# FER-Net: Leveraging Attention Mechanisms in EfficintNetB5 model for Enhanced Emotion Recognition in Augmented Images

Balla Charishma Sulochana, Bhavya Sri Pragada, Gaddam Anvith Reddy, *Rimjhim Padam Singh, Suja Palaniswamy

*Department of Computer Science and Engineering,*
*Amrita School of Computing, Bengaluru,*
*Amrita Vishwa Vidyapeetham, India*

charisma.sulochana@gmail.com, bhavyasri.pragada@gmail.com, ganvithreddy3052@gmail.com, *ps_rimjhim@blr.amrita.edu, p_suja@blr.mrita.edu

***Abstract*— Facial emotion recognition is a critical component in the development of advanced human-computer interaction systems, with applications spanning security, healthcare, and social robotics. This study explores the efficacy of advanced deep learning models, specifically EfficientNetB5 model in identifying emotions from facial images. Leveraging with augmented data, our approach integrates attention mechanisms, batch normalization, convolutional layers, global average pooling layers, and dropout techniques to enhance model performance. The proposed model demonstrated significant performance accuracy of 87.5%. , highlighting the importance of emphasizing relevant image regions for better emotion detection. Additionally, we conducted a comparative analysis with other prominent deep learning architectures, including AlexNet, XceptionNet, ResNet, MobileNet and EfficientNetV2M and proved the superiority of the proposed model against them.**

***Keywords - Deep Learning Algorithms, Emotion Recognition, Facial Expression Recognition, Attention mechanism***

## I. INTRODUCTION

Society has been embracing an increased use of computers in the modern world thus the use of chatbots as a mode of communication is encouraged. Chatbots these days are excellent in deciphering text messages and understanding it efficiently but when it comes to the emotional aspect of the user, they remain unreliable. Now with the help of chat-bots, it became possible to analyze the face of a client and make necessary correctives to suit their needs and wants.

Facial emotion recognition is a relevant area of computer vision and affective computing subject to more and more research. The field of interest is emotion recognition through the facial image. This technology has some great applications in many industrial areas such as HCI, security and mental health and E-gaming. Conventional Facial emotion recognition entailed the handcrafted features as well as classical machine learning techniques which in some cases were ineffective in addressing the increasing variation and intensity of human expressions. But with advancements in the field of deep learning it is like Facial emotion recognition has been turned true since models can learn hierarchical features directly from massive datasets which lead to improved accuracy and robustness.

There have been many advancements on how deep learning can be applied for the task of Facial emotion recognition and one of the greatest advancements is the introduction of an attention block into the model. Extra attention layers assist the network by searching for slightly detailed pixels that are irrelevant to a human eye but will aid the prediction of what type of emotion is expressed in a particular picture. They help the model focus on the parts of the face needed for decoding emotions and counteracting the problem of feature imbalance and uneven distribution of character features in the images. This improves the model's prediction as the interpretation of the decision is possible based on the attention maps where the decisions about the model prediction are highly dependent on the face-region features.

Hence, the paper proposes a significant advancement in Facial emotion recognition system with the employment of Efficient Net architectures. The major contributions of the paper are:

1. A novel Facial emotion recognition system, FER-Net, that extends EfficientNetB5 model with attention mechanisms and other fine-tuning mechanisms like batch normalization, dropouts, L2 regularization etc.

2. A comprehensive implementation and comprehensive analysis of five state-of-art deep learning models like AlexNet, MobileNet, XceptionNet, ResNet and EfficientNet models against the proposed approach.

## II. RELATED WORKS

SN Gourav et al [1] implemented a voting procedure enhancement system using mood sensor via face recognition. Their approach entails real-time facial expressions monitoring of voters through using the Python with TensorFlow and Keras utilizing OpenCV with machine learning algorithms. It uses face recognition for security and zero fraud and also has multiple data security and privacy features. Technologies that are suggested for use in system development are Express. js, Node. js, MongoDB, and React. js. One of the ways of evaluation is the user surveys and feedback that the team refers to as an ethical form of privacy because of its inherent consideration of the users' privacy.

Hill by Almeida Silva et al. [2] framework is the system for extrapolating facial expressions through the use of computer vision with CNN framework mostly. Examples within the methodology can be grouped into layers such as 'relu', 'kernel regularization', 'batch normalization', 'max pooling', 'dropout' and so on. Hill is used in such things like smart cars, games, monitoring health status, education, managing advertisement, recognition the emotions on the films – happy, sad, surprise, neutral, anger, disgust, fear. TensorFlow is mentioned for superior perform of AI training and Python as the very best programming language. The system combines emotion recognition AI with text and speech AI and is applied to communication with a chatbot for mental health care services.

Feng N. V et al. [3] showed the effectiveness of Convolutional Neural Networks (CNNs) for emotion recognition on the FER 2013 dataset based on facial expressions. Their hybrid training approach first involves a transcript corpus; later, the model is implemented based on user response feedback. The findings show the possibility for the development of emotion-enhanced chatbot systems that could improve the users' experience through greater expressiveness in the interactions. It is recommended to do further research on the deep learning models of emotion recognition and content suggestions models.

Kumar et al. [4] proposed an autonomous music selection system by using Facial Expression Recognition by using CNN and FER2013 dataset. The recognition rate of their model was 62.1%. The music recommendation algorithm performs matching via the cosine similarity algorithm against expressions and the music content. The study says older methods of machine learning such as random forest struggled to recognize facial expressions and the CNN. These findings support the findings of the need for better algorithms that help in improved emotion detection and further handling noise.

E. R. Vimina et al. [5] proposed the "Automated Reverse Turing Test via Facial Expressions" CAPTCHA system which uses AI face detection. Participants guess the emotions of people from a set of facial shots and score 95.82%. Such a high rate of efficiency is proved through the ample time-saving capacity of the system in comparison to other CAPTCHAs. This is useful in strengthening the security of the internet because it becomes hard for robotic account generators to pass the test and proves useful in a wide range of applications.

D. Lawrance et al. [6], whose work is called STLEV, uses LSTM and SNN with Triplet Loss for emotion recognition in videos. The elaborated method includes using the SNN algorithm to analyze the feature of emotion-related videos and LSTM to classify emotions, and the final average accuracy of the two algorithms reaches 87.5% accuracy on the BU-4DFE dataset and MI-4DFE dataset. This framework enhances the issue of deep learning models that need large training datasets. STLEV seems useful in cognitive human-computer interaction, with the recommendation that further efforts might concentrate on the domain of meta-learning and the process of model simplification.

Suja P et al. [7] proposed a complex technique MREAP for facial emotion recognition which was based on meta-learning technique using Siamese Networks. MREAP was able to give an accuracy of 80% when 80% of poses were changed in the dataset. The model at least exhibited discriminative ability since a series of different poses were given emotion predictions during testing. The distance measure method in Siamese network constructs a feature vector through the Euclidean distance. Further enhancements are to be made through balancing loss function, augmenting data, and testing meta-learning.

T. Keshari et al. [8] implemented a system which addressed emotion recognition by using a combination of facial expressions and upper body movements. Some of the feature-level used fusion schemes include concatenation and PCA with classifiers like ANN, SVM, and HHM. Quantitative assessment made within MATLAB demonstrated enhancement in performance due to combining multiple modalities. Further studies can be conducted to define the roles of deep learning and data augmentation for the assessment of eye movement under varied conditions.

Aly et al. [9] created a disability-based learning system using face expression recognition through ResNet-50 framework. The system achieved 87.62% accuracy on RAF-DB and 88.13% on FER2013. It achieved further accuracy by using a positional attention mechanism to pay attention to the critical regions. The research confirmed that the algorithm is useful as compared to standard literature and datasets used in recent publications and the contributions for the scientific world and online learning.

The automatic systems of sentiment analysis are related to machine learning-based systems, rule-based systems, and generative systems. The model proposed in [10] uses NLP to classify emotions and user behavior by stemming/ tokenizing. It should not be surprising that the chatbot model is suitable for various sectors – the effectiveness of the user engagement and the satisfaction related to emotions recognition and personalized service recommendations [10].

Perveen et al. [11] proposed recognition of facial micro-expressions based on the element's classification in conjunction with the multi-stream deep CNNs. After dimensionality reduction done using PCA and using architectures like ResNet-50, DenseNet-121 and VGG-16 the system is based on stacking method with random tree, J48 and random forest classifier. It performed well for the databases like CASME-II, CASME 2, SMIC and SAMM which is a necessary quality for real life emotion detection applications.

