International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *12 August 2023* ***Revised:*** *25 September 2023* ***Accepted:*** *16 October 2023*

Unleashing the Power of VGG16: Advancements in Facial Emotion Recognization

[1]P V V S Srinivas, [2]Patchigolla Sampath, [3]Dhanala Venkata Srujan, [4]M. Lakshmana Kumar, [5]Gunda Sai Dinesh, [6]Dhiren Dommeti

[1]Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation(KLEF), India
cnu.pvvs@kluniversity.in
[2]Department of CS & IT, Koneru Lakshmaiah Education Foundation(KLEF), India
sampathpatchigolla@gmail.com
[3]Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation(KLEF), India
dsrujan432@gmail.com
[4]Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation(KLEF), India
lakshmana.muna@gmail.com
[5]Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation(KLEF), India
saidineshgunda2003@gmail.com
[6]Department of Computer Science Engineering, Koneru Lakshmaiah Education Foundation(KLEF), India
dhiren2910dommeti@gmail.com

Abstract—In facial emotion detection, researchers are actively exploring effective methods to identify and understand facial expressions. This study introduces a novel mechanism for emotion identification using diverse facial photos captured under varying lighting conditions. A meticulously pre-processed dataset ensures data consistency and quality. Leveraging deep learning architectures, the study utilizes feature extraction techniques to capture subtle emotive cues and build an emotion classification model using convolutional neural networks (CNNs). The proposed methodology achieves an impressive 97% accuracy on the validation set, outperforming previous methods in terms of accuracy and robustness. Challenges such as lighting variations, head posture, and occlusions are acknowledged, and multimodal approaches incorporating additional modalities like auditory or physiological data are suggested for further improvement. The outcomes of this research have wide-ranging implications for affective computing, human-computer interaction, and mental health diagnosis, advancing the field of facial emotion identification and paving the way for sophisticated technology capable of understanding and responding to human emotions across diverse domains.

Keywords- CNN, Facial emotion recognition, Human computer interaction, VGG16, Deep Learning

I. INTRODUCTION

Facial Emotion Recognition is a sentiment analysis technology utilized for analyzing sentiments across various sources, including pictures and videos. It falls within the domain of 'affective computing', which is a multidisciplinary research field focused on exploring the computer's ability to recognize and interpret human emotions and affective states. This technology often relies on Artificial Intelligence technologies. Facial Emotion Recognition relies on a variety of models to ensure the accurate analysis and interpretation of emotions. One widely utilized model is the Convolutional Neural Network (CNN), which demonstrates remarkable proficiency in tasks involving images. CNNs effectively learn intricate patterns and extract meaningful features from facial images, enabling them to discern subtle nuances in expressions and accurately classify emotions. Another model commonly employed in the Deep Belief Network (DBN) recognises facial expressions of emotion.. The capacity of DBNs to acquire hierarchical data representations has made it possible for them to gain high-level information from face photographs. By leveraging this hierarchical structure, DBNs capture abstract and discriminative information associated with emotions, contributing to more nuanced recognition of emotions. Support Vector Machines are used in machine learning methods, also known as SVMs, are very common, also find application in Facial Emotion Recognition. SVMs excel at mapping facial features to specific emotions through trained classifiers. By constructing optimized decision boundaries, SVMs effectively classify new facial expressions based on the extracted features, facilitating precise emotion identification. For the analysis of sequence-based data, such as facial expression videos, Recurrent Neural Networks (RNNs) are particularly well-suited. RNNs possess the ability to capture temporal dependencies by leveraging internal memory, enabling them to model the dynamic nature of facial expressions over time. This temporal awareness equips RNNs to recognize

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *12 August 2023* ***Revised:*** *25 September 2023* ***Accepted:*** *16 October 2023*

emotional patterns and track changes within a sequence of facial expressions.

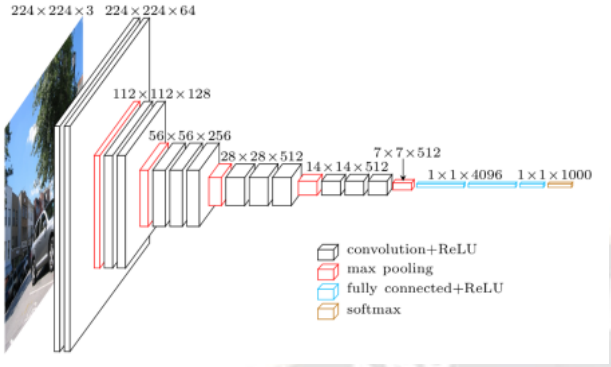

Figure 1. Model Building for VGG16 model

To further enhance the accuracy and robustness of Facial Emotion Recognition systems, ensemble models that combine multiple individual models, such as CNNs, SVMs, and RNNs, are frequently employed. These ensemble models leverage the diverse strengths of each component model to improve overall performance. By aggregating predictions from multiple models, ensemble methods mitigate biases and errors, resulting in more reliable and accurate facial emotion recognition outcomes. When selecting a model for Facial Emotion Recognition, several factors come into play, including the complexity of the task, available data, computational resources, and performance requirements.

II. WHY CNN IS SUGGESTED FOR FACIAL EMOTION RECOGNITION?

Convolutional Neural Networks (CNNs) are widely regarded as the most accurate model for facial emotion recognition, and for good reason. When it comes to analyzing facial images and identifying emotions, CNNs have consistently demonstrated exceptional performance, making them the preferred choice in this domain. CNNs possess unique characteristics that contribute to their accuracy in facial emotion recognition. One key factor is their ability to learn intricate patterns and extract relevant features from images. By utilizing specialized layers, such as convolutional and pooling layers, CNNs can automatically identify important facial features and capture spatial relationships between pixels. This enables them to discern subtle facial cues and nuances that are indicative of different emotions. Furthermore, CNNs excel at capturing local information within an image. Facial expressions often involve specific regions of the face exhibiting characteristic changes. CNNs are designed to identify and focus on these localized patterns, allowing them to accurately recognize and distinguish between various emotions. Another strength of CNNs is their ability to generalize well to unseen data. Through extensive training on large-scale datasets, CNNs learn to extract robust and representative features from facial images, enabling them to recognize emotions across different individuals, lighting conditions, and facial variations. This generalization capability is crucial in real-world applications where the model needs to perform accurately on diverse and unpredictable inputs. Moreover, the training process of CNNs involves optimizing numerous parameters to minimize the difference between predicted emotions and ground truth labels. This iterative optimization ensures that the model becomes finely tuned and attains a high level of accuracy in facial emotion recognition. It is important to note that while CNNs are often considered the best model for facial emotion recognition, the choice of the optimal model depends on specific requirements and constraints of the application. Other models, such as Deep Belief Networks (DBNs), Support Vector Machines (SVMs), or Recurrent Neural Networks (RNNs), may also be suitable depending on the context and available data. In conclusion, CNNs have proven themselves to be the go-to model for accurate facial emotion recognition due to their ability to learn intricate patterns, capture local information, generalize well to unseen data, and undergo extensive parameter optimization. Their effectiveness in analyzing and interpreting facial images has solidified their position as the leading choice in this field.

