Cognitive Resonance in Emotional Decoding: A CNN-ResNetB0 Synergy

Vidya P
Computer Science and Engineering
REVA University
Bengaluru, India
R20200579.VIDYAP@cse.reva.edu.in

Sharon Chattopadhyay
Computer Science and Engineering
REVA University
Bengaluru, India
2000069@reva.edu.in

Bhavatarini N
Computer Science and Engineering
REVA University
Bengaluru, India
bhavatarini.n@reva.edu.in

***Abstract*—Emotion recognition from facial expressions has gained significant attention for its abeyant applications in human-computer synergy and affective computing. In this research paper, we extant a human face emotion detection model planted on Convolutional Neural Networks (CNNs). Our model weights the power of CNNs to automatically learn and excerpt discriminative lineaments from facial portraits. We introduce a comprehensive dataset containing a diverse range of emotions, enabling robust model training and evaluation. Additionally, we explore the interpretability of the model by visualizing activation maps to identify the regions of the face contributing most to emotion classification. The proposed model's efficiency and real-time applicability make it a promising candidate for emotion-aware systems in real-world scenarios. This work offers a robust and interpretable solution for human face emotion detection, utilizing Convolutional Neural Networks, with high accuracy and potential applications in domains where understanding human emotions is crucial.**

Keywords—CNN, facial expressions, emotion recognition, emotion detection

I. INTRODUCTION

Convolutional Neural Networks (CNNs) have revolutionized emotion recognition by using deep learning to automatically interpret facial expressions. These models analyze fine-grained details, grasping nuanced and overt signals for accurate emotion classification. Trained on diverse datasets, CNNs performs well in recognizing distinct patterns associated with various emotions, enhancing adaptability and accuracy. Their hierarchical structure enables effective feature extraction, making them proficient in real-time applications across diverse lighting conditions. CNN-based emotion detection models have extensive effects, spanning human-computer interaction, entertainment, and psychological research. This introduction explores how CNNs have transformed emotion recognition, providing insights into the complex domain of human emotional expression. The impact of CNