International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7636
***Article Received**: 25 April 2023 **Revised:** 24 June 2023 **Accepted:** 10 July 2023*

Multi-Network Feature Fusion Facial Emotion Recognition using Nonparametric Method with Augmentation

Vandana Devi[1], Avinash Sharma[2]
[1]PhD Scholar, Computer Science & Engineering
[2]Professor, Computer Science & Engineering
[1,2]Maharishi Markandeshwar Deemed University Mullana, Ambala, Haryana, India
[1]vandukhanchi@gmail.com[1], asharma@mmumullana.org[2]

Abstract— Facial expression emotion identification and prediction is one of the most difficult problems in computer science. Pre-processing and feature extraction are crucial components of the more conventional methods. For the purpose of emotion identification and prediction using 2D facial expressions, this study targets the Face Expression Recognition dataset and shows the real implementation or assessment of learning algorithms such as various CNNs. Due to its vast potential in areas like artificial intelligence, emotion detection from facial expressions has become an essential requirement. Many efforts have been done on the subject since it is both a challenging and fascinating challenge in computer vision. The focus of this study is on using a convolutional neural network supplemented with data to build a facial emotion recognition system. This method may use face images to identify seven fundamental emotions, including anger, contempt, fear, happiness, neutrality, sadness, and surprise. As well as improving upon the validation accuracy of current models, a convolutional neural network that takes use of data augmentation, feature fusion, and the NCA feature selection approach may assist solve some of their drawbacks. Researchers in this area are focused on improving computer predictions by creating methods to read and codify facial expressions. With deep learning's striking success, many architectures within the framework are being used to further the method's efficacy. We highlight the contributions dealt with, the architecture and databases used, and demonstrate the development by contrasting the offered approaches and the outcomes produced. The purpose of this study is to aid and direct future researchers in the subject by reviewing relevant recent studies and offering suggestions on how to further the field. An innovative feature-based transfer learning technique is created using the pre-trained networks MobileNetV2 and DenseNet-201. The suggested system's recognition rate is 75.31%, which is a significant improvement over the results of the prior feature fusion study.

Keywords-: Augmentation, DenseNet-201, Emotion, Face Expression Recognition Dataset, MobileNetV2, Neighbourhood Component Analysis

I. INTRODUCTION

Understanding and reading facial expressions is a crucial part of nonverbal communication. It improves the clarity of verbal communication and facilitates the retention of information [1][2]. Human attention may be detected in a number of ways, including conduct, mental state, personality, criminal propensity, deception, and so on. Most humans are adept at reading facial expressions, regardless of their gender, culture, ethnicity, or location. Unfortunately, automating facial emotion recognition and categorization is a difficult problem. Some common emotions used in scientific studies include anxiety, anger, sadness, and enjoyment [3]. Unfortunately, it is also very difficult for robots to distinguish between a wide range of emotions. More than that, robots need to be taught to comprehend their context and the human mind's goals. Robots and computers are also included when the word "machine" is used. Robots are distinct in that their autonomy is baked into their design, allowing for more creative use of communication skills [4]. The biggest difficulty in labelling people's feelings is that there are so many different factors to consider, such as gender, age, race, ethnicity, and the quality of the images or videos being used. A system that can recognise facial expressions with the same level of understanding as humans are essential. In the past several decades, FER (Facial Expression Recognition) has emerged as a promising new area of study. To begin the FER process, a video or still picture must first have its faces detected. The photos include more than just people's faces—they also show elaborate settings. While humans have a natural talent for reading people's expressions and other facial cues from photographs, computers have a far harder time doing so. Face detection's main goal is to extract faces from their respective backgrounds (non-faces). Teleconferencing, tagging, and face identification are only a few of the many applications of face detection [5]. On the contrary, the ability to see a face from a variety of angles is lost when using a window-based approach. For FER, face identification is among the most often utilised applications of model matching methods. The window-based

