Beyond Basic Emotions: Deep Neural Networks for Compound Facial Expression Detection

[1]Shabanam Bano*, [2]Pawan Bhambu

[1, 2]Department of Computer Science and Engineering, Vivekananda Global University, Jaipur, Rajasthan, India.

[1]shabanam.bano.professional@gmail.com

[2]pawan.bhambu@vgu.ac.in

Abstract. In this research, a fundamental method for deep learning-based emotion identification from facial photographs is presented. Happiness, sorrow, surprise, rage, and other basic emotions are automatically extracted and classified using a Convolutional Neural Network (CNN) model. The model's performance in emotion identification tasks is assessed after it has been trained on a dataset of photos. The CNN model successfully captures facial features for categorization and achieves adequate accuracy, according to experimental data. A fundamental framework for emotion recognition is provided by this work, which can be expanded with more sophisticated methods for better results. According to experimental data, our sequential CNN-based approach performs more accurately than conventional machine learning models, especially when detecting ambiguous or insignificant emotional cues. The results demonstrate how CNNs can improve emotion detection models for social media applications and provide more in-depth understanding of online user behavior, content engagement, and social dynamics with approximately 98% accuracy.

Keywords: Emotion Detection, Basic & complex emotion detection, Machine Learning (ML), Deep Learning (DL), Convolutional Neural Network (CNN).

1 Introduction

Basic emotions such as joy, grief, fear, wrath, and disdain are often combined with secondary emotions resulting from the interpretation and rating of the experience to create complex emotions. These secondary emotions include feelings of pride, envy, shame, guilt, empathy, and many more. Remorse, ambivalence, wonder, compassion, gratitude, envy, nostalgia, and moral anger are examples of complex emotions. Even so, they typically include a range of emotions, both good and bad, as well as mental assessments and physical adjustments. These emotions can be challenging to precisely characterize due to their complexity and subjectivity. One of the most significant issues facing the scientific study of emotion is the connection between basic emotions and these complex emotion experiences [1], [6]. In the continuation of this, the machine learning (ML) community has recognized the deep learning (DL) computer paradigm as the gold standard in recent years. Additionally, it has steadily emerged as the most popular computational method in machine learning, thereby accomplishing exceptional performance on a number of challenging cognitive tests, matching or even surpassing

B. Singh et al. (eds.), *Proceedings of the International Conference on Advances in Computing Technology and Artificial Intelligence (COMPUTATIA 2026)*, Atlantis Highlights in Intelligent Systems 18,
https://doi.org/10.2991/978-94-6239-713-2_15

human performance [2],[7]. Deep learning algorithms are made to behave similarly to how the neural network of humans does. Deep neural networks, or neural networks with numerous hidden layers, are represented by these methods. Convolutional neural networks, a type of deep learning technique, can train enormous datasets with millions of parameters using 2D images as input and convolve them with filters to produce the desired outcomes [3],[8]. We can do a lot of work with CNNss (DL algorithms). We can work with complex architectures like ResNet and Inception as well as simple picture classification problems [5]. When you require excellent performance but don't have a lot of data, domain-specific tasks (like facial recognition or medical imaging), bespoke behavior during training, several custom layers, etc.