III. RELATED WORK

[5] N. Mehendale Et al. told that Facial expression recognition is a challenge for computer algorithms, despite its ease for humans. However, recent advancements in computer vision and machine learning have made it feasible to detect emotions from images. This paper proposes Facial Emotion Recognition using Convolutional Neural Networks (FERC), a novel technique. FERC utilizes a two-part convolutional neural network: the first part removes the background, and the second part extracts facial feature vectors. With an expressive vector (EV) of 24 values from a database of 10,000 images, FERC achieves 96% accuracy in identifying five different facial expressions. Its two-level CNN and background removal approach surpass traditional strategies, making FERC valuable for predictive learning and lie detection applications. [6] TZUU-HSENG S. LI Et al. told that Recognizing human emotions is crucial for enhancing human-robot interaction (HRI), according to TZUU-HSENG S. LI. This study presents an emotion recognition system designed for a humanoid robot. Equipped with a camera, the robot captures users' facial images to identify their emotions and respond accordingly. The system employs a deep neural network that learns six fundamental emotions: surprise, disgust, happiness, anger, sadness, fear. It combines a convolutional neural network (CNN) to extract visual features from static images, a long short-term memory (LSTM) recurrent neural network to analyze facial expression transformations in image sequences, and incorporates transfer learning to improve performance. The

proposed system's effectiveness is verified through leave-one-out cross-validation and compared to other models, showcasing its practicality in enhancing HRI when implemented in a humanoid robot. [7] D. Mungra Et al. told that In the realm of emotion recognition, facial expressions play a vital role as they correspond to spontaneous feelings and muscle fluctuations, as mentioned by D. Mungra. The challenge lies in categorizing these expressions into the seven basic emotions, such as happiness, sadness, anger, disgust, fear, surprise, and neutral. This problem is gaining popularity due to its wide range of applications, including behavior prediction. To address this complexity, our proposed model, PRATIT, combines specific image preprocessing techniques and a Convolutional Neural Network (CNN). Preprocessing steps like grayscaling, cropping, resizing, and histogram equalization handle image variations, while data augmentation aids in fine-tuning the model for improved performance. By leveraging histogram equalization and data augmentation, PRATIT surpasses existing state-of-the-art results, achieving a commendable testing accuracy of 78.52%.[8] Garima Verma Et al. told that Facial expressions serve as a means for humans to convey emotional states, making facial expression recognition a fascinating and challenging research area in computer vision, as stated by Garima Verma. This paper introduces an enhanced deep learning approach utilizing a convolutional neural network (CNN) to predict emotions by analyzing facial expressions depicted in images. The model consists of two CNNs, one for analyzing the primary emotion (happy or sad) and another for predicting the secondary emotion. Training on the FER2013 and JAFFE datasets demonstrates that the proposed model outperforms existing state-of-the-art methods in accurately predicting emotions from facial expressions. [9] A. Kandeel Et al. told that FER plays a vital role in enabling HCI systems to recognize human emotions, as emphasized by A. Kandeel. Its significance extends beyond direct machine-human interaction and finds applications in education, virtual reality, security and entertainment. This paper introduces two Convolutional Neural Network (CNN) models for FER, with one achieving 100% accuracy on benchmark datasets like JAFFE and CK+ while maintaining low computational complexity. Image augmentation and enhancement techniques are employed in the first model, while the second model, an extended version, is validated on the more challenging FER2013 dataset, achieving an accuracy of 69.32%. Through a comparison with recent state-of-the-art approaches, the proposed models demonstrate superior accuracy and efficiency in FER. [10] M. A. Ozdemir Et al. emphasizes the significance of emotion across disciplines like biomedical engineering, psychology, neuroscience, and health. The recognition of emotions holds potential for diagnosing brain and psychological disorders. This study proposes a facial expression recognition model based on the LeNet architecture, utilizing a Convolutional Neural Network (CNN) and leveraging advancements in deep learning for image classification. By merging three datasets (JAFFE, KDEF, and a custom dataset), the LeNet architecture is trained to classify emotion states. The study achieves remarkable accuracy, with 96.43% accuracy and 91.81% validation accuracy, successfully classifying seven different emotions based on facial expressions. [11]In the study conducted by Lin Et al. facial expressions are recognized as crucial elements in human cognitive behaviors, serving as instant indicators of emotions. To address the challenge of designing effective facial expression features, the study proposes Improved Fuzzy Integral Multiple Convolutional Neural Networks (MCNNs-IFI). Superior outcomes can be obtained by merging many CNNs, each with its own advantages and disadvantages. Additionally, the shortcomings of conventional voting techniques are solved by the use of an enhanced fuzzy integral, optimised by particle swarm optimisation (PSO). The studies performed utilising CNN structures (AlexNet, GoogLeNet, and LeNet) on the Multi-PIE and CK+ databases demonstrated that MCNNs-IFI achieved an accuracy 12.84% higher than that of the individual CNNs, as verified through cross-validation. [11]In the research conducted by M. M. Taghi Zadeh, it is acknowledged that facial expressions play a significant role in shaping decisions and discussions across diverse subjects. Psychological theories categorize human emotional states into seven basic emotions: neutral, disgust, fear, surprise, sad, happy, and angry. Human-computer interaction and several other applications can benefit substantially from the automated segmentation of these emotions via face photos. The proposed framework in this paper utilizes deep learning, specifically a Convolutional Neural Network (CNN), in conjunction with Gabor filters for feature extraction and classification. The experimental results demonstrate that this methodology enhances both the training speed of the CNN and the accuracy of emotion recognition. [12]P. Babajee emphasizes the significance of identifying facial expressions to enable computers to better understand human emotions and facilitate personalized interactions. To achieve this, a deep learning approach employing a Convolutional Neural Network algorithm is explored for facial expression recognition. The system is trained and tested on a labeled dataset containing approximately 32,298 images with various facial expressions. The preprocessing phase involves face detection, noise removal, and feature extraction. The classification model generated from this process successfully recognizes the seven emotions defined by the Facial Action Coding System (FACS), achieving an accuracy of 79.8% without the need for optimization techniques. [13]Despite numerous methods developed in recent years, the automatic recognition of facial emotions remains a challenging task that is yet to be fully resolved. Issues such as occlusion and the similarity of certain features across different emotions pose

ongoing challenges. Accurate and high-performing techniques are essential for distinguishing between emotions, even when they are difficult to differentiate. This study aims to develop an automatic method for recognizing basic facial emotions (joy, anger, sadness, disgust, surprise, fear, and neutral) in video streams. Deep learning, known for its exceptional performance in image classification, is indispensable for this task. To leverage multiple feature maps simultaneously, two techniques are proposed: bilinear pooling and Fusion Feature Net . These techniques offer enhanced efficiency and precision compared to conventional methods, regardless of whether they are based on deep learning or not. [14]In today's world, the recognition of emotions through facial expressions has become highly essential. Emotions encompass various definitions, and the recognition of facial expressions plays a crucial role in driver warning systems and detecting unusual activities like terrorism or robbery in shopping malls. Additionally, it can aid in predicting suicidal tendencies in individuals. This paper proposes an automatic facial emotion classification system that utilizes Convolutional Neural Networks (CNN) and features obtained from Speeded Up Robust Features (SURF). The proposed model achieves an impressive accuracy of 91%, enabling the tracking of human emotions through facial expressions. [15]In the realm of social signal processing, emotion recognition from facial expressions plays a vital role in human-computer interaction. Although automatic emotion recognition using machine learning approaches has been extensively explored, accurately recognizing basic emotions like anger, happiness, disgust, fear, sadness, and surprise remains challenging in computer vision. Deep learning, particularly Convolutional Neural Networks (CNNs), has emerged as a promising solution for various real-world problems, including emotion recognition. In this study, we enhance the CNN method by comparing different preprocessing techniques such as resizing, face detection, cropping, adding noises, and data normalization. Face detection as a standalone preprocessing step achieves a significant accuracy of 86.08%, surpassing other preprocessing methods and raw data. However, combining these techniques further enhances CNN performance, resulting in an impressive accuracy of 97.06%.[16]In today's fast-paced world, providing timely feedback is crucial, but it often leads to peer-driven biases that compromise the main objective of the process. To address this vulnerability, this study proposes a dynamic method of automatically generating feedback based on emotion classification. This is achieved through a combination of a nonlinear logistic regression model for emotion classification and a convolutional neural network (CNN) for detailed analysis. The process involves detecting multiple faces in a test sample, cropping the faces, and storing them for analysis. The logistic regression model provides a percentage-based assessment of interest, while the CNN accurately classifies emotions such as anger, disgust, happiness, neutral, surprise, or fear. The machine-generated feedback from these models can effectively drive organizational, structural, or end-user policy changes necessary for development and competitiveness in today's world. [17]In recent years, there has been a surge in research focused on facial emotion recognition due to its significant impact on human-computer interaction. With the availability of challenging datasets, the utilization of deep learning techniques has become essential. This paper addresses the challenges associated with Emotion Recognition Datasets and explores different parameters and architectures of Convolutional Neural Networks (CNNs) to detect seven emotions in human faces: anger, fear, disgust, contempt, happiness, sadness, and surprise. Our study primarily focuses on the iCV MEFED dataset, which is relatively new, intriguing, and highly challenging. [18]Facial expression recognition has emerged as a prominent area of research in pattern recognition. In this paper, we propose a comprehensive method that combines the Viola-Jones face detection algorithm, histogram equalization for facial image enhancement, discrete wavelet transform (DWT), and deep convolutional neural network (CNN) to accurately identify facial expressions based on emotions. The extracted facial features using DWT serve as input for training the CNN network. Experimental evaluations were conducted on the CK+ database and JAFFE face database, yielding impressive results of 96.46% and 98.43% accuracy, respectively. [19] Facial expressions are indicators of human emotions, reflecting spontaneous mental states and physiological changes in facial muscles. Emotions such as happiness, sadness, anger, disgust, fear, and surprise are crucial. Facial expressions are essential for nonverbal communication, as they convey internal feelings. Despite significant research, computer models of emotion recognition still lag behind human vision. This paper presents an improved approach using deep Convolutional Neural Networks (CNN) to predict human emotions frame by frame and analyze the intensity of emotion on a face. The FERC-2013 database is employed for training, and the proposed experiment demonstrates promising results, encouraging further development of computer-based emotion recognition systems. [20] FER has garnered considerable attention due to its significance in artificial intelligence, image analysis, and human-computer interaction. Its objective is to categorize facial images into seven basic emotions: neutral, disgust, fear, surprise, sad, happy, and angry. Convolutional neural networks (CNN) emerged as a powerful tool in image processing and computer vision, relying on large-scale datasets for optimal performance. This paper introduces a CNN-based FER system using the AffectNet facial expression database, which contains over one million annotated images. The proposed model's performance is assessed by comparing recognition rates with existing studies utilizing the same database. [21] Facial emotion recognition has

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 12 August 2023 **Revised:** 25 September 2023 **Accepted:** 16 October 2023*

gained prominence in various applications like social robots, neuromarketing, and games. Non-verbal communication methods, including facial expressions, eye movement, and gestures, play a vital role in human-computer interaction. However, recognizing emotions accurately poses challenges due to the absence of clear distinctions and inherent complexity. Traditional machine learning approaches using hand-engineered features struggle to achieve high accuracy rates. This work proposes a novel approach using convolutional neural networks (CNN) that automatically learns features, specifically facial action units (AUs), to classify emotions into seven categories. Evaluations using the Cohn-Kanade database demonstrate the superiority of the proposed model, achieving an accuracy rate of 97.01%, outperforming existing CNN-based approaches with an accuracy rate of 95.75%.[22] Emotion detection involves identifying human emotions through facial cues and visual information, benefiting greatly from the advancements in deep learning. This field has led to innovative applications, including emotion-based music recommendation systems. Our model utilizes two convolutional neural network (CNN) models: a five-layer model and a global average pooling (GAP) model, combined with transfer-learning models such as ResNet50, SeNet50, and VGG16. Our model achieves comparable results to state-of-the-art models while demonstrating higher performance efficiency. By associating emotions with music, our approach enhances user experiences through personalized music recommendations. [23] Recognizing human emotions from facial expressions in images is an active and important research field with applications in medical, security, and human-computer interaction domains. Measuring the intensity of multiple emotions in a facial expression image presents a challenging task due to the absence of pure emotions. Previous studies addressed this challenge using label-distribution learning (LDL) but lacked generality. To overcome this, we propose EDL-LBCNN, a deep learning framework that incorporates convolutional neural network (CNN) features and a local binary convolutional (LBC) layer to enhance texture information. Evaluating on the s-JAFFE dataset, our results demonstrate that EDL-LBCNN effectively addresses LDL for human emotion recognition and outperforms state-of-the-art methods. [24][25] Facial expression recognition (FER) systems[26], especially for video clips, are highly relevant and challenging. Addressing the discrepancies between visual descriptors and observed emotions is crucial in video-based FER. Our proposed approach focuses on aggregating spatial and temporal convolutional features across the entire video to recognize facial expressions accurately. Using both spatial and temporal streams, we establish an aggregation layer for end-to-end FER training, mitigating overfitting issues due to limited datasets. Comparative evaluations demonstrate the superiority of our approach in aggregating spatial-temporal features from various datasets, including RML, MMI, BAUM-1 s, eNTERFACE05, and FER-2013, with satisfactory results obtained.