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7636
***Article Received**: 25 April 2023 **Revised:** 24 June 2023 **Accepted:** 10 July 2023*

method, on the other hand, is limited in its ability to capture multi-perspective facial expressions. Modern classifiers including are employed in a variety of applications nowadays, such as face detection in security, biometrics, tumour diagnosis in healthcare, and handwriting study [6]. It is a delicate and challenging undertaking to extract traits from one face and apply them to another in order to improve categorization. A system for describing facial motions using a set of standardised units called Action Units (AUs). While automatic FER has received the greatest attention from studies [7] noting that "each individual communicates his emotion by his manner," this is no simple undertaking. In this work, together with the neutral state, we provide a unique fusion-based technique for identifying human facial expressions of the six main emotions. As both the MobileNetV2 and DenseNet-201 deep learning pre-trained networks extract features, but use different data sources, we suggest fusing the features from both networks to improve identification accuracy. Moreover, we use a Neighbourhood Component Analysis (NCA) based feature selection approach to zero down on the most important characteristics. Confusion matrices have then been used to evaluate the performance of four different classifiers (KNN, Naive Bays, Ensemble, and Support vector machine) with respect to identifying each of the identified emotions. Also, reprocessing the input picture with the augmentation that was applied to it is a potential new pre-processing step that might improve the output's accuracy. By fusing two separate deep learning networks and then filtering the results with the aid of an advanced feature selection approach called NCA, the identification rate is much improved. This paper will be organised as follows. In Part II, we saw several similar writings. In Part III, we explain the suggested fusion-based approach to recognising facial emotion expressions and provide a full explanation of the technique. After describing the procedure for conducting the experiments in Part IV, Section V presents the findings and their interpretations. The final thoughts and next projects described in Section IV.

II. RELATED WORKS

In this work, Mehmood et al. [4] used the best approaches for emotion identification using EEG sensors implanted in the brains of living humans. In this work, we optimize the face recognition approach and apply it to the extraction and selection of EEG features. Emotions are categorised into four broad categories: joy, serenity, melancholy, and fear. Extraction of features uses an optimally chosen feature, such as, to improve the accuracy of emotional categorization. Improved EEG recognition may be achieved with the use of supplementary methods like the arousal-valence space. In this study, Chen et al. [8] developed a Soft-max Regression-based Deep Sparse Auto-encoder Network for Human-Robot Collaboration in Expression Recognition. This study employs the SRDSAN method to aid in minimizing deformation and determining the learning effectiveness and dimensional complexity, with softmax prediction being used to classify the input signal. In this study, Hassouneh et al. [9] used electroencephalograms and facial expressions and a real-time emotional recognition system is developed using deep neural networks and machine learning.. In this work, we use digital markers to develop the optical flow technique for face regression analysis. As the optical flow algorithm method acknowledges reduced computing complexity, it is useful for physically challenging folks. In this study, Tan et al. [10] developed spiking neural network models of spatiotemporal EEG signals using neuro-sense for interpreting and identifying short-term emotions. Now, for the first time, the SNN method is really put into practice. In doing so, the brain's many roles may be better understood. Two methods, including arousal-valence space, are used to quantify the EEG data. Methods that are high in arousal, methods that are low in arousal, and methods that are both high and low in valence make up the four columns that make up the arousal-valence space. Here, Li et al. [5] construct a thorough face expression recognition database. Identification of facial expressions is one of the network system's biggest challenges. These neural pipeline approach divide up the dataset in a particular manner first. This will aid in alleviating some of the difficulties inherent to the FER method.

In this study, Sati et al. [11] used NVIDIA for both face detection and identification, as well as facial emotion recognition. Jetson Face-to-face interactions are still one of the most challenging ways to identify emotions. This study develops nano, which enables both recognizing faces and face detection of feelings by incorporating unique features into this method. The artificial neural network method makes it easier to recognise and categorise facial emotional expressions. In this study, Wang et al. [6] put into practice a cutting-edge deep learning approach. In this study, we use the deep learning four-category model. Convolutional neural networks and other deep architectures make up the first group. The deep learning paradigm has a significant impact on neural networks. It plays a vital role in the algorithm used for machine learning. It's crucial for the accuracy of the data, and both nonlinear and linear characteristics are included in the classification.

III. PROPOSED METHOD

The proposed face emotion prediction system consists of five stages: image pre-processing, augmentation, feature extraction using a deep learning pre-trained network, feature selection using the NCA technique, and classification. This part serves as an introduction to the comprehensive facial expression

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7636
***Article Received**: 25 April 2023 **Revised:** 24 June 2023 **Accepted:** 10 July 2023*

recognition system. Although the CNN model is the major emphasis, we also include a number of feature extraction and pre-processing procedures to see how much of an increase in accuracy can be accomplished. First, we extract features from all of the training photographs in the database, and then we use those features to construct feature vectors for each image in the database. First, features are extracted from the enhanced dataset using an already-trained deep-learning network; then, features from multiple pre-trained networks are fused; then, the NCA technique is applied for additional feature selection to improve accuracy; and, finally, the classifier is used by the facial emotion recognition system to recognise the facial emotions category.

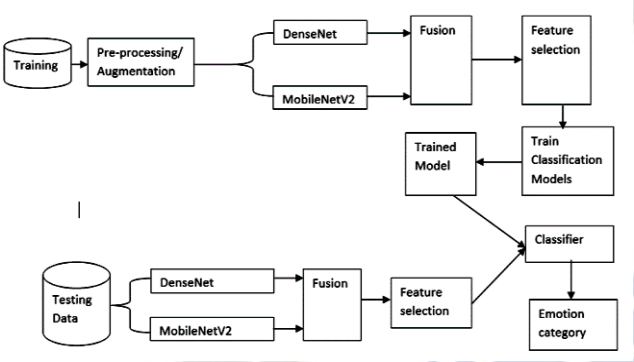


Fig. 1. Demonstrates the proposed method's process.

One facial image is used to generate one of seven labels representing an individual's emotional state (i.e., fear, anger, sad, disgust, happy, surprise, and neutral). There are no suggested novel neural network designs or training methods for FER in this paper. To circumvent this problem, we first use any readily-available pre-trained deep learning network architecture to provide a crude FER result, which we subsequently refine. The next section will cover how to assess the reliability of the first result and make adjustments to the classification. The complexity of our FER method may be reduced to a simple seven-class problem. When trained, the FER network generates a 7-dimensional vector where each dimension indicates the probability of an emotion. Output vectors containing images of people displaying strong yet distinguishable facial expressions, such as a broad smile, tended to have one element with a noticeable peak and the rest with relatively modest values. In other words, if the FER network's output vector does not have a clear peak, we may consider the classification result to be neutral. The NCA (Neighbourhood Component Analysis) feature selection method will be discussed in detail below. To evaluate the viability of our suggested technique on the Face Expression Recognition Dataset, we implement a 7-category facial expression classification using cutting-edge technology.

A. Pre-processing

We did not alter the photographs in any way before applying them to the datasets, and we used the same parameters regardless of whatever dataset we were working with. The accuracy of object recognition in grayscale images is higher than in their RGB equivalents, which is why they are increasingly used for rendering. In addition, computational expenses may be lowered by using grayscale images. Here are some approaches to cleaning and organizing raw data for analysis face detection, cropping, image normalization, image resizing, and image segmentation. As MobileNetV2 and DenseNet need pictures to be 224x224 pixels in size, while the photographs in the dataset are only 48x48 pixels, only scaling is performed. This adjustment is made so that the DenseNet201 and MobileNetV2 models can handle the input data.

B. Augmentation

The quality of a convolutional neural network's results is greatly improved by providing it with a massive amount of data. A bigger dataset allows for the possibility of extracting more features, which is useful for boosting productivity. Sometimes it's just not feasible to get all the data needed, thus data augmentation is utilized to make up the difference and make the system work better. "Image augmentation" involves the generation of new visual content by modifying existing visual assets. Data augmentation is used to improve the efficiency of neural network models by compensating for inherent data bias and expanding the model's applicability. photographs from various sources warped at random (rotation, translation, shear, flipping, etc.). Public image-labeled databases often use data augmentation techniques to improve database dimensions. Because of their ease of use, geometric data augmentation techniques have become the most popular. To improve training images, using geometric transformations (such as translations, rotations, and scaling) might have a number of implications. The process of augmenting data is simplified by the imageDataAugmenter function, which accepts as arguments the ability to translate and scale the data as well as rotate it by a specified angle.

C. Feature Generation

As a result of its robustness against the effects of distance and size, our method makes use of CNN. Moreover, the output feature map from a convolutional neural network is often less in size than the input. In the field of deep learning, "transfer learning" refers to the practise of reusing a model that has previously been trained. Due to time and effort limitations, building a model from scratch

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7636
***Article Received**: 25 April 2023 **Revised:** 24 June 2023 **Accepted:** 10 July 2023*

is not practical. The pre-trained network is made ready for use as a generic method for emotion detection by means of transfer learning on a large facial expression recognition dataset. The Facial Expression Recognition Dataset was used to train all of the networks used in this paper, which is a relatively new dataset for face verification. The model of the MobileNetV2 architecture [13] is shown in Fig. 2. In this era of ubiquitous mobile connectivity, networks must be both fast and lightweight. When used in a real-time framework, MobileNetv2's speed and accuracy provide reliable results in a timely manner. The model does quite well on the ImageNet database [14]. So, it might be utilized to determine a person's identity based on their facial features and emotional stat [15].

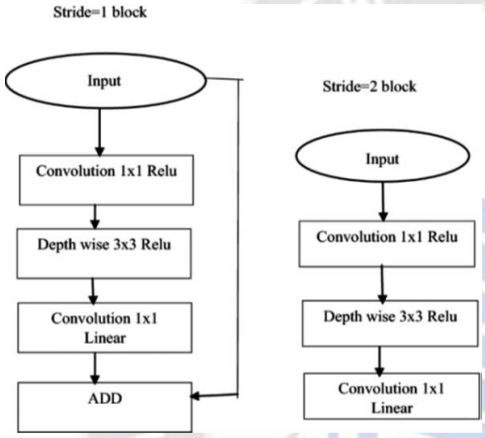


Fig. 2 MobileNetV2 architecture

DenseNet's advantages lie in the fact that it reduces the number of variables in use, has a low gradient issue, and allows for the reuse and implementation of features [16]. The DenseNet-201 is a 201-layer pre-trained convolutional neural network. The maximum allowed picture size on the network is 224 by 224 pixels. As can be seen in Fig. 3 [17], each layer is made up of a 3*3 filter, ReLU activation, and batch normalisation (BN). Instead of adding the output feature maps from all preceding layers as indicated by equations [18], DenseNet contributed new features to the model by concatenating them all progressively.

$$X1l = Hl\,(x1l-1) \quad (1)$$

$$Xl1 = Hl\,(x1l-1)+xl-1 \quad (2)$$

$$X1l = Hl\,([xo\,x1, x2, \ldots, xl-1]) \quad (3)$$

The feature of the lth layer is represented by xl. l for the layer index and H0 for the non-linear operation are written as above.

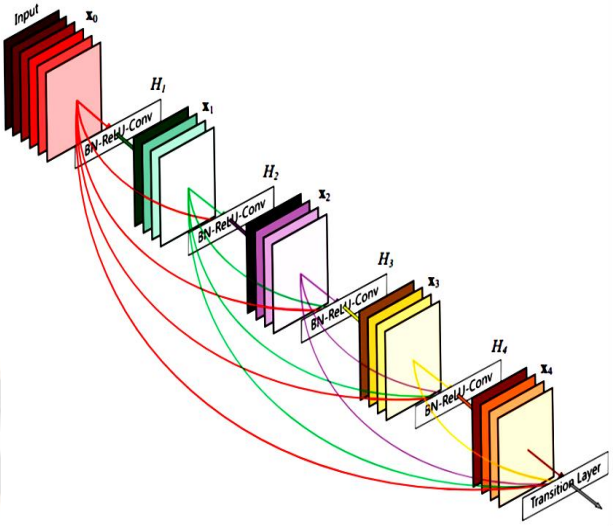


Fig. 3 MobileNetV2 architecture

This study used the DenseNet-201 [5 + (6 + 12 + 48 + 32) 2) = 201] architecture, one of the three DenseNet models (DenseNet-121, DenseNet-160, and DenseNet-201). Here are some DenseNet-201 specifics: 5 layers of convolution and pooling, 3 layers of transition (6,12,48), 1 layer of classification (32), and 2 layers of dense-block (1-1 and 3-3 conv) make up the stack as shown in table 1 [17].

TABLE 1. DENSENET-201 STRUCTURE

Layers	**Output Size**	**DenseNet-201**
Convolution	112 × 112	7 × 7 conv, stride 2
Pooling	56 × 56	3 × 3 max pool, stride 2
Dense Block (1)	56 × 56	[1×1 conv; 3×3 conv] ×6
Transition Layer (1)	56 × 56	1 × 1 conv
	28 × 28	2 × 2 average pool, stride 2
Dense Block (2)	28 × 28	[1×1 conv; 3×3 conv] ×12
Transition Layer (2)	28 × 28	1 × 1 conv
	14 × 14	2 × 2 average pool, stride 2
Dense Block (3)	14 × 14	[1×1 conv; 3×3 conv] ×48
Transition Layer (3)	14 × 14	1 × 1 conv
	7 × 7	2 × 2 average pool, stride 2
Dense Block (4)	7 × 7	[1×1 conv; 3×3 conv] ×32
Classification layer	1 × 1	7 × 7 global average pool
		1000D fully-connected, SoftMax

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7636
***Article Received**: 25 April 2023 **Revised:** 24 June 2023 **Accepted:** 10 July 2023*

D. Feature Fusion and Selection

The fusing of the generated feature by various layers or deep learning networks has become a crucial part of today's deep convolutional learning networks. These feature vectors are catenated together to create higher dimensional fusion features after the two networks used to extract features from each image have finished. And this can make up for a single network's feature extraction shortcomings. However, there is always some connection and information duplication among these data, and more complex computations are required for features with higher dimensions. As a result, we use neighborhood component analysis (NCA) to pick features and reduce their dimensionality. The features extracted by DenseNet-201 and MobileNetV2 are denoted as $F_{D,}$ F_M respectively, and then the fused features F are concatenated according to Equation (4), which is shown as follows:

$$F=\text{Concatenate}\ (F_M, F_{D,}) \quad (4)$$

Using linear methods to minimize dimensionality, the original data may be represented more briefly, like Neighboring Components Analysis (NCA) use a linear operator on the raw data. The purpose of this research is to evaluate and contrast several NCA methods for reducing the dimension of high-dimensional datasets [19]. We used the cutting-edge method of NCA to choose more relevant features, which improved the quality of our study. An easy learning algorithm may benefit from this technique since it improves the efficiency with which it classifies data. Selecting features for face emotion identification based on how they look uses neural coherence analysis. It's possible that the size of the resulting descriptor vectors or features will be rather huge. Overfitting, longer training times, and worse accuracy might also result if the feature vectors had noisy or duplicated properties. As part of our method, we provide a feature selection strategy that ranks attributes and shrinks vectors. To maximize the prediction or recognition accuracy of classification algorithms, we used an NCA-based strategy, a non-parametric approach to picking features. Combining the outputs of MobileNetV2 and DenseNet into a single feature vector allows us to use Neighbourhood Component Analysis (NCA) to extract the reduced but most significant aspects of facial photos, which might be used to identify nuanced expressions of emotion. Several classifiers are fed the resulting descriptor vector.

E. Classification

Data mining is often used for tasks involving classification. The main goal of classification is to assign a class label to an unknown instance made up of a number of variables that vary from a range of possible class values. In particular, a classifier model is applied, which is created by running an algorithm for learning on a training set of past instances that have the same characteristics as the unknown case. However, the class designations for the training set are completely predetermined. The final classifier model is evaluated for effectiveness on a different data set once the training phase is over. There are plenty of tools at the ready when it comes to categorising. As a result, convolutional neural networks (CNNs) and deep CNNs have been used in recent research.To mention a few, these include Bayesian belief networks, rule-based systems, discriminant analysis, logistic regression, support vector machines, artificial neural networks, decision tree algorithms, and support vector machines. The K-nearest Neighbor (KNN), Support Vector Machine (SVM), Naive Bayes (NB), and Ensemble models are just a few of the several models used in the study. Classification is key to the suggested fusion-based automated FER technique. After being trained on a dataset of labelled examples, it can recognise a wide variety of facial expressions. As such, we use a set of supervised machine learning methods including SVM, KNN, NB, and Ensemble. Furthermore, the strength and scalability of this kind of categorization method are what inspired us to start employing these. This study tries to identify and detect certain emotions, which are a fundamental part of how we communicate our judgment and decision-making in everyday life. Face recognition is a method of determining or confirming the identification of a person in still or moving visual media. Emotions including happiness, sadness, anger, fear, surprise, neutral and disgust may all be detected by this study. In this effort, we use the aforementioned four machine learning algorithms to identify and categorize face expressions.

IV. RESULTS AND INTERPRETATIONS

A. Collection of information

Machine learning algorithms need a dataset before they can learn to understand and identify human emotions. There are a plethora of datasets available to researchers currently, including AffectNet, the Extended Cohn-Kanade Dataset (CK+), Emotic, and many more. All the pictures from the FER database are 48x48 grayscale images. The facial features of each subject have been automatically registered to provide a consistent cant and proportional fill of the frame across all images. Pictures are captioned with descriptions of the subject's emotions, which might range from anger to contempt to fear to happiness to sadness to surprise to indifference. There are about 36,000 images in

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7636
***Article Received**: 25 April 2023 **Revised:** 24 June 2023 **Accepted:** 10 July 2023*

all. To keep things interesting, we shuffled the split of the dataset with each iteration of our technique.