Y. Tang et al. [12] proposed the use of FreNet: a deep-learning architecture for frequency-domain analysis of facial expressions. The vectorization process used in FreNet extracts highly abstract representations in reduced length dimensions and summary layer. The model that was presented had faster processing speeds and less computational costs as compared to GoogleNet and other classical models. The trait of FreNet to perform feature learning and compact representation through dimension reduction can be used for HCl applications.

Anand et al. in [13] used EfficientNetB0 and WKELM in combination with Red Fox Optimizer for facial emotion recognition. The methodology includes image normalization and histogram equalization, achieving high accuracy on FER2013 (95.82%) and EMITIS datasets (96.98%). The model ranks better than transformers and ResNet-18. Future work can be extended to increase the size of the dataset and use selection methods to improve the accuracy of the classifier.

Deore et al. [14] proposed SongRec: facial emotion recognition for the recommendation of music using CNN. The one trained was associated with a facial dataset and the system matched the moods with appropriate songs and determined an accuracy of 62. 88% accuracy in real-time. Future enhancements include application of the fuzzy concept and use of MSE and RMSE to measure accuracy. It is possible to identify the real-life applications of Song Rec since the system bases its recommendation on the analysis of the facial expressions of the users.

Nair et al. [15] proposed a playlist recommendation system based on the sentiment analysis models, which is trained on Twitter data in an interactive manner. The system utilizes LSTM, Bidirectional LSTM, and 1-D CNN, achieving 79.29% correct classifications with Bidirectional LSTM. The model differentiates user's emotions into positive, negative, and neutral, to propose the relevant playlist/group based on the sentiment.

Bakariya et al. [16] presented a real time system which integrates facial emotion recognition with music recommendation via deep learning such as CNN. The validation accuracy of the models was also checked for Model-A and Model-B, and Model-B proves to be better than Model-A. The system proved its efficiency by indicating the ability to recognize facial emotions and make music recommendations; therefore, it has potential for technical use.

Kim et al. [17] used the children with PWS and ASD to assess emotional recognition. The ex post facto comparative design of the study was used to investigate emotional recognition, Theory of Mind, Working Memory as well as ASD traits. Disgust became a higher response in children with PWS compared to ASD children though the latter encountered social interaction difficulties. Such information has documented the need for establishing specific interventions for PWS and ASD children.

Lu et al. [18] proposed a system for real time recognition of human emotions, ages and genders. It identifies six emotions and a neutral emotion and classifies age into children, young adults and middle adults, and old adults. The algorithms such as NFC and BLAC used as pre-processing techniques positively influence recognition performance. The system achieved a mean accuracy rate of 69.15% for age recognition and 91.75% for gender recognition, demonstrating its effectiveness. Several other researchers have also thoroughly explored the problem having different challenges [19-21].

## III. Dataset Description

The Dataset Facial Emotion Recognition Image Dataset, derived from Kaggle contains a collection of 15,452 RGB photos capturing face expressions each connected with one of 6 distinct emotions : Smile, Anger, Sadness, Neutral, Surprise and Disgust. The images were gathered from social networks as Facebook and Instagram, from scrapping YouTube videos.

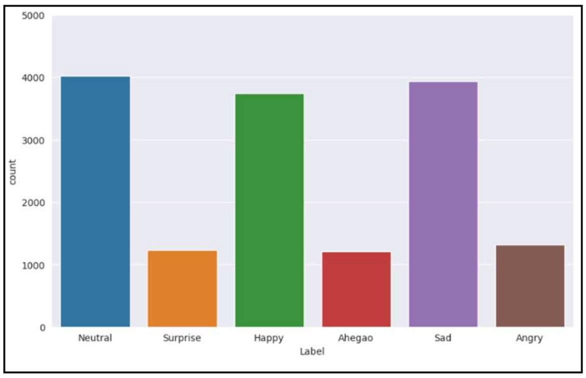


Fig 1: Bar graph representing no. of images from each class in the original dataset

The photographs include a diversity of individuals; different genders, different age groups, different cultures. The dataset is further subdivided into folders within the dataset. zip, with each folder corresponding to one of the six emotion classes. Additionally, data. csv has the emotion labels attached to the file paths of the photographs. One of the important features of the dataset is its imbalanced nature. The imbalance in the emotion class data can be seen in Figure 1.

## IV. Methodology

The complete system architecture from dataset loading to dataset augmentation and the proposed FER-Net model has been graphically presented in Figure 2.