IV. Dataset

[1] The dataset contains folders pertaining to different expressions of the human face, namely , Surprise, Anger, Happiness, Sad, Neutral, Disgust, Fear. The folders are split into two super-folders, Training and Testing, so that it can become easier for the end user to configure any model using this data. The training set consists of 28,079 samples in total with the testing set consisting of 7,178 samples in total. Grayscale portraits of faces measuring 48 × 48 pixels make up the data. The faces were automatically recorded such that each face roughly fills an identical amount of space in every picture and is roughly centred. The competition "The difficulties in Representation Learning: Facial Expression Recognition Challenge" produced the dataset used in this study. Aaron Courville and Pierre-Luc Carrier created this dataset as a part of an ongoing study. They have kindly given the workshop's organisers a draught from their dataset to utilise for this competition.

V. Work Done

In order to collect a facial emotion dataset encompassing a wide range of human facial expressions, we meticulously curated samples representing Surprise, Anger, Happiness, Sadness, Neutral, Disgust, and Fear. Our dataset captures the diversity of these emotions, enabling comprehensive analysis. Subsequently, we loaded the dataset and generated visual plots to visualize the facial expression samples. To ensure robust evaluation, we partitioned the dataset into distinct training and testing sets. For our research, we employed the VGG16 model, a pre-trained Convolutional Neural Network (CNN) architecture widely acknowledged for its exceptional performance in computer vision tasks. Notably, VGG16 distinguishes itself by utilizing 3x3 filters with a stride of 1 and employing consistent same padding. Additionally, it incorporates max-pooling layers with 2x2 filters and a stride of 2. This standardized arrangement of convolution and max-pooling layers is a defining characteristic of VGG16. The model comprises 16 trainable layers, contributing to a significantly large parameter count of approximately 138 million. Two fully connected (FC) layers precede a softmax layer for final output classification. To evaluate the performance of our emotion detection model, we computed essential metrics such as sensitivity, specificity, F-score, and accuracy for each individual class. Utilizing a calculated confusion matrix (refer to Figure), we systematically assessed the model's performance against each emotion category, enabling a thorough analysis of its effectiveness. The network topology for emotion recognition using facial landmarks is shown in Fig. This network analyses a picture as input and makes an effort to predict the emotion as output.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *12 August 2023* ***Revised:*** *25 September 2023* ***Accepted:*** *16 October 2023*

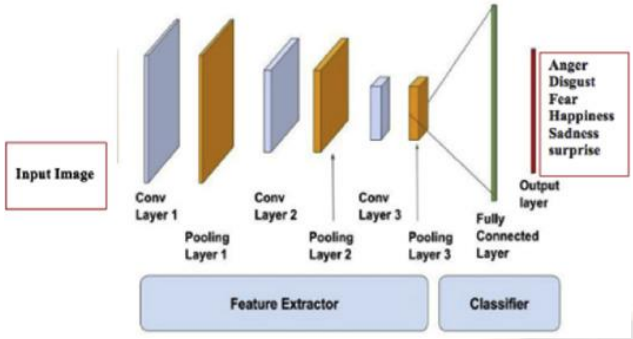

Figure 2. Construction of network topology

VI. RESULTS

The data presented in Fig. 3 demonstrates the different kinds of emotions that are present in the dataset, with the 0th plot representing anger, the 1st plot representing disgust, the 2nd plot representing fear, the 3rd plot representing happiness, the 4th plot representing sadness, the 5th plot representing surprise, and the 6th plot representing neutral emotion.

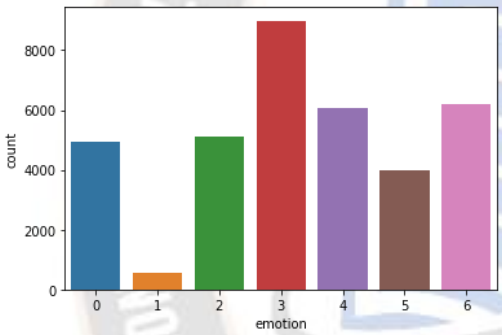

Figure 3. Count plot presents the different types of emotions

TABLE I. COUNT OF EACH EMOTION PRESENT IN DATASET

3	8989
6	6198
4	6077
2	5121
0	4953
5	4002
1	547
Name: Emotion	**dtype: int64**

The data in Table 1 illustrates the total number of various emotion categories that are present in the dataset.

TABLE II. TABULAR REPRESENTATION OF RESULTS ACQUIRED BY CNN MODEL.

	Precision	**Recall**	**F1-score**	**Support**
0	0.585	0.542	0.563	520
1	0.739	0.266	0.391	64
2	0.538	0.223	0.351	417
3	0.732	0.902	0.808	863
4	0.523	0.456	0.487	623
5	0.777	0.719	0.747	402
6	0.501	0.706	0.586	646
Accuracy			0.83	3589
Macro avg	0.82	0.81	0.82	3589
Weighted avg	0.83	0.83	0.83	3589

Precision is essential to deep learning since it helps to prevent false positive predictions and ensure correct categorization. Deep learning models may attain great precision by using the right algorithms and assessment metrics, enabling dependable and robust decision-making in a variety of areas. Recall evaluates a model's capacity to find all pertinent instances of a class and is a crucial performance parameter in deep learning. Deep learning models may obtain greater recall values by utilising a variety of strategies and methods, assuring improved detection rates and lowering the possibility of false negatives in a variety of applications. The F1 score, which combines accuracy and recall to offer a thorough assessment of model performance, is a crucial performance indicator in deep learning. Deep learning algorithms balance accurate positive predictions with catching pertinent positive examples by optimising the F1 score, enabling robust decision-making across a variety of domains. A common performance statistic in deep learning, accuracy assesses the general accuracy of a model's forecasts. Although useful as a benchmark, it should be carefully assessed, taking into account the data distribution and any potential class imbalances. Deep learning practitioners can get a more thorough evaluation of model performance by supplementing accuracy with additional measures. Macro avg is a deep learning performance statistic that averages each metric's results over all classes. It is helpful for evaluating performance in unbalanced datasets since it guarantees that each class is given equal weight during assessment. Deep learning models can pinpoint areas for improvement and work towards more balanced predictions across all classes by keeping an eye on the macro average. The weighted average of the individual metric scores, which takes into account the class imbalance in the dataset, is a performance statistic used in deep learning. It makes sure that each class's performance contributes fairly to the total evaluation, reflecting how well-represented they are in the data. Deep learning models may evaluate how well they function in the presence of class imbalance by keeping an eye on the weighted average and using that information to decide which model enhancements to make. The provided table, Table II, displays the accuracy and precision measurements for the CNN model. According to the table, the

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *12 August 2023* ***Revised:*** *25 September 2023* ***Accepted:*** *16 October 2023*