Fig. 4. Sample images in face expression recognition dataset

B. *Experimental environment*

The effectiveness of each classifier is assessed using the Matlab platform and the following hardware:

(i) Intel Core i5 9th version 16GB memory device (

(ii) 4xNVIDIA Pro Version Quadro, P6000 RTX PCIe 3.0-24GB

C. *Performance Evaluation*

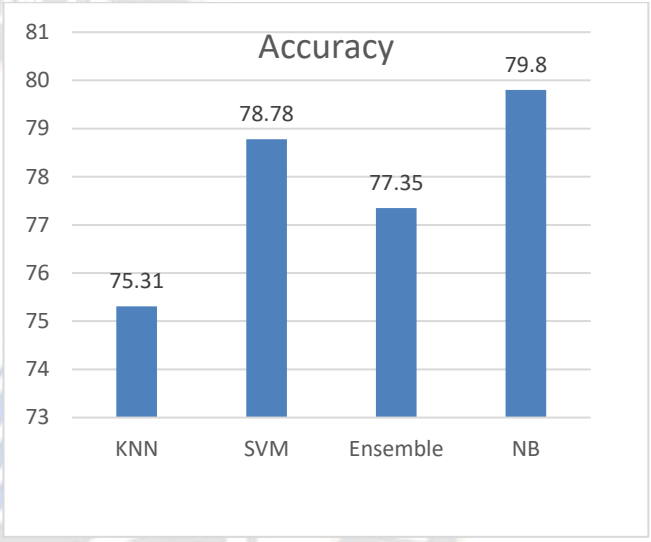
A comparison of the classification approach's efficacy for identifying the six fundamental emotions and the neutral state of face photographs is shown in a graph after the corresponding trials have been conducted. The face picture collection was split into 80 percent for use in development and 20 percent for use in testing. Due to the uneven distribution of face photos across emotion categories, we used a dataset of 700 images, 100 from each category, to train and test our networkF1-measure and Accuracy are two performance metrics used to assess how well the categorization model is performing. The percentage of incidents that can be accurately predicted is known as accuracy. In some cases, accuracy alone is insufficient to assess a model's performance in classification tasks, thus we additionally determined the accuracy of a test's F-measure (F-score).

$$\text{Accuracy} = ((TP + TN) / (TP + TN + FP + FN))$$

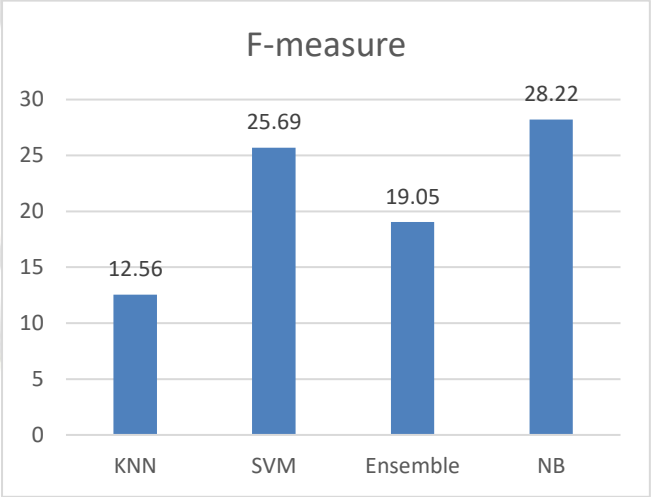
True positive (TP) occurs when the model accurately predicts the positive class. If the model recognizes the negative class, true negative (TN) results. False positives (FP) and false negatives (FN) occur when the model correctly detects the positive and negative classes, respectively.

$$\text{F-measures} = 2\times ((\text{Precisions} \times \text{Recalls}) / (\text{Precisions} + \text{Recalls}))$$

Where F-measure is the calculated median of recall and precision measurements, which balances the performance of both metrics and is effective for disproportionate classification. Recall evaluates a classifier's ability to recognise positive examples, whereas precision counts the percentage of properly predicted positive classifications.



Graph 1. Performance Evaluation Measure on Accuracy Metric



Graph 2. Performance Evaluation Measure on F-measure Metric

Using the validation data, the models' accuracy and f-measure values are calculated. Evaluation metrics are used to compare the effectiveness of different classifiers. The test dataset is consistent with the findings of the classifiers. By comparing the results of the models utilized in this research, we find that they are able to accurately categorize facial expressions of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7636
***Article Received**: 25 April 2023 **Revised:** 24 June 2023 **Accepted:** 10 July 2023*

emotion. Accuracy values, which measure the efficacy of the models, are all at least about 80%. In terms of accuracy, Naive Bayes is superior to SVM, Ensemble, and KNN. Again, the F-measure, which evaluates predictive power, favoured Naive Bayes as the top classifier. But, the SVM also exhibited strong predictive power. See how our strategy stacks up against other approaches to training accuracy in table 1 below. On the other hand, our project's usage of the sophisticated feature selection methodology NCA, applied to the fusion of features retrieved by several pre-trained deep learning networks, yields more accuracy than the traditional approach. We do comparison with the papers mentioned in below table based on fusion based process or similar datasets of face images.

TABLE 2. PERFORMANCE EVALUATION COMPARISON WITH THE PREVIOUS RESEARCH STUDIES

Authors	**Significance of respective research method**	**Accuracy**
Our proposed method (Overall accuracy on an average of used classifiers)	Multi network feature fusion using DenseNet-201 & MobileNetV2 and classified by KNN, NB, SVM, Ensemble with NCA and augmentation	75.31%
[20]	Facial expression classification by optimized MobileNet model with feature fusion	74.35%
[21]	Ensembled based classification using Custom CNN, ResNET50, InceptionV3	72.3%
[22]	Feature generated by three sub-networks AlexNet, ResNet and VGGNet are fed to SVM classifier to integrate the output of the three networks to get the final result	71.27%
[23]	Ensembled AlexNet, ResNet and VGGNet CNN for global and local feature extraction using original and cropped data	70.74%

V. CONCLUSION AND FUTURE SCOPE

Feature fusion and neighbourhood component analysis were used in our effort to enable the recognition of facial emotions. The topic of facial expression recognition is examined, with the end objective being to label images of people's faces with one of seven distinct emotional states or face expression classes. Emotion, social interaction, and cognitive science all make use of the examination of facial expressions. The MobileNetV2 model was trained on many different classes, so it's reasonable to assume that it has picked up some universal traits that may be used to deduce people's emotions and intentions. Next, the localized face may be employed in age, gender, and facial gesture recognition processes. One major drawback of earlier studies was their inadequate use of training data. In contrast, this approach uses a large number of training examples. The method used provided a potent machine learning-based solution to reduce the likelihood of judgmental errors made by humans and provide a more effective route to cost savings. The NCA's attempts to integrate attributes resulted in a number of useless qualities remaining in the final product. To learn more, we need additional data and other ways to extract features, which might result in different classifier performance. Alternate deep learning methodology or modified deep learning approaches might be used, evaluated, and built for better categorization in future research. Combining text-based, sophisticated sentiment analysis with facial-photo fusion is a future possibility.

References

[1] F. Noroozi, M. Marjanovic, A. Njegus, S. Escalera, and G. Anbarjafari, "Audio-Visual Emotion Recognition in Video Clips," IEEE Trans. Affect. Comput., vol. 10, no. 1, pp. 60–75, 2019, doi: 10.1109/TAFFC.2017.2713783.

[2] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," J. Big Data, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[3] B. J. Park, C. Yoon, E. H. Jang, and D. H. Kim, "Physiological signals and recognition of negative emotions," Int. Conf. Inf. Commun. Technol. Converg. ICT Converg. Technol. Lead. Fourth Ind. Revolution, ICTC 2017, vol. 2017-Decem, pp. 1074–1076, 2017, doi: 10.1109/ICTC.2017.8190858.

[4] R. M. Mehmood, R. Du, and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors," IEEE Access, vol. 5, no. c, pp. 14797–14806, 2017, doi: 10.1109/ACCESS.2017.2724555.

[5] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," IEEE Trans. Affect. Comput., vol. 13, no. 3, pp. 1195–1215, 2022, doi: 10.1109/TAFFC.2020.2981446.

[6] X. Wang, Y. Zhao, and F. Pourpanah, "Recent advances in deep learning," Int. J. Mach. Learn. Cybern., vol. 11, no. 4, pp. 747–750, 2020, doi: 10.1007/s13042-020-01096-5.

[7] E. Owusu, J. A. Kumi, and J. K. Appati, "On Facial Expression Recognition Benchmarks," Appl. Comput. Intell. Soft Comput., vol. 2021, 2021, doi: 10.1155/2021/9917246.

[8] L. Chen, M. Zhou, W. Su, M. Wu, J. She, and K. Hirota, "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction," Inf. Sci. (Ny)., vol. 428, pp. 49–61, 2018, doi: 10.1016/j.ins.2017.10.044.

[9] A. Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a Real-Time Emotion Recognition System Using Facial Expressions and EEG based on machine learning and deep neural network methods," Informatics Med. Unlocked, vol. 20, p. 100372, 2020, doi: 10.1016/j.imu.2020.100372.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 8s
DOI: https://doi.org/10.17762/ijritcc.v11i8s.7636
***Article Received**: 25 April 2023 **Revised:** 24 June 2023 **Accepted:** 10 July 2023*

[10] C. Tan, M. Šarlija, and N. Kasabov, "NeuroSense: Short-term emotion recognition and understanding based on spiking neural network modelling of spatio-temporal EEG patterns," Neurocomputing, vol. 434, pp. 137–148, 2021, doi: 10.1016/j.neucom.2020.12.098.

[11] V. Sati, S. M. Sánchez, N. Shoeibi, A. Arora, and J. M. Corchado, "Face detection and recognition, face emotion recognition through nvidia jetson nano," Adv. Intell. Syst. Comput., vol. 1239 AISC, no. September, pp. 177–185, 2021, doi: 10.1007/978-3-030-58356-9_18.

[12] O. Ekundayo and S. Viriri, "Multilabel convolution neural network for facial expression recognition and ordinal intensity estimation," PeerJ Comput. Sci., vol. 7, no. Cv, 2021, doi: 10.7717/peerj-cs.736.

[13] Ezenwobodo and S. Samuel, "International Journal of Research Publication and Reviews," Int. J. Res. Publ. Rev., vol. 04, no. 01, pp. 1806–1812, 2022, doi: 10.55248/gengpi.2023.4149.

[14] Y. Nan, J. Ju, Q. Hua, H. Zhang, and B. Wang, "A-MobileNet: An approach of facial expression recognition," Alexandria Eng. J., vol. 61, no. 6, pp. 4435–4444, 2022, doi: 10.1016/j.aej.2021.09.066.

[15] A. Michele, V. Colin, and D. D. Santika, "Mobilenet convolutional neural networks and support vector machines for palmprint recognition," Procedia Comput. Sci., vol. 157, pp. 110–117, 2019, doi: 10.1016/j.procs.2019.08.147.

[16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, vol. 2017-Janua, pp. 2261–2269, 2017, doi: 10.1109/CVPR.2017.243.

[17] N. Hasan, Y. Bao, A. Shawon, and Y. Huang, "DenseNet Convolutional Neural Networks Application for Predicting COVID-19 Using CT Image," SN Comput. Sci., vol. 2, no. 5, 2021, doi: 10.1007/s42979-021-00782-7.

[18] S. H. Wang and Y. D. Zhang, "DenseNet-201-Based Deep Neural Network with Composite Learning Factor and Precomputation for Multiple Sclerosis Classification," ACM Trans. Multimed. Comput. Commun. Appl., vol. 16, no. 2s, 2020, doi: 10.1145/3341095.

[19] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood components analysis," Adv. Neural Inf. Process. Syst., 2005.

[20] Y. Chen and J. He, "Deep Learning-Based Emotion Detection," J. Comput. Commun., vol. 10, no. 02, pp. 57–71, 2022, doi: 10.4236/jcc.2022.102005.

[21] E. G. Moung, C. C. Wooi, M. M. Sufian, C. K. On, and J. A. Dargham, "Ensemble-based face expression recognition approach for image sentiment analysis," Int. J. Electr. Comput. Eng., vol. 12, no. 3, pp. 2588–2600, 2022, doi: 10.11591/ijece.v12i3.pp2588-2600.

[22] C. Jia, C. L. Li, and Z. Ying, "Facial expression recognition based on the ensemble learning of CNNs," ICSPCC 2020 - IEEE Int. Conf. Signal Process. Commun. Comput. Proc., pp. 0–4, 2020, doi: 10.1109/ICSPCC50002.2020.9259543.

[23] S. Gu, C. Xu, and B. Feng, "Facial expression recognition based on global and local feature fusion with CNNs," 2019 IEEE Int. Conf. Signal Process. Commun. Comput. ICSPCC 2019, pp. 5–9, 2019, doi: 10.1109/ICSPCC46631.2019.8960765.