### *A. Data Augmentation:*

Data augmentation is a manipulation of the original image by adding initialized transformations to the training images of the dataset for increasing the training samples. This makes the trained model more accurate as well as generalizable.

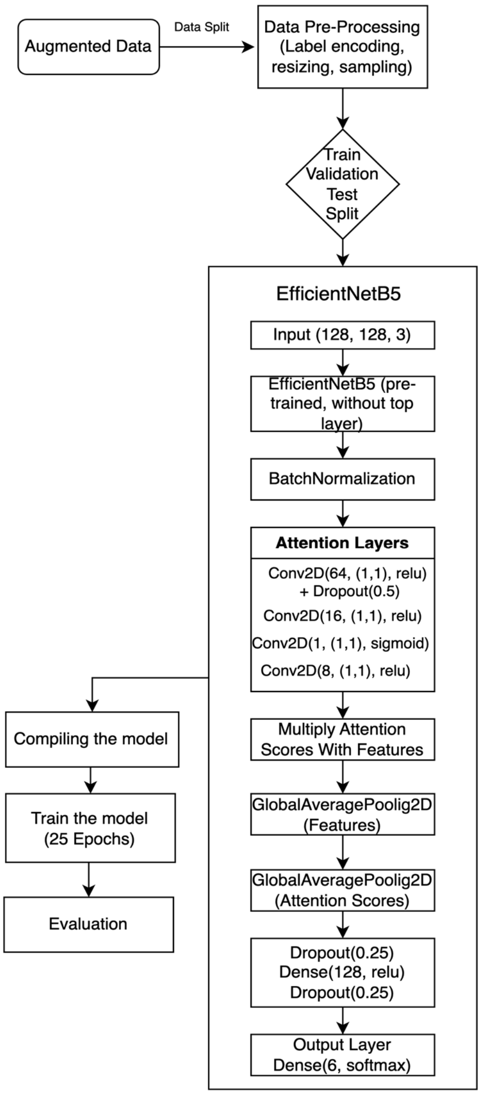


Fig. 2: Flow diagram of methodology

Data augmentation techniques such as rotation, width shift, height shift, shear range, zoom range and horizontal flip are performed and 4500 images from all the six classes are augmented until the desired values are reached and the classes become balanced

Finally, the source dataset and the synthetic dataset are merged into a single augmented dataset. This involves replicating the images from the two datasets into a new classification structure with several classes. The number of images in each class of the augmented dataset is represented in Figure 3.

### *B. Data Preprocessing:*

The created dataset is then further pre-processed to apply training to it. Images are resized into 128X128 pixels, and the labels identified for each are also converted through one hot encoding. Data mean and standard deviation are computed and subtracted from them from the original data. Finally, the preprocessed data are divided into 70% training, 15% validation and 15% test sets.

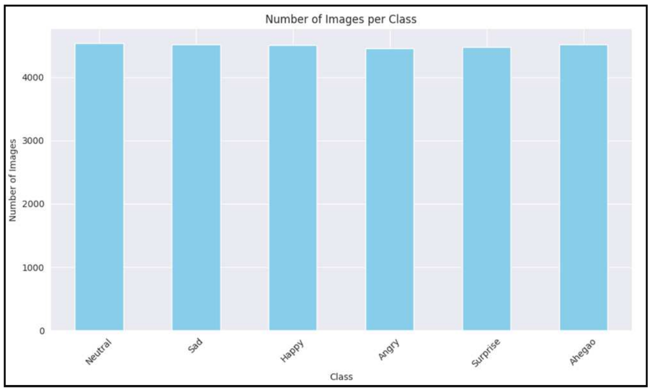


Fig 3: Bar graph representing no. of images from each class in augmented dataset.

### *C. Proposed FER-Net model*

The proposed FER-Net model focuses on leveraging the efficiency of the recent EfficientNetB5 model for facial emotion recognition in images. The model exploits the efficacy of deveral model enhancement techniques like attention mechanisms, dropout techniques, normalization techniques and regularization techniques to develop the final FER-Net model

The EfficientNetB5 [22] model pre-trained on ImageNet is loaded as a base model from the EfficientNet family of neural networks that has a greater parameter capacity to balance the load between the capacity of the data and the complexity of the model to achieve a substantially better accuracy and complexity trade-off for the image classification task. The Efficient Net models concludes that when scaling the network width, depth, and image resolution together as a single scaling group on the basic CNN backbone, computational efficiency in computer vision can be increased exponentially. This is because the Efficient Net is effective when real-time processing for devices that have low information systems resources that might need deployment to the devices.