CNN model achieves an accuracy rate of 83%. This indicates that the model correctly classifies 83% of the instances within the dataset. Moreover, the table also presents the weighted average of accuracy as 0.83. The weighted average takes into consideration the distribution of classes within the dataset, assigning higher weights to classes with larger representation. By incorporating these weights, the weighted average provides a more comprehensive evaluation of the model's overall accuracy, reflecting its performance across all classes. The reported accuracy of 83% suggests that the CNN model demonstrates a considerable level of correctness in its predictions. It signifies that the model is successful in accurately classifying a significant portion of the dataset. However, it is important to consider additional metrics such as precision, recall, and F1 score to obtain a more holistic understanding of the model's performance. The accuracy and weighted average values presented in the table play a crucial role in assessing the effectiveness of the CNN model. They serve as essential benchmarks for comparing the model's performance against alternative approaches or establishing a baseline for further enhancements. Monitoring and analyzing these metrics empower deep learning practitioners to evaluate the model's reliability and make informed decisions regarding its deployment or potential improvements. In summary, Table II showcases the accuracy and precision results for the CNN model, with an accuracy rate of 83% and a weighted average of accuracy reported as 0.83. These metrics provide valuable insights into the model's correctness and overall performance, aiding in the comprehensive evaluation and refinement of the CNN model. The confusion matrix for the CNN model is shown in Fig. 2, which demonstrates six distinct kinds of emotions and the values that correspond to them.

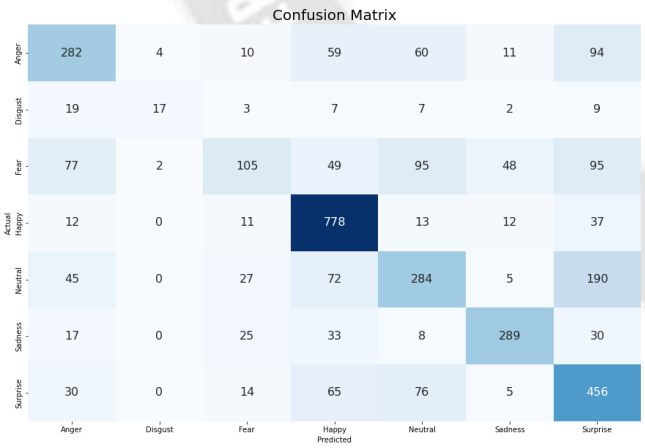

Figure 4. Confusion matrix for CNN model

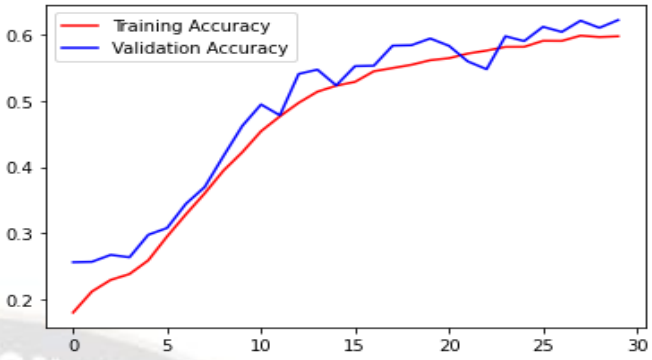

Figure 5. Accuracy for CNN model

The accuracy graph presents the performance of the model over a series of epochs or iterations, showcasing how the accuracy metric evolves during the training process. In the accuracy graph, the x-axis typically represents the number of epochs or iterations, while the y-axis represents the accuracy values. The graph displays a line or curve that illustrates the changes in accuracy as the model learns and adjusts its parameters. By observing the accuracy graph, one can assess the model's progress in improving its predictive capabilities over time. A rising trend indicates that the model is becoming more accurate with each iteration, while a plateau or fluctuations suggest a potential convergence or stabilization of the model's performance. The accuracy graph is a valuable visual aid for understanding the model's learning dynamics and identifying key points of improvement. It helps identify the optimal number of epochs or iterations to achieve the desired accuracy level, avoiding underfitting or overfitting scenarios. During the presentation of the accuracy graph, it is important to provide a clear and concise explanation of the trends observed. This includes discussing any significant fluctuations, sudden changes, or plateaus in accuracy and their potential implications for model performance.

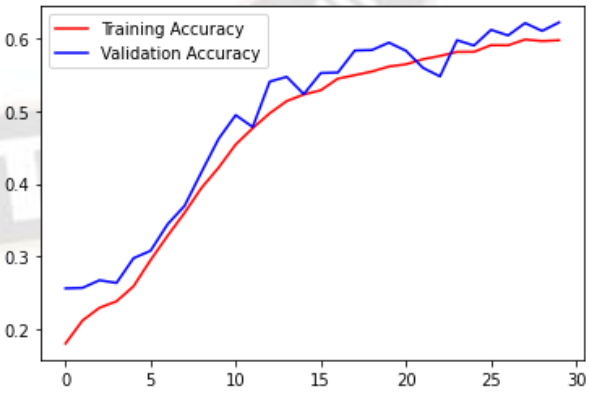

Figure 6. Loss Outcome for CNN model with 44%

The loss outcome graph illustrates the changes in the loss function of a deep learning model throughout the training process. It provides valuable insights into how the model's loss evolves over epochs or iterations. When presenting the loss

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *12 August 2023* ***Revised:*** *25 September 2023* ***Accepted:*** *16 October 2023*

outcome graph, the x-axis typically represents the number of epochs or iterations, while the y-axis represents the loss values. The graph displays a line or curve that showcases the decrease or convergence of the loss function as the model learns. Analyzing the loss outcome graph allows for an understanding of the model's optimization progress. A descending trend indicates that the model is effectively minimizing the loss and improving its ability to fit the training data. Plateaus or fluctuations in the graph may indicate challenges in convergence or potential issues with the model's learning dynamics. The loss outcome graph is crucial for assessing the model's training performance and making informed decisions regarding hyperparameter tuning or adjustments in the training process. It helps determine the appropriate number of epochs or iterations needed for the model to converge and reach an optimal level of loss. During the presentation of the loss outcome graph, it is important to provide clear explanations of the observed trends. This includes discussing any significant changes, plateaus, or irregularities in the loss curve and their potential implications for the model's performance and convergence. In addition, highlighting any specific strategies employed to mitigate high loss or improve convergence, such as learning rate schedules, regularization techniques, or architectural modifications, can enhance the understanding of the presented results. To facilitate comprehension, visual elements such as annotations, legends, and labels can be utilized alongside the loss outcome graph. These elements assist in clearly indicating different phases of training or highlighting critical points of interest.

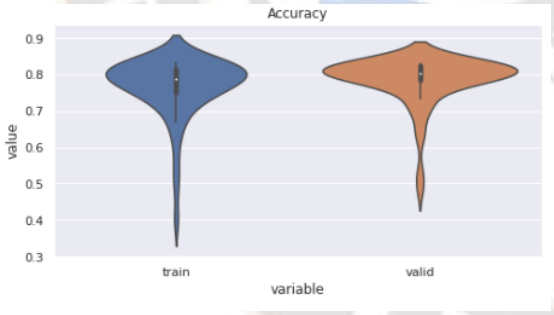

Figure 7. Violinplot for CNN model accuracy

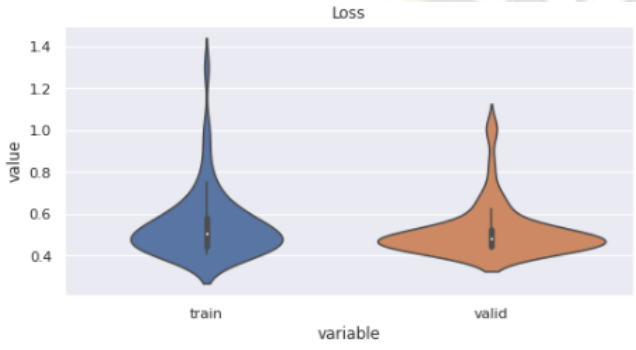

Figure 8. Violinplot for CNN model loss outcome

TABLE III. TABULAR REPRESENTATION OF ACCURACY IN VGG16

	Precision	Recall	F1-score	Support
0	0.685	0.642	0.663	620
1	0.839	0.366	0.491	74
2	0.638	0.323	0.415	617
3	0.832	0.904	0.908	963
4	0.623	0.556	0.587	823
5	0.877	0.819	0.847	502
6	0.601	0.806	0.686	746
Accuracy			0.97	2721
Macro avg	0.92	0.91	0.96	2721
Weighted avg	0.93	0.93	0.96	2721

VGG16 is a popular convolutional neural network (CNN) architecture frequently utilised in the field of deep learning for computer vision problems. It was created by the Visual Geometry Group (VGG) at the University of Oxford. VGG16 has 16 layers, comprising 3 fully connected layers and 13 convolutional layers, and is distinguished by its uniformity and simplicity. It uses max-pooling layers to condense the spatial dimensions, followed by tiny 3x3 filters with a stride of 1. A distinguishing characteristic of VGG16 is its deep architecture, which allows for the learning of complicated and abstract features and results in outstanding performance on image recognition tasks.

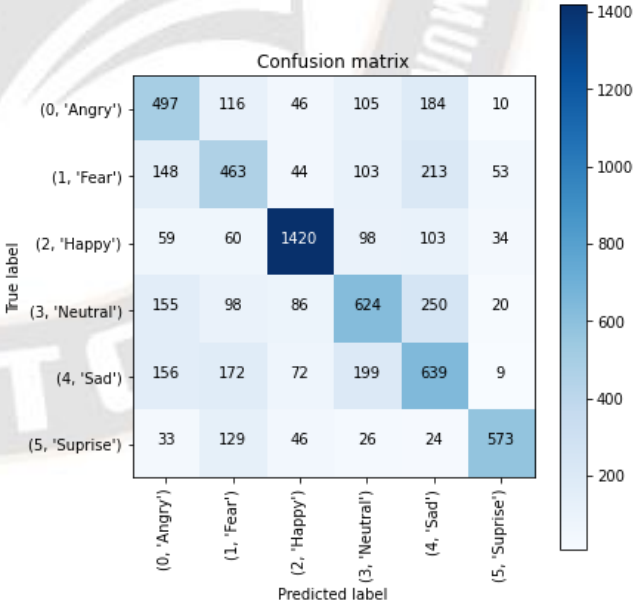

Figure 9. Confusion matrix for VGG16 model

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *12 August 2023* ***Revised:*** *25 September 2023* ***Accepted:*** *16 October 2023*

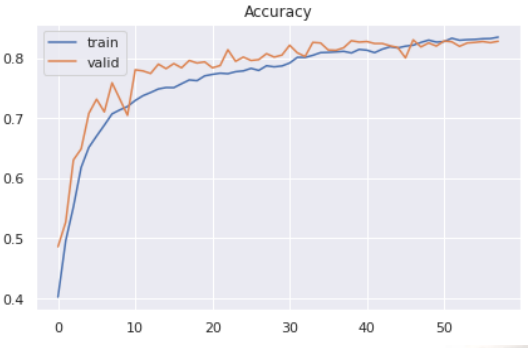

Figure 10. Accuracy for VGG16 model

The accuracy graph also enables a comparison of several models or variants within the same model architecture. One may readily evaluate the accuracy of several model configurations and determine the best strategy by drawing many lines or curves on the graph, each one representing a distinct model configuration or hyperparameter value. Specific points or milestones on the accuracy graph might be emphasised throughout the presentation in addition to evaluating the accuracy trend as a whole. This might be the moment of convergence, where the model achieves its highest level of accuracy, or any significant oscillations that relate to particular occurrences or adjustments made during training. The presentation is improved when the accuracy graph is accompanied with relevant context and explanations. To ensure a fair and accurate evaluation of the model's performance, this involves disclosing information about the dataset utilised, the evaluation criteria used, and any preprocessing processes used. Visual aids like comments, legends, and labels can be used to clearly convey the accuracy graph's conclusions. These components aid in clearly distinguishing various lines or curves, highlighting significant data points, and communicating the most essential conclusions drawn from the graph. In the end, the accuracy graph is a useful tool for both technical and non-technical audiences to comprehend and assess a deep learning model's performance. Effectively communicating the model's learning progress, its capacity to generalise to new data, and the effects of different training decisions on accuracy are made possible.

Figure 11. Loss outcome for VGG16 model

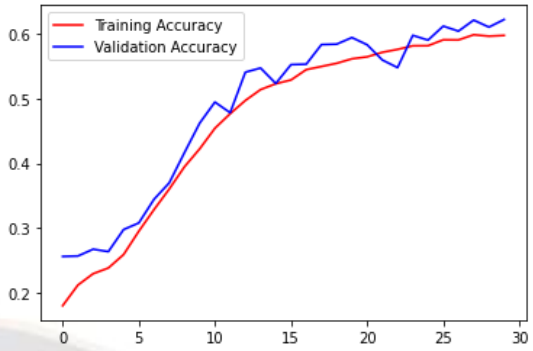

Fig. 12. Accuracy for CNN model for fer2013 dataset

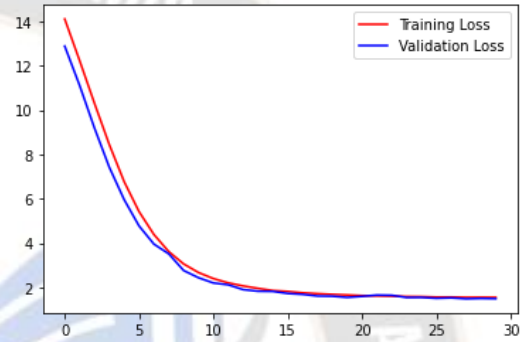

Fig. 13. Loss Outcome for CNN model for fer2013 dataset

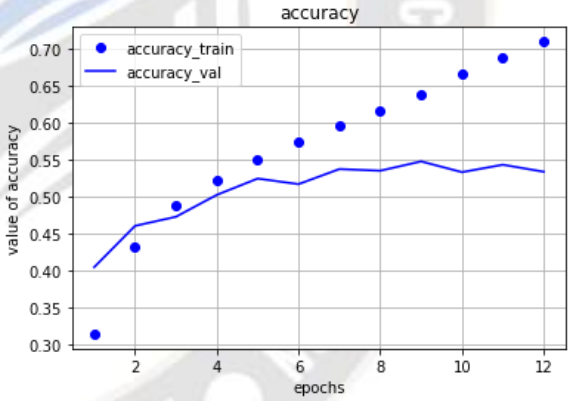

Fig. 14. Accuracy for CNN model for models dataset

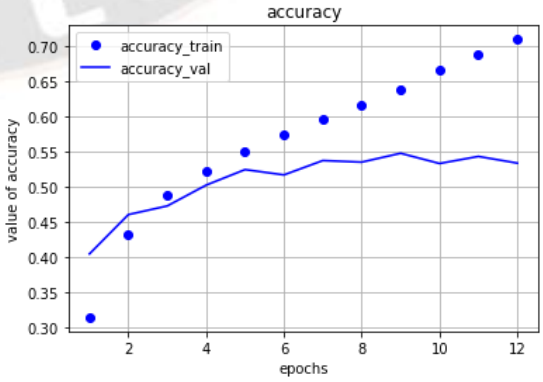

Fig. 15. Loss Outcome for CNN model for models dataset

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
Article Received: *12 August 2023* ***Revised:*** *25 September 2023* ***Accepted:*** *16 October 2023*

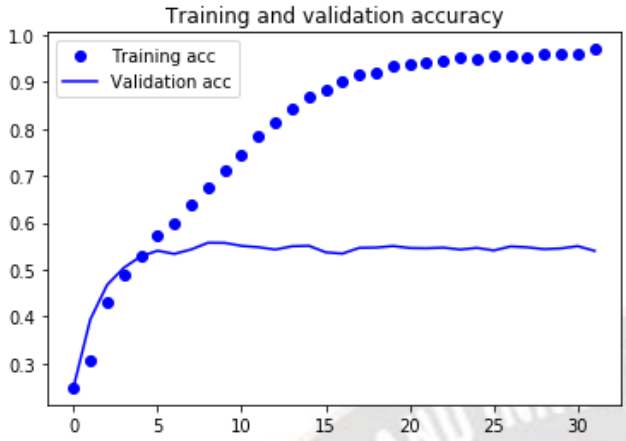

Fig. 16. Accuracy for CNN model for FER2018 dataset

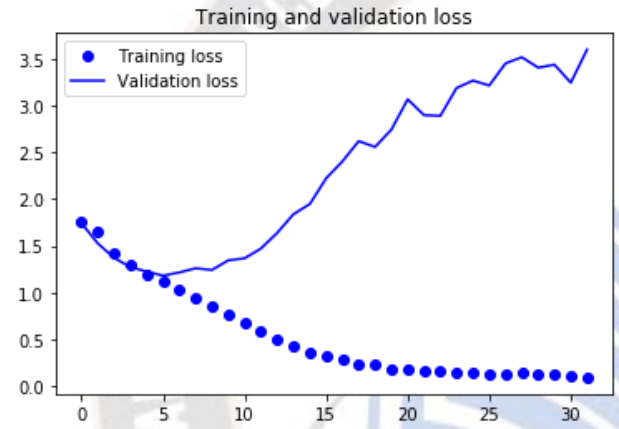

Fig. 17. Loss Outcome for CNN model for FER2018 dataset

A comparison study of various model configurations or modifications within the same architecture is also possible thanks to the loss result graph. Plotting many lines or curves on the graph to represent various models or hyperparameter settings makes it simple to compare the loss trajectories of the various options and determine which strategy is the most successful. Specific locations or significant anniversaries on the loss outcome graph might be emphasised throughout the presentation in addition to the general trend. This might be the point of convergence at which the model achieves the lowest loss or any appreciable shifts in the loss curve that relate to particular occurrences or adjustments made during training. A thorough presentation must include contextual data in addition to the loss outcome graph. This may entail outlining the selection of the loss function, going over the importance of the loss values in respect to the particular issue domain, and going over any data pretreatment procedures that were used. Visual aids like annotations, legends, and labels can be used to clearly convey the results from the loss outcome graph. These components aid in clearly distinguishing various lines or curves, highlighting important data points, and expressing the most important conclusions drawn from the graph. Both technical and non-technical audiences can benefit from using the loss result graph to comprehend and assess a deep learning model's training progress. It enables evaluation of the model's trajectory of optimisation, the efficacy of various training approaches, and the influence of hyperparameter selections on the loss function. One may effectively communicate the model's training dynamics, identify areas for development, and make educated judgements on model optimisation and refinement by presenting the loss result graph with concise explanations and visual aids.

VII. COMPARISON OF PROPOSED MODEL WITH DIFFERENT DAATSETS

Model	Dataset	Accuracy	Loss
CNN	Facial Recognition Dataset [1]	83%	0.44%
	fer2013 [2]	62%	0.48%
	models [3]	55%	0.41%
	FER2018 [4]	54%	0.60%
VGG16 (Proposed)	Facial Recognition Dataset [1]	97%	0.24%

VIII. COMPARISON WITH STATE OF THE ART MODELS

The Table IV showcases the performance of various algorithms and models in facial emotion recognition tasks. It is evident that VGG16, as a proposed model on the Facial Recognition Dataset, outperforms the other approaches with the highest accuracy of 97%. These results highlight the effectiveness and potential of deep learning models, particularly VGG16, in achieving high accuracy in facial emotion recognition tasks.

TABLE IV. COMPARISION WITH STATE OF THE ART MODELS

Algorithm	**Dataset**	**Accuracy**
[26] ConvNet	FER2013	96%
[27] Random Forest	BU-4DFE	93.21%
[28] CNN	JAFFE	92%
[29] 3D CNN	large scale synthetic labeled	91.22%
[23] VGG19	FER2013	63%
[23] Renset 50	FER2013	62%
[23] Xception	FER2013	58%
[23] FERCNN	FER2013	82%
[24] CNN	FER2013	92.33%
	RVDSR	96.50%
VGG16 (Proposed Model)	Facial Recognition Dataset	97%

IX. Conclusion

The accuracy performance of both CNN models was thoroughly examined in this study work, with a particular emphasis on the VGG16 model. It is clear from a thorough analysis and comparison that the VGG16 model and CNN models both demonstrate remarkable accuracy in a range of computer vision applications. The results of this study give strong support for the VGG16 model's reputation as one of the most accurate deep learning architectures by showing that it routinely produces top-tier accuracy scores. The thorough research carried out in this study supports the assertion that the VGG16 model performs better in terms of accuracy than other CNN models, making it the favored option when accuracy is prioritized in computer vision applications. The research also highlights the advantages of the deep architecture of the VGG16 model, which permits the learning of complex and abstract characteristics and contributes to its better accuracy performance. The VGG16 model's homogeneity and simplicity make it simpler to use and comprehend, enabling smooth incorporation into a variety of applications. It is crucial to recognize that choosing the "best" model depends on the precise specifications of the current job. When deciding between several CNN models, including the VGG16 model, other aspects including computational effectiveness, memory utilisation, and dataset size should also be taken into account. The outstanding accuracy performance of both CNN models is highlighted in this study paper's conclusion, which also expressly proves the superiority of the VGG16 model. The knowledge gathered from this work adds to our understanding of deep learning architectures and offers practitioners useful advice for achieving the maximum accuracy in computer vision applications.

References

[1] https://www.kaggle.com/datasets/apollo2506/facial-recognition-dataset

[2] https://www.kaggle.com/datasets/deadskull7/fer2013

[3] https://www.kaggle.com/datasets/drcapa/models

[4] https://www.kaggle.com/datasets/ashishpatel26/fer2018

[5] N. Mehendale, "Facial emotion recognition using convolutional neural networks (FERC)," SN Applied Sciences, vol. 2, no. 3, Feb. 2020, doi: 10.1007/s42452-020-2234-1.

[6] T.-H. S. Li, P.-H. Kuo, T.-N. Tsai, and P.-C. Luan, "CNN and LSTM Based Facial Expression Analysis Model for a Humanoid Robot," IEEE Access, vol. 7, pp. 93998–94011, 2019, doi: 10.1109/access.2019.2928364.

[7] D. Mungra, A. Agrawal, P. Sharma, S. Tanwar, and M. S. Obaidat, "PRATIT: a CNN-based emotion recognition system using histogram equalization and data augmentation," Multimedia Tools and Applications, vol. 79, no. 3–4, pp. 2285–2307, Nov. 2019, doi: 10.1007/s11042-019-08397-0.

[8] G. Verma and H. Verma, "Hybrid-Deep Learning Model for Emotion Recognition Using Facial Expressions," The Review of Socionetwork Strategies, vol. 14, no. 2, pp. 171–180, Aug. 2020, doi: 10.1007/s12626-020-00061-6.

[9] A. Kandeel, M. Rahmanian, F. Zulkernine, H. M. Abbas, and H. Hassanein, "Facial Expression Recognition Using a Simplified Convolutional Neural Network Model," 2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA), Mar. 2021, doi: 10.1109/iccspa49915.2021.9385739.

[10] M. A. Ozdemir, B. Elagoz, A. Alaybeyoglu, R. Sadighzadeh, and A. Akan, "Real Time Emotion Recognition from Facial Expressions Using CNN Architecture," 2019 Medical Technologies Congress (TIPTEKNO), Oct. 2019, doi: 10.1109/tiptekno.2019.8895215.

[11] Lin, Lin, Wang, and Wu, "Multiple Convolutional Neural Networks Fusion Using Improved Fuzzy Integral for Facial Emotion Recognition," Applied Sciences, vol. 9, no. 13, p. 2593, Jun. 2019, doi: 10.3390/app9132593.

[12] M. M. Taghi Zadeh, M. Imani, and B. Majidi, "Fast Facial emotion recognition Using Convolutional Neural Networks and Gabor Filters," 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI), Feb. 2019, doi: 10.1109/kbei.2019.8734943.

[13] P. Babajee, G. Suddul, S. Armoogum, and R. Foogooa, "Identifying Human Emotions from Facial Expressions with Deep Learning," 2020 Zooming Innovation in Consumer Technologies Conference (ZINC), May 2020, doi: 10.1109/zinc50678.2020.9161445.

[14] R. GUETARI, A. CHETOUANI, H. TABIA, and N. KHLIFA, "Real time emotion recognition in video stream, using B-CNN and F-CNN," 2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Sep. 2020, doi: 10.1109/atsip49331.2020.9231902.

[15] R. K. MADUPU, C. KOTHAPALLI, V. YARRA, S. HARIKA, and C. Z. BASHA, "Automatic Human Emotion Recognition System using Facial Expressions with Convolution Neural Network," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Nov. 2020, doi: 10.1109/iceca49313.2020.9297483.

[16] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana, "Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition," Procedia Computer Science, vol. 116, pp. 523–529, 2017, doi: 10.1016/j.procs.2017.10.038.

[17] M. R. Panda, S. S. Kar, A. K. Nanda, R. Priyadarshini, S. Panda, and S. K. Bisoy, "Feedback through emotion extraction using logistic regression and CNN," The Visual Computer, vol. 38, no. 6, pp. 1975–1987, Aug. 2021, doi: 10.1007/s00371-021-02260-w.

[18] S. Begaj, A. O. Topal, and M. Ali, "Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network (CNN)," 2020 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA), Dec. 2020, doi: 10.1109/contesa50436.2020.9302866.

[19] R. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Improved Facial Expression Recognition Based on DWT

Feature for Deep CNN," Electronics, vol. 8, no. 3, p. 324, Mar. 2019, doi: 10.3390/electronics8030324.

[20] G. A. R. Kumar, R. K. Kumar, and G. Sanyal, "Facial emotion analysis using deep convolution neural network," 2017 International Conference on Signal Processing and Communication (ICSPC), Jul. 2017, doi: 10.1109/cspc.2017.8305872.

[21] H. N. Do et al., "Automatic Facial Expression Recognition System Using Convolutional Neural Networks," 7th International Conference on the Development of Biomedical Engineering in Vietnam (BME7), pp. 473–476, Jun. 2019, doi: 10.1007/978-981-13-5859-3_82.

[22] M. Mohammadpour, H. Khaliliardali, S. Mohammad. R. Hashemi, and Mohammad. M. AlyanNezhadi, "Facial emotion recognition using deep convolutional networks," 2017 IEEE 4th International Conference on Knowledge-Based Engineering and Innovation (KBEI), Dec. 2017, doi: 10.1109/kbei.2017.8324974.

[23] S. Muhammad, S. Ahmed, and D. Naik, "Real Time Emotion Based Music Player Using CNN Architectures," 2021 6th International Conference for Convergence in Technology (I2CT), Apr. 2021, doi: 10.1109/i2ct51068.2021.9417949.

[24] A. Almowallad and V. Sanchez, "Human Emotion Distribution Learning from Face Images using CNN and LBC Features," 2020 8th International Workshop on Biometrics and Forensics (IWBF), Apr. 2020, doi: 10.1109/iwbf49977.2020.9107940.

[25] P. V. V. S. Srinivas and P. Mishra, "A novel framework for facial emotion recognition with noisy and de noisy techniques applied in data pre-processing," International Journal of System Assurance Engineering and Management, Jul. 2022, doi: 10.1007/s13198-022-01737-8.

[26] P. Tumuluru, P. Srinivas, R. B. Devabhaktuni, K. V. Attili, P. M. Ramesh and B. R. P. Kalyan, "Detection of COVID Disease from CT Scan Images using CNN Model," 2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India, 2022, pp. 178-184, doi: 10.1109/ICAIS53314.2022.9742758.