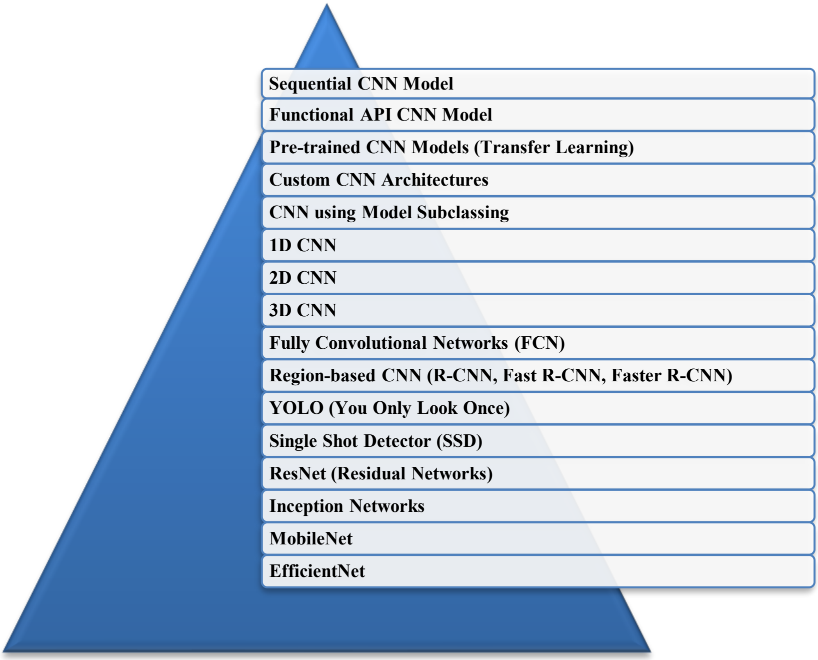

Fig. 1 CNN Models

Fig. 1 is showing the list of CNN models, and I'm using the sequential CNN model for the proposed work.

1.1 Sequential Model

A "sequential model" is just a linear structure for layers in a neural network; however, I will load pre-trained weights into its layers within a sequential architecture. The output of one layer is used as the input for the subsequent layer in sequential models, a kind of neural network architecture where layers are stacked linearly. The trainable parameters of a neural network that dictate how the model processes and interprets picture data are referred to as model weights in image processing. To reduce the loss and raise the accuracy of the model, these weights are changed throughout training. Multiple layers make up a sequential model, which processes input data incrementally. The layers are stacked in a linear fashion, with each layer's output serving as its input. The main levels of a sequential model, particularly in image processing applications, are listed below. Table 1 is showing how a complex emotion is generated by the combination of basic emotions.

Table 1: Emotions Mapping

Basic Emotions	Complex Emotion	Physical Description of Complex Emotion
Happy+Neutral	Love	Soft eyes, slightly raised eyebrows, a warm, genuine smile
Happy+Surprise	Delight	Wide smile, eyes crinkling at the corners, slightly raised eyebrows
Neutral+Sad	Hope	Eyes gazing upward or forward, slightly parted lips, relaxed features
Disgust and Sad	Remorse	Downcast eyes, furrowed brow, frown, tense jaw.
Disgust+Angry	Contempt	One side of the mouth raised (sneer), narrowed eyes, and a slight head tilt.
Angry+Sad	Frustration	Furrowed brow, lips pressed together or open in exasperation, and flared nostrils.
Surprise+Sad	Shock	Wide-open eyes, raised eyebrows, mouth agape.
Surprise+Fear	Awe	Eyes widened, slightly parted lips, relaxed or lifted eyebrows
Surprise+Sadness	Disapproval	Furrowed brow, lips pursed or turned downward, slight frown.
Surprise + Joy	Amazement	Eyes wide open, raised eyebrows, mouth slightly open in a smile.

2 Methodology

Emotion identification from photos is a crucial task for comprehending user mood and interaction because of the growth of visual content brought about by the growth of social media platforms [4],[9],[10]. Even though fundamental emotion recognition has been the subject of several research, it is still quite difficult to identify complicated emotions like irony, mixed emotions, or ambiguous emotional states. In order to identify complex emotions in social media photographs, this research suggests a convolutional neural network (CNN)-based deep learning framework. Developing a CNN-based deep learning framework to identify subtle emotions in social media photographs is the aim of this project. Our method classifies emotions that fall into several categories (e.g., joy, sadness, surprise, and their combinations) by using a sequential CNN model that has already been trained and refined using a specific dataset of annotated social media photographs. To enhance the model's generalization across various visual material categories, we use data augmentation techniques. According to experimental data, our Sequential CNN-based approach performs more accurately than conventional machine learning models, especially when detecting ambiguous or insignificant emotional cues. The data sets serve as the foundation for this section. The information has been gathered from Kaggle-hub for this section. A simple API for accessing Kaggle resources from anywhere is provided via the KaggleHub Python module. For training and testing purposes about 80k frontal face photos have been dowloaded. There are two main stages to the research project. Using a CNN architecture, a rudimentary emotion detection model was created in the first phase to categorize seven universal emotions: anger, dis-gust, fear, happiness, neutrality, sadness, and surprise. Standard measures, such as accuracy, confusion matrix, and classification re-port, were used to train, validate, and assess the model. The trained basic model served as the basis for identifying complex emotions in the second phase. In order to identify higher-order emotional states, this required expanding the architecture or mapping logic and utilizing the acquired properties from the basic model. The progress in emotion understanding from basic to complicated levels was highlighted by comparing and analyzing the final results from both models

2.1 Basic Emotion Detection Model

To categorize images into six basic emotional categories—happiness, sadness, anger, fear, disgust, and surprise—the basic emotion model was created. A convolutional neural network (CNN) with a sequential design was used to train it. For every class, evaluation metrics like F1-score, accuracy, and recall were calculated.

2.1.1 Model Performance

The performance of the suggested convolutional neural network (CNN) model for the fundamental categorization of emotions from facial photos is shown in this section. Seven emotion categories—angry, disgusted, fearful, pleased, neutral, sad, and surprised—were used to train and assess the model. Standard classification criteria, such as accuracy, precision, recall, and F1-score, were used to evaluate the model's performance. Furthermore, the distribution of correctly and erroneously identified samples was visualized using a confusion matrix. A thorough assessment of the model's accuracy and capacity for generalization on the test dataset is given in the next subsections

2.1.2 Model Training Details

A convolutional neural network (CNN) architecture was used to train the emotion detection model across 100 epochs with a batch size of 128. Using the Adam optimizer and category cross-entropy as the loss function, the model was trained on pre-processed grayscale facial photographs that had been scaled to 48 by 48 pixels. By using both training and validation datasets during the learning phase, the model was able to prevent overfitting and improve generalization. The model's convergence behavior was assessed by tracking the training process over all 100 epochs and recording the accuracy and loss values at each one. The efficacy of the architecture and hyper parameter selections was confirmed by the model's steady improvement in training and validation accuracy.

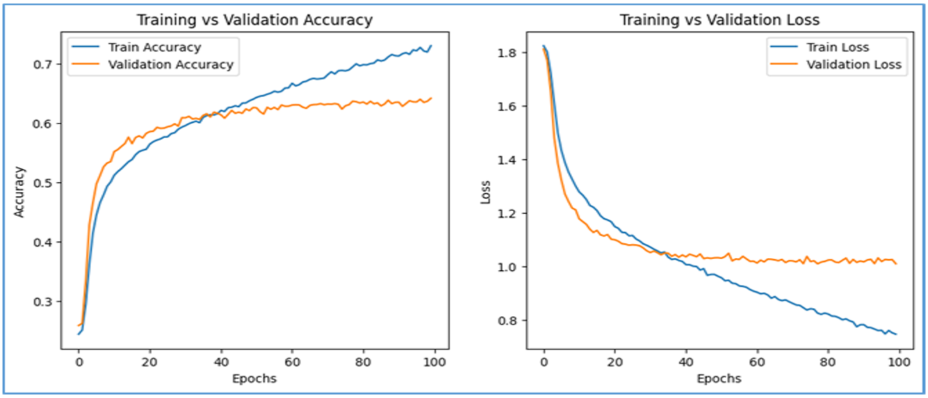

Fig. 2 Basic Emotion Detection Model Training vs Validation Accuracy and Loss

Fig. 2 is showing about the accuracy and loss of basic emotion detection model training vs validation over epochs. A quantitative assessment of the emotion detection model is given in this section. For every emotion category, key performance metrics like recall, accuracy, precision, and F1-score are presented. These measures aid in evaluating how well the model recognizes complicated emotions from visual inputs. The performance of two models created for social media image emotion identification is assessed and contrasted in this section. A baseline sequence neural network that has been trained to identify fundamental emotions makes up the first model. By identifying complex emotions—which can involve overlapping emotional states or context-dependent interpretations—the second model expands on this. To evaluate their performance and comprehend the difficulties of complicated emotion recognition, a variety of evaluation criteria are employed.

2.1.3 Model Performance

To assess how well the trained CNN model classified basic emotions from grayscale facial photos, a confusion matrix was created. It helps determine which emotions are most commonly misclassified by giving a visual depiction of the actual versus expected emotion classifications. While off-diagonal entries suggest misclassifications, the diagonal of the matrix clearly displays the right predictions. The model demonstrated some misunderstanding between sad and neutral, which are visually similar in facial expressions, but it did well in identifying emotions like surprise and happiness.

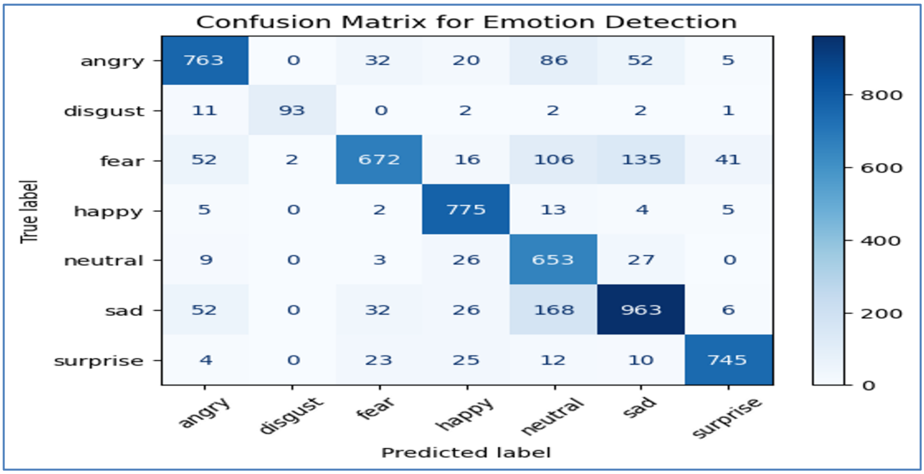

Fig.3 Confusion Matrix

Fig. 3 is showing the confusion matrix of the suggested CNN model, which displays the classification performance across seven emotion classes, is shown in Fig. 3. While off-diagonal values show incorrect classifications among related emotions like fear and sadness, diagonal elements show accurate predictions.

2.1.4 Classification Report of Basic Emotion Detection Model

Precision, recall, and F1-score are used in the classification report to assess the trained emotion detection model's performance across seven emotion categories. With an overall accuracy of 82%, the model demonstrated a high degree of prediction accuracy. Despite its relatively low support, disgust had the highest precision (0.98) among the individual emotions, indicating that nearly all of the cases predicted as disgust were in fact right. With F1-scores of 0.91 and 0.92, respectively, happy and surprise both demonstrated excellent performance, demonstrating the model's ability to identify positively expressive emotions. The model appears to over predict neutrality, potentially misclassifying more subtle expressions as neutral. This is indicated by the neutral emotion's high recall (0.91) and comparatively low precision (0.63). Fear's lowest recall (0.66) indicates that the model has trouble correctly identifying this emotion, most likely because it visually resembles emotions of grief or anger. The model's constant performance across both balanced and imbalanced emotion classes is confirmed by the weighted-average F1-score of 0.82 and the macro-average F1-score of 0.83. The model's overall robustness is supported by these results, but they also point out areas that need development, especially in terms of differentiating between closely related negative emotions. The model's overall accuracy was 82%, and it had high F1-scores for emotions including surprise, disgust, and happiness. Fear and neutral were shown to have minor misclassifications, suggesting that there is a need for improved feature separation among visually comparable emotions.

2.2 Basic Emotion Detection Model

The output of a pre-trained convolutional neural network (CNN) that categorizes common emotions like happy, sad, angry, and surprise serves as the foundation for the complex emotion recognition model. A lightweight fully connected neural network (FCNN) was created in order to expand this capacity and infer more complex or higher-order emotional states. The model transfers ten predetermined complex emotions, such as love, delight, frustration, and awe, to a seven-dimensional vector of base emotion probabilities. Through supervised training with devised data, the model discovers the patterns that link each complicated emotion to a pair of simple emotions. Using a multi-label classification method, the finished model can identify several complicated emotional states at once.

2.2.1 Motivation and Design

The fundamental seven (angry, disgusted, fearful, pleased, neutral, sad, and surprised) are not the only ways to understand human emotions. More complicated or subtle emotions, like wonder, hope, or regret, which are mixtures of fundamental emotions, are frequently ex-pressed in the real world. As a result, identifying complex emotions ofers a more expressive and realistic comprehension of affective states, particularly in applications such as social media analysis, human-computer interaction, and mental health monitoring. Complex emotions were conceptualized as mixtures of simple emotions in order to ac-accomplish this. For instance, delight was described as a combination of surprise and happiness, whereas frustration was described as a combination of melancholy and rage. By using this mapping, we were able to construct a second-level model, known as a stacked architecture, in which the basic emotion model's predictions were used as input features for a different, more sophisticated emotion classifier. This design extends the underlying model's emotional granularity in a modular and interpretable manner while utilizing its strengths.

2.2.2 Model Performance and Training

In the second phase, a fully connected neural network was used to construct a sophisticated emotion detection model. It mapped ten predefined complex emotions, such as shock, delight, and regret, to the output probabilities of the base model. The complex model handled multi-label classification using a binary cross-entropy loss function and was trained for 50 epochs using deliberately manufactured data. The model accurately predicted complex emotional states based on the co-occurrence of base emotions and demonstrated consistent training/validation performance during evaluation. Overall, the integrated system showed great promise for practical affective computing applications by effectively demonstrating the capacity to identify both fundamental and complex emotional states from facial expressions.

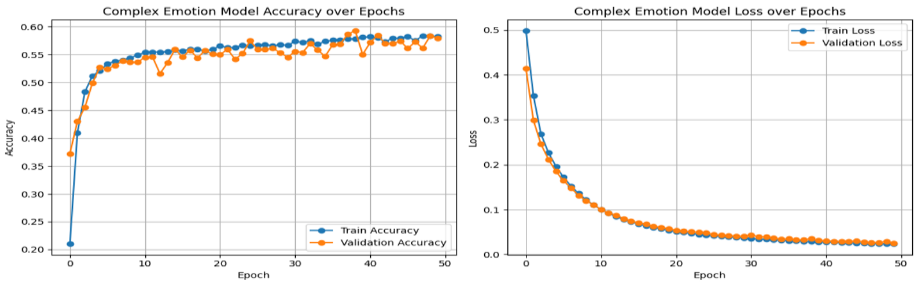

Fig. 4 Complex Emotion Model Accurecy And Loss over Epochs

Fig. 4 is showing the complicated emotion recognition model's performance was qualitatively evaluated using a limited set of 10–15 real-world facial photo-graphs. These photos were chosen by hand to capture a range of authentic emo-tional emotions. The complex emotion model was given the emotion probabilities that were obtained after each image was run through the base emotion detector. Each test case contains the input image, the identified complex emotions, and the predicted base emotions (with probability distribution) for display. Explana-tions explaining the contributions of particular base emotion combinations to each complex emotion prediction were also produced in order to improve inter-pretability. This made it easier to confirm that the model and the planned emotion mappings make sense.

3 Research Result & Discussion

The findings of the complex emotion recognition model applied to social media photos are shown and explained in this chapter. Several metrics are used to assess the model's performance, and the results are presented in light of the goals of the study and the body of current literature. To examine the model's advantages and disadvantages, both quantitative and qualitative studies are offered.

3.1 Sample Predictions

The trained CNN-based emotion recognition model was applied to sample test photos in order to verify the model's practicality. The algorithm was highly confident in its ability to predict a wide range of emotions, including surprise, anger, sadness, and happiness. Visual examples show how well the model recognizes the aspects of facial expressions and labels them with the proper emotion. The majority of the time, the expected label and the projected emotion matched, demonstrating how well the trained network handled unseen input. The findings in the classification report and confusion matrix are further supported by these findings.

3.1.1 Basic Emotion Prediction Sample

The following figure shows the result of a randomly selected image from sample data, and the result is the following: Angry

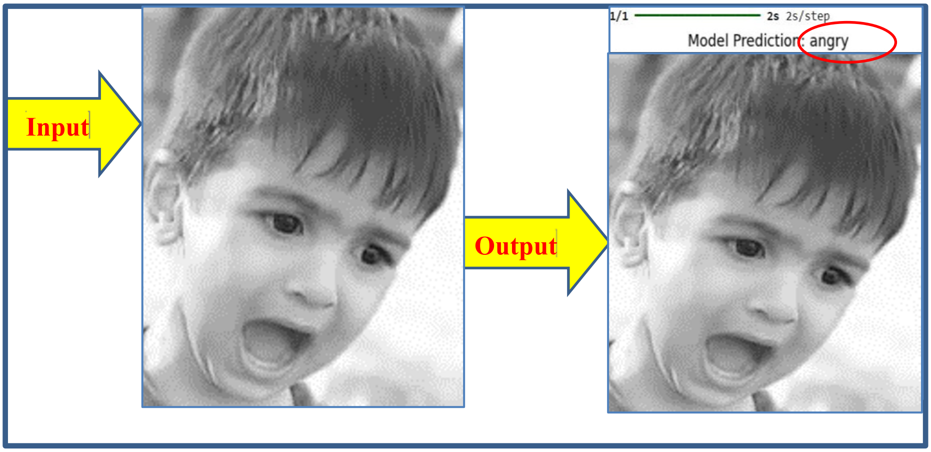

Fig.5 Basic Emotion Prediction

Fig. 5 shows the result of randomly selected image from sample data and the re-sult is: Angry.

3.1.2 Complex Emotion Prediction S ample

Following Fig. 6 shows the result of randomly selected image from sample data and the result is: **Hope** (according to % of neutral and sad)

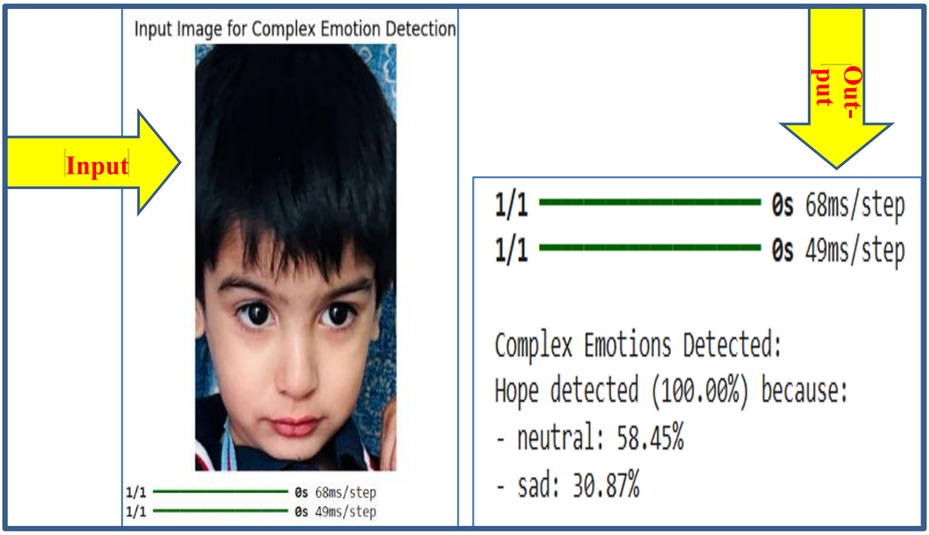

Fig. 6 Complex Emotion Prediction

3.2 Comparative Discussion and Limitation

The behavior and performance of the basic and complex emotion detection models are contrasted in this section. The base model does well on basic emotions like happy, sad, furious, etc. because it was trained on standard facial emotion datasets (such as Kaggle-Hub). However, the necessity for a complex emotion model was prompted by the fact that human emotional states are frequently complicated and overlapping. Combinations of base emotions are used to anticipate complicated emotions (e.g., hope = neutral + sad, regret = sad + disgust). The lack of a specific complex emotion dataset for training and assessment limits this stacked model, even though it provides a superior interpretation of nuanced emotions. Thus, probabilistic mappings were used to synthetically train the model. This aids in proving viability, but it has an impact on generalization, particularly when evaluated on actual photos. Moreover, thorough validation is constrained by a small test set (10–15 genuine photos). This conversation emphasizes the need for more robust generalization methods, better emotional annotation standards, and richer datasets in subsequent research. The lack of publicly accessible datasets explicitly labelled for complex emotions like awe, regret, irritation, or hope is one of the main obstacles this research has faced. Complex emotions are still under-represented in academic literature and data resources, in contrast to fundamental emotions, which have been extensively researched and are accessible in datasets such as FER-2013. Only 100 to 150 real-world photos were thus personally gathered from social media sites and annotated according to personal opinion. The model's capacity to generalize across a range of facial expressions and lighting conditions is impacted by this small and unbalanced dataset. A synthetic dataset was created to address this issue by combining base emotion probabilities; nevertheless, this does not take the place of actual annotated data.

4 Conclusion

In this work, we developed an efficient method for emotion recognition from facial photos using convolutional neural networks (CNNs). The suggested model uses a deep learning-based categorization framework to identify basic emotions, including anger, happiness, sadness, neutrality, fear, disgust, and surprise. The testing results show that the model achieves an overall accuracy of 82%, demonstrating its ability to reliably classify emotions and extract pertinent facial features. Due to slight differences and overlapping facial expressions, the model performs relatively poorly for emotions like fear and neutral; however, it performs well for distinct emotions like delight and surprise, according to the classification performance. These results demonstrate the strengths and weaknesses of the suggested method in addressing practical emotion recognition problems. Additionally, CNN-based learning offers a solid basis for identifying face patterns and can be expanded to address more intricate emotional states. With potential uses in affective computing, mental health analysis, and human–computer interaction, this research advances the creation of emotion recognition systems. To improve generalization and real-time performance, the model can be upgraded in subsequent work by adding sophisticated architectures, better datasets, and optimization strategies.

Acknowledgements. The authors would like to express their gratitude to their institution for providing the necessary support and resources to carry out this research work. No external funding was received for this study.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] A. Joseph and P. Geetha, "Facial emotion detection using a modified EyeMap–MouthMap algorithm on an enhanced image and classification with TensorFlow," *The Visual Computer*, vol. 36, no. 3, pp. 529–539, Mar. 2019.

[2] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, no. 1, Art. no. 53, 2021.

[3] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," in *Proc. 2018 1st Int. Conf. Secure Cyber Comput. Commun. (ICSCCC)*, 2018, pp. 278–282.

[4] J. Nyanjom and D. Naylor, "Performing emotional labour while teaching online," *Educational Research*, vol. 62, no. 2, pp. 147–163, 2020.

[5] C.-C. Carbon, "Wearing face masks strongly confuses counterparts in reading emotions," *Frontiers in Psychology*, vol. 11, Art. no. 589497, 2020.

[6] Z. Wang, S.-B. Ho, and E. Cambria, "A review of emotion sensing: categorization models and algorithms," *Multimedia Tools and Applications*, vol. 79, no. 47–48, pp. 35553–35582, 2020.

[7] R. Berrios, "What is complex/emotional about emotional complexity?" *Frontiers in Psychology*, vol. 10, Art. no. 1606, 2019.

[8] R. Schmøkel and M. Bossetta, "FBAdLibrarian and Pykognition: open science tools for the collection and emotion detection of images in Facebook political ads with computer vision," *Journal of Information Technology & Politics*, vol. 19, no. 1, pp. 118–128, 2021.

[9] K. Yang et al., "Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets," *The Visual Computer*, vol. 37, no. 6, pp. 1447–1466, 2021.

[10] D. Reney and N. Tripathi, "An efficient method to face and emotion detection," in *Proc. 2015 5th Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, 2015, pp. 493–497.