The pre-trained EfficientNetB5 model is loaded and the data is then passed as input through the base model to extract the high-level efficient features. Batch normalization is used to normalize the feature that aids effectively in model training and learning process. This layer is then followed by an attention mechanism that is applied on the features with the help of the convolutional layers. A spatial attention module that convolves over the layers to compute attention scores that are element wise multiplied with the input feature maps to reduce noise and focus on the important aspects of the image. A self-attention layer is created by defining an object of the MultiHeadAttention class to capture dependencies between various units in the input feature maps. The output of the model features is passed to the global average pooling layer to generate global representations which are transformed to the shape needed by self-attention model. The spatial attention mechanism is then used to highlight and attend to the spatial features of the output. The attended output from spatial attention and the self-attention output are combined by multiplying the two using element-wise to incorporate the spatial and self-attention mechanism.

The attention scores are computed by the next convolutional layer which is followed by using a sigmoid activation function for normalizing the output values. The attention scores are then multiplied element wise with the batch normalized features to get the masked features. Global Average Pooling is applied separately to the masked features and the attention scores to get the global representations. The global average pooling is followed by scaling the features by dividing them by the corresponding attention scores to highlight the relevant features. A dense layer which has a ReLu activation function is applied for further model fine-tuning. A Dropout layer with the probability factor of 0.25 is used for reducing model over-fitting. The output layer is a dense layer with SoftMax with the last layer having 6 nodes that indicate the probability of the presence of each of the specified 6 classes.

## V. RESULTS AND ANALYSIS

For the model training and learning on the augmented FER dataset, it has been trained using categorical cross-entropy loss and Adam optimizer and accuracy as the evaluation metrics. The model is trained with the training data and 25 epochs along with the batch size of 64 and the validation data. The trained model is then evaluated for testing new data using the model to make predictions after the model is being evaluated. Statistics such as accuracy (Acc), precision (Pre), recall (Rec) and F1-score (F1) are calculated, and the confusion matrices and the classification reports and the plots of the accuracies and the losses are evaluated [24].

| Models | Acc | Pre | Rec | F1 | Loss |
|---|---|---|---|---|---|
| EfficientNetVM | 86.4 | 86.7 | 86.1 | 86.2 | 0.58 |
| XceptionNet | 85.0 | 86.2 | 85.1 | 85.1 | 0.66 |
| AlexNet | 67.2 | 71.7 | 67.3 | 67.8 | 1.29 |
| MobileNet | 82.0 | 82.5 | 82.0 | 81.9 | |
| ResNet50 | 78.9 | 79.9 | 78.9 | 78.9 | 1.06 |
| **Proposed FER-Net model** | **87.4** | **87.3** | **87.4** | **87.3** | **0.61** |

Table1: Evaluation metrics of each state-of-art model on augmented FER dataset

The final augmented dataset is trained using multiple models, namely, EfficientNet, AlexNet, MobileNet, XceptionNet and ResNet by fine-tuning each model with a standard fixed architecture. The fine-tuning architecture includes using Global Average Pooling, standard activation functions, adding dropout layers, performing batch normalization and L2 Regularization, which could improve the overall result with the above fixed set of parameters.

After every pre-trained model architecture there is a Global Average Pooling layer added, which reduces spatial dimensions of the feature map. The dense layers are fully connected with ReLu activation function. The Batch Normalization layers after dense layers help in normalizing the activation layer, there by accelerating the training process. At last, the output layer is activated using SoftMax activation function, to normalize the probability distribution of the output which helps in multi class classification of the 6 labels of different emotions.

The performance results obtained by different models fine-tuned using the above architecture are presented in Table 1. The proposed FER-Net model enhancing EfficientNetB5 model with attention mechanisms and other model enhancement techniques has secured an accuracy of 87.4% and F-score of 87.3%, which is the best model among all the trained models. It shows that the attention mechanism helped in improving the performance of the model. The EfficientNetVM and XceptionNet models performed competitively in terms of Precision, recall and F-score but still performed less efficiently as compared to the proposed model. On the other hand, ResNet50 and MobileNetv2 models performed decently with around 80% F-score but the simple AlexNet model performed poorly on the augmented FER dataset.

Hence, the results clearly claim the superiority of the proposed FER-Net integrating EfficientNet model with attention mechanisms.

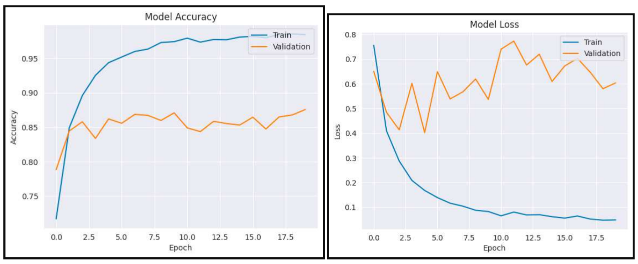


Fig 4 : Training and validation plots obtained by the proposed FERNet model Accuracy Plot (left) and Loss Plot (right)

Figure 4 presents the model training trends on training and validation data for the proposed FER-Net model. It can be seen that during the initial epochs, both accuracy and loss optimized steeply while for the latter epochs the plots became much more stable and got restricted into a fixed range. This clearly depicts a good model fit over the data for the proposed FER-Net model.

## VI. CONCLUSION AND FUTUREWORKS

As a result of applying both self-attention and spatial attention in the proposed FER-Net model, the work was able to increase the localization information of the images thereby, leading to the highest performance metrics. In light of the work conducted, it is clear that pre-trained model EfficientNetB5 classification model can be successfully employed with the inclusion of attention mechanism and model enhancement techniques like dropout, normalization and regularization mechanisms for enhanced image classification for facial emotion recognition. All other models implemeted and tested on FER dataset for classifying the slight emotional signals could not outperform the proposed architecture.

In further studies there are several approaches that could be used as future directions for the improvement of emotional recognition systems. Furthermore, the potential to enhance the data with the other sensory indicators (audio or text) might contribute to the development of more enhanced systems for emotion recognition. It is evident that such models have significant utility potential to be further developed and translated for use in such practical functions as HCI, personal psychology, interactive media. The results that have been achieved in this project provide the right grounds for further developing highly efficient, accurate, and real-time emotion-recognition systems in other spheres and fields of application to further enhance and advance human-computer interactions.

## References


[1] S. N. Gourav, K. Arora, S. Khan, A. Dulat and S. Sinha, "Facial Recognition-Based Emotion Detection for Enhanced Voter Experience," 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Gautam Buddha Nagar, India, 2023, pp. 75-81, doi:10.1109/UPCON59197.2023.10434600.

[2] T. T. Almeida Silva, A. O. Andrade and N. P. da Silva, "System for Recognition of Facial Expressions Using Machine Learning," 2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE), Natal, Brazil, 2020, pp. 1-6, doi:10.1109/LARS/SBR/WRE51543.2020.9307127.

[3] N. V, R. G B, N. L. B, S. Kumar C P, R. S. Reddy A and P. J, "Chatbot Based on Emotions Using Deep Learning," 2023 1st International Conference on Optimization Techniques for Learning (ICOTL), Bengaluru, India, 2023, pp. 1-5, doi:10.1109/ICOTL59758.2023.10435281

[4] S. Kumar, R. K. Nandi, S. Roy and A. Sarkar, "Design and Implementation of AI Based Efficient Emotion Detection and Music Recommendation System," 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, 2023, pp. 1-5, doi: 10.1109/ICCCNT56998.2023.10306938.

[5] E. R. Vimina, "An Automated Reverse Turing Test Using Facial Expressions," 2009 2nd Conference on Human System Interactions, Catania, Italy, 2009, pp. 314-317, doi: 10.1109/HSI.2009.5090998.

[6] D. Lawrance and S. Palaniswamy, "Emotion recognition from facial expressions for 3D videos using siamese network," 2021 International Conference on Communication, Control and Information Sciences (ICCISc), Idukki, India, 2021, pp. 1-6, doi: 10.1109/ICCISc52257.2021.9484949.

[7] S. Ramakrishnan, N. Upadhyay, P. Das, R. Achar, S. Palaniswamy and A. A. Nippun Kumaar, "Emotion Recognition from Facial Expressions using Images with Arbitrary Poses using Siamese Network," 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2021, pp. 268-273, doi: 10.1109/ICOSEC51865.2021.9591919.

[8] T. Keshari and S. Palaniswamy, "Emotion Recognition Using Feature-level Fusion of Facial Expressions and Body Gestures," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1184-1189, doi: 10.1109/ICCES45898.2019.9002175.

[9] Aly, Mohammed & Ghallab, Abdullatif & Fathi, Islam. (2023). Enhancing Facial Expression Recognition System in Online Learning Context Using Efficient Deep Learning Model. IEEE Access. 11. 10.1109/ACCESS.2023.3325407.

[10] Balaji, M. & N., Yuvaraj. (2019). Intelligent Chatbot Model to Enhance the Emotion Detection in social media using Bi-directional Recurrent Neural Network. Journal of Physics: Conference Series. 1362. 012039. 10.1088/1742-6596/1362/1/012039.

[11] Perveen, Gulnaz & Ali, Syed & Ahmad, Jameel & Shahab, Sana & Adnan, Muhammad & Anjum, Mohd & Khosa, Ikramullah. (2023). "Multi-Stream Deep Convolution Neural Network With Ensemble Learning for Facial Micro-Expression Recognition." IEEE Access. PP. 1-1. 10.1109/ACCESS.2023.3325108.

[12] Y. Tang, X. Zhang, X. Hu, S. Wang and H. Wang, "Facial Expression Recognition Using Frequency Neural Network," in IEEE Transactions on Image Processing, vol. 30, pp. 444-457, 2021, doi: 10.1109/TIP.2020.3037467.

[13] Anand, M. & Babu, S.. (2024). Multi-class Facial Emotion Expression Identification Using DL-Based Feature Extraction with Classification Models. International Journal of Computational Intelligence Systems. 17. 10.1007/s44196-024-00406-x.

14] Shalaka Prasad Deore, "SongRec: A Facial Expression Recognition System for Song Recommendation using CNN." International Journal of Performability Engineering, vol. 19, no. 2, Feb. 2023, pp. 115-121.

[15] A. Nair, S. Pillai, G. S. Nair and A. T, "Emotion Based Music Playlist Recommendation System using Interactive Chatbot," 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatre, India, 2021, pp. 1767-1772, doi:10.1109/ICCES51350.2021.9489138.

[16] Bakariya, Brijesh & Singh, Arshdeep & Singh, Harmanpreet & Raju, Pankaj & Rajpoot, Rohit & Mohbey, K..(2023). Facial emotion recognition and music recommendation system using CNN-based deep learning techniques. Evolving Systems. 15. 1-18. 10.1007/s12530-023-09506-z.

[17] Oh, S.; Kim, D.-K. Comparative Analysis of Emotion Classification Based on Facial Expression and Physiological Signals Using Deep Learning. Appl. Sci. 2022, 12, 1286. https://doi.org/10.3390/app12031286

[18] Lu, Ta-Te & Yeh, Sheng-Cheng & Wang, Chia-Hui & Wei, Min-Rou. (2021). Cost-effective Real-time Recognition for Human Emotion-Age-Gender Using Deep Learning with Normalized Facial Cropping Preprocess. Multimedia Tools and Applications. 80. 10.1007/s11042-021-10673-x.

[19] K. Soumya and S. Palaniswamy, "Emotion Recognition from Partially Occluded Facial Images using Prototypical Networks," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 491-497, doi: 10.1109/ICIMIA48430.2020.9074962.

[20] Reddy, R.T., Singh, R.P. and Kumar, P., 2023, December. Gender Recognition using Central Difference Convolution in AlexNet with Haar Cascades. In *2023 IEEE 20th India Council International Conference (INDICON)* (pp. 1404-1409). IEEE.

[21] C. Vaishnavi and S. Palaniswamy, "Emotion Recognition from Online Classroom Videos Using Meta Learning," 2022 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC), Bhubaneswar, India, 2022, pp. 1-6, doi: 10.1109/ASSIC55218.2022.10088292.

[22] Sruthi, S., Emadaboina, S., Machavarapu, P., Singh, R.P. and Kanchan, S., 2024, April. Covid-19 Classification using Fine-tuned EfficientNet Architecture. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)* (pp. 1-6). IEEE.

[23] Singh, R.P., Sharma, P. and Madarkar, J., 2019. Compute-extensive background subtraction for efficient ghost suppression. *IEEE Access*, *7*, pp.130180-130196.