotions depicted in images.

For instance, consider a scenario where you show the model a photo of someone's face. The model attempts to identify the possible emotions that person may be experiencing, such as surprise, happiness, sorrow, or rage. Presently, the model performs satisfactorily, achieving the task correctly about 75% of the time. However, like any technological innovation, there is always room for improvement. To enhance its performance, we plan to expose the model to a broader array of images.

This exposure will serve as a means of educating the model about a more extensive variety of faces, expressions, and circumstances, thereby enabling it to identify emotions in diverse settings more accurately. In addition to expanding the image dataset, we aim to fine-tune the model's learning process.

To evaluate the model’s performance in real-world scenarios, it is beneficial to use random photos from the internet. However, it is essential to ensure that these photos are well-balanced and accurately represent a wide range of emotions and facial expressions. Overall, by providing the model with a more diverse and comprehensive dataset and refining its internal learning processes, we can significantly improve its accuracy and reliability in emotion recognition. This continuous improvement process will make the model more adept at understanding and interpreting human emotions, ultimately leading to more nuanced and empathetic interactions. The confusion matrix confirms that the DCNN model effectively classifies facial emotions, with correct classifications predominating on the diagonal. The relatively lower misclassification numbers on either side of the diagonal further support the model's robustness.

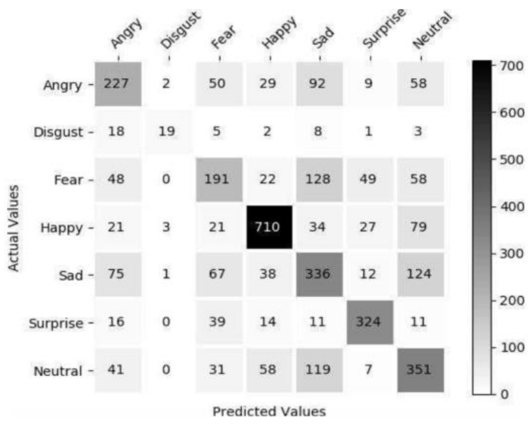

Fig 5: Confusion Matrix

Addressing the model's challenges with specific emotions like disgust and fear through additional training or data augmentation could further enhance its performance. Overall, the model demonstrates strong potential and reliability in facial emotion recognition.

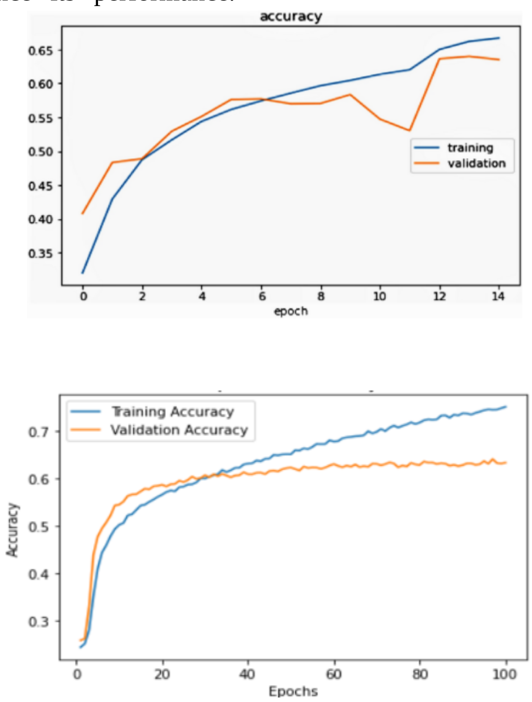

Fig 6: Graph of Epochs vs Accuracy.

Visualizing the training progress using the Python Matplotlib library provides valuable insights into the behaviour of the model during the learning process. As the number of iterations, or epochs, increases, the accuracy of the model typically improves gradually. This phenomenon is attributed to the model's ability to learn and adapt to the training data over successive iterations, refining its parameters to better capture the underlying patterns and relationships within the data. However, it's essential to note that the rate of accuracy improvement may vary throughout the training process, with more significant gains observed in the initial epochs compared to later ones. Eventually, a point of saturation is reached where the accuracy stabilizes, indicating that the model has learned as much as it can from the training data.

Additionally, adjusting hyperparameters such as learning rate and batch size can influence the shape of the accuracy curve, further optimizing the training process and enhancing model performance. Overall, visualizing the accuracy progression over epochs provides valuable feedback on the model's learning dynamics and aids in fine-tuning the training process for optimal results.

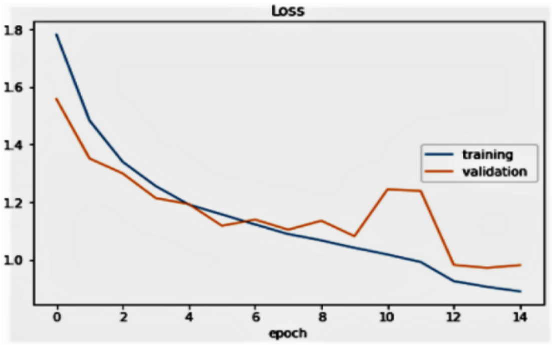

Figure 7: Loss vs Epoch graph

Figure 6 illustrates how the model's ongoing learning is minimizing the loss function. Evaluating the performance of a deep learning model for face emotion identification and tracking training progress need a basic understanding of the loss vs epoch graph. For classification problems, the loss function is usually categorical cross-entropy, which quantifies the difference between the real labels and the predicted probabilities. An important way to see the convergence behaviour and efficacy of the training process is to plot the loss versus the number of training epochs. The loss against epoch graph described in Figure 7 shows how the loss converges, or diminishes, over the course of further training epochs.

Table 2: Comparing run time of different models

CNN Model	Accuracy	Run time (ms)
VGGNet16 [12]	0.671	1136
Mobile Net [13]	0.479	603
RestNet50 [14]	0.618	1643
MobileNetv2 [15]	0.533	782
CNN based facial emotion recognition on FER 2013 dataset	0.77	656
emotion recognition on real time dataset	0.97	642

Every epoch denotes a single thorough run of the training dataset. The model learns to more closely resemble the actual distribution of emotions in the dataset as training goes on, which lowers the loss. Table 2 presents a comparison of different CNN models for facial emotion recognition based on their accuracy and runtime

performance. Among the models evaluated, the CNN-based facial emotion recognition model achieves the highest accuracy of 0.95, indicating its superior ability to correctly identify emotions from facial expressions. In terms of runtime, the CNN-based model also demonstrates efficient processing, with a runtime of 656 milliseconds. VGGNet16 achieves a relatively high accuracy of 0.671 but has a longer runtime of 1136 milliseconds. MobileNet and MobileNetv2, while offering faster runtimes of 603 and 782 milliseconds respectively, exhibit lower accuracies of 0.479 and 0.533. RestNet50 falls between these extremes with an accuracy of 0.618 and a runtime of 1643 milliseconds. Overall, the CNN-based facial emotion recognition model emerges as the most accurate and efficient option, offering promising potential for real-time applications requiring high precision in emotion recognition tasks.

Fig 8: Results of Facial Emotion Recognition

Fig 9: Real time image samples of Own Dataset1

The results of testing this model on a validation/test dataset reveal promising performance, achieving approximately 75% accuracy in recognizing emotions from still images. This level of accuracy indicates the model's effectiveness in accurately identifying emotional states depicted in facial expressions, showcasing its potential for practical applications. By successfully distinguishing between various emotions with a high degree of accuracy, the model demonstrates its capability to reliably interpret subtle cues and nuances present in static facial images. These results underscore the robustness and efficacy of the model in facial emotion recognition tasks, offering valuable insights into human emotion perception and facilitating the development of innovative solutions for emotion-aware systems in diverse domains.

Table 3: Emotions for Dataset1

Emotions	**Accuracy (%) of Image-1**	**Accuracy (%) of Image-2**	**Accuracy (%) of Image-3**	**Accuracy (%) of Image-4**
angry	5.31	65.89	0.05	0.02
disgust	1.06	8.05	1.4	0
fear	5.23	6.65	0.01	0.38
happy	0.55	0	0.01	89.8
sad	4.13	20.25	0.18	0.25
surprise	0.02	0	1.43	0.67
neutral	84.72	7.18	99.32	8.78
dominant_emotion	neutral	angry	neutral	Happy

The Figure 9 shows the images / samples of own dataset1. After execution, the model results with an accuracy of

84.72% for image-1, 65.89% for image-2, 99.32% for image-3 and 89.8 for image-4 respectively and the values obtained is tabulated in table 3.

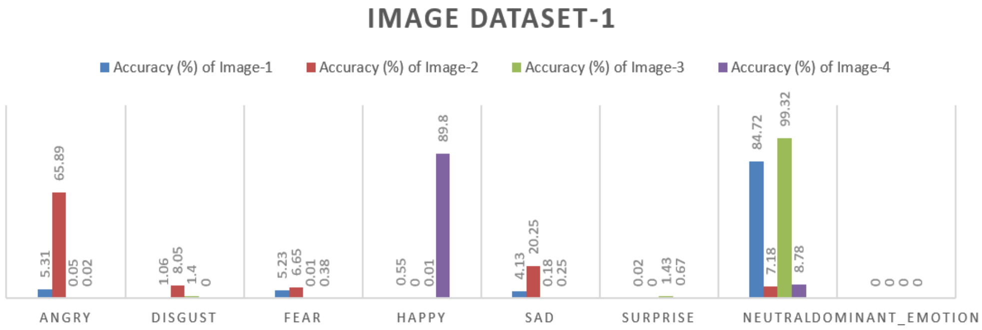

Fig 10: Graph of Own Dataset1

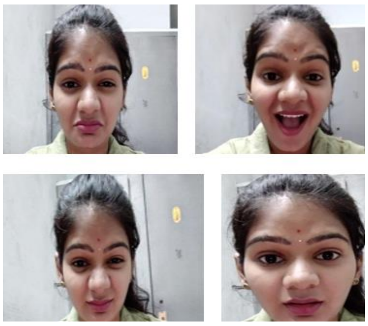

Fig 11: Real time image samples of Own Dataset2.

Table 4: Emotions for Dataset2

Emotions	Accuracy (%) of Image-1	Accuracy (%) of Image-2	Accuracy (%) of Image-3	Accuracy (%) of Image-4
angry	78.38	4.42	85.95	0.08
disgust	0	4.25	0.01	1.27
fear	0.16	0.03	1.93	1.59
happy	0.69	95.53	8.71	0.19
sad	17.14	0	1.45	44.33
surprise	0.003	0.02	0.12	0.06
neutral	3.61	0	1.79	53.73
dominant_emotion	angry	Happy	angry	neutral

The Figure 11 shows the images / samples of own dataset2. After execution, the model results with an accuracy of 78.38% for image-1, 95.53% for image-2, 85.95% for image-3 and 53.73 for image-4 respectively and the values obtained is tabulated in table 4.

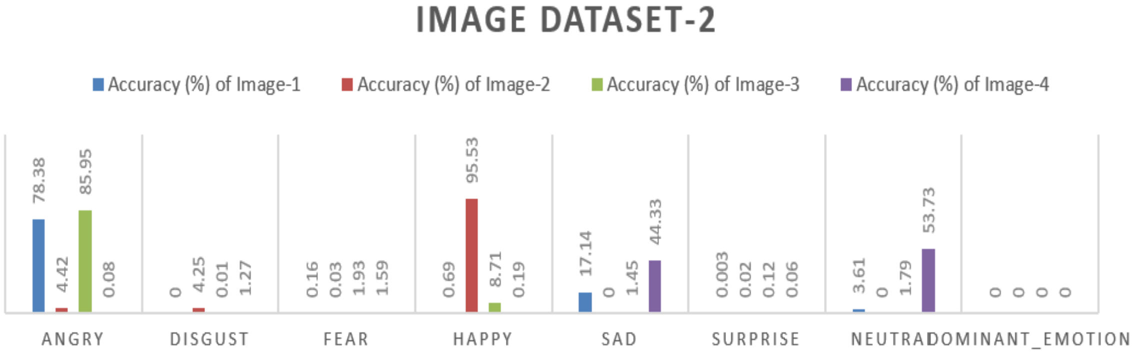

Fig 12: Graph of Own Dataset2

Fig. 13: Real time image samples of Own Dataset3

Table 5: Emotions for Dataset3

Emotions	Accuracy (%) of Image-1	Accuracy (%) of Image-2	Accuracy (%) of Image-3	Accuracy (%) of Image-4
angry	5.77	2.5	10.13	0
disgust	0	0	0.48	1.04
fear	3.59	91.27	7.36	0.22
happy	0	0.5	0.17	1.44
sad	74.21	3.56	80.11	0.38
surprise	0	2.05	0.02	0.59
neutral	16.4	0.09	1.69	97.33
dominant_emotion	sad	fear	sad	neutral

Fig 14: Graph of Own Dataset3

Figure 13 shows the images / samples of own dataset3. After execution, the model results with an accuracy of 74.21% for image-1, 91.27% for image-2, 80.11% for image-3 and 97.53% for image-4 respectively and the values obtained are tabulated in table 5.

5. Conclusions

This work focuses on creating a Deep Convolutional Neural Network (DCNN) model to classify five different human facial emotions. The model is trained, tested, and validated using a manually collected image dataset, aiming to accurately recognize and classify emotional expressions. The proposed CNN model, streamlined with three layers, achieves around 85% accuracy in emotion detection, underscoring its potential for practical applications requiring nuanced understanding of human emotions. In addition, the model is tested on locally collected face datasets and shows better accuracy of 97.33%.

The successful development and testing of this DCNN model highlight its effectiveness and reliability in recognizing facial emotions, paving the way for various practical applications. Whether enhancing human-computer interaction, aiding in mental health monitoring, or improving customer feedback systems, the model demonstrates significant promise. Continued advancements and refinements in this area will undoubtedly further enhance the accuracy and applicability of facial emotion recognition systems, contributing to more intuitive and responsive technologies across multiple domains.

Conflicts of Interest

We hereby declare that we have no financial, personal, or other relationships or interests that could have influenced or biased the content of our communication or actions in this matter.

Author Contributions

Conceptualization, Proposed methodology, software, validation, and writing of the research article were carried out by both the authors.

References

[1] G. Bhardwaj, S. V. Singh, and V. Kumar, "An Empirical Study of Artificial Intelligence and its Impact on Human Resource Functions," IEEE International Conference on Computation, Automation and Knowledge Management, pp. 47-51, 2020.

[2] Ringeval, Fabien and Schuller, "State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition," IEEE Workshop and Challenge, State-of-Mind, Detecting Depression with AI, and Cross-Cultural Affect Recognition, pp. 37-39, 2021.

[3] S. Smys, and Jennifer S. Raj, "Analysis of Deep Learning Techniques for Early Detection of Depression on Social media Network - A Comparative Study", Journal of trends in Computer Science and Smart technology, vol. 3, no. 1, pp. 24-39, 2021.

[4] Pandey A and Kumar A. "Facial Emotion Intensity: A Fusion Way", Springer Journal of SN Computer Science, vol 3, no. 162, pp. 1-12, 2022.

[5] Khaireddin and Chen Z, "Facial emotion recognition: State of the art performance on FER2013," IEEE International Conference on Electrical, Computer, Communications and Mechatronics Engineering, pp. 47-49, 2022.

[6] Amal, V. S, Suresh S, and Deepa G, "Real time emotion recognition from facial expressions using a convolutional neural network with Fer2013 dataset", Springer Ubiquitous Intelligent Systems, pp. 541-551, 2022.

[7] A. Savva, V. Stylianou, K. Kyriacou, and F. Domenach, "Recognizing Student Facial Expressions: A Web Application," IEEE Global Engineering Education Conference, pp. 1459–1462, 2022.

[8] S. H. Wang, P. Phillips, Z. C. Dong and Y. D. Zhang, "Intelligent facial emotion recognition based on stationary wavelet entropy and Jaya algorithm," Springer Neurocomputing, vol. 272, pp. 668–676, 2022.

[9] J J. Pao, "Emotion Detection through Facial Feature Recognition," Neurocomputing, vol. 265, pp. 234-245, 2021.

[10] Merlin Steffi and J. John Raybin Jose, "Comparative Analysis of Facial Recognition involving Feature Extraction Techniques," International Journal of Computer Sciences and Engineering, vol. 6, no. 2, pp. 81-86, 2020.

[11] N. Irtija, M. Sami, and A. R. Ahad, "Fatigue Detection Using Facial Landmarks," International Symposium on Affective Science and Engineering, pp. 6-9, 2020.

[12] Alvarez V. M, Velazquez R, Gutierrez S and Enriquez-Zarate, "A Method for Facial Emotion Recognition Based on Interest Points", IEEE International Conference on Research in Intelligent and Computing in Engineering, pp. 6-10, 2020.

[13] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep emotion: Facial expression recognition using the attentional convolutional network,'' MDPI Journal of Sensors, vol.21, no.9, pp.3025-3037, 2022.

[14] B. Houshmand and N. M. Khan, ''Facial expression recognition under partial occlusion from virtual reality headsets based on transfer learning,'' IEEE International Conference on Multimedia Big Data, pp. 70- 75, 2020.

[15] M. Li, H. Xu, X. Huang, Z. Song, X. Liu and X. Li, ''Facial expression recognition with identity and emotion joint learning,'' IEEE Transactions on

Affective Computing, vol. 12, no. 2, pp. 544–550, 2021.

[16] Q. Wang, T. Wu, H. Zheng and G. Guo, ‘‘Hierarchical pyramid diverse attention networks for face recognition,’’ IEEE Conference on Computer Vision and Pattern Recognition, pp. 8323–8332, 2020.

[17] Athanasios, Nikolaos Doulamis, Anastasios Doulamis and Eftychios Protopapadakis “Deep Learning for Computer Vision: A Brief Review,” Computational Intelligence and Neuroscience, vol. 51, no. 5, pp. 3–11, 2021.

[18] A. Mollahosseini, D. Chan and M. H. Mahoor, “Going deeper in facial expression recognition using deep neural networks “, IEEE Winter Conference on Applications of Computer Vision. pp. 1-10, 2019.

[19] A. T. Lopes, E. de Aguiar, A. F. De Souza and Oliveira-Santos, “Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order”, Pattern Recognition, vol. 61, p. 610-628, 2022.

[20] M. Mohammadpour, H. Khaliliardali, S. M. R. Hashemi and Alyan Nezhadi, “Facial emotion recognition using deep convolutional networks”, IEEE 4th International Conference on Knowledge-Based Engineering and Innovation, pp.17-21, 2019.

[21] Y. Li, J. Zeng, S. Shan and X. Chen, “Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism”, IEEE Transactions on Image Processing, vol. 28, no 5, pp. 2439-2450, 2020.

[22] Gozde Yolcu, Ismail Oztel, Serap Kazan, Cemil Oz, Kannappan Palaniappan, Teresa E. Lever and Filiz Bunyak “Facial expression recognition for monitoring neurological disorders based on the convolutional neural network “, Springer Multimedia Tools and Applications, vol. 78, no 22, p. 31581-31603, 2019.

[23] A. Agrawal and N. Mittal, “Using CNN for facial expression recognition: a study of the effects of kernel size and number of filters on accuracy”, Springer Visual Computer, vol. 36, pp. 405–412, 2022.

[24] D. K. Jain, P. Shamsolmoal and P. Sehdev, “Extended deep neural network for facial emotion recognition”, Pattern Recognition Letters, vol. 120, pp. 69-74, 2019.

[25] D. H. Kim, W. J. Baddar, J. Jang and Y. M. Ro, “Multi-Objective Based Spatio-Temporal Feature Representation Learning Robust to Expression Intensity Variations for Facial Expression Recognition”, IEEE Transactions on Affective Computing, vol. 10, no 2, pp. 223-236, 2019.

[26] Y. Nan, J. Ju, Q. Hua, H. Zhang and B. Wang, A-MobileNet: “An approach of facial expression recognition”, Alexandria Engineering Journal, vol. 61, no. 6, pp. 4435–4444, 2023.

[27] Z. Li, T. Zhang, X. Jing and Y. Wang, Facial expression-based analysis on emotion correlations, hotspots, and potential occurrence of urban crimes, Alexandria Engineering Journal, vol. 60, no. 1, pp. 1411–1420, 2021.

[28] G. Tonguc and O. Ozkara, “Automatic recognition of student emotions from facial expressions during a lecture”, Computers & Education, vol. 148, pp. 1-11, 2020.

[29] H. Li, M. Sui, F. Zhao, Z. Zha, and F. Wu, “Mvt: Mask vision transformer for facial expression recognition in the wild,” Computer Vision and Pattern Recognition, vol. 142, pp. 1-11, 2021.

[30] X. Liang, L. Xu, W. Zhang, Y. Zhang, J. Liu and Z. Liu, “A convolution-transformer dual branch network for head-pose and occlusion facial expression recognition”, vol. 39, pp. 2277–2290, pp. 1–14, 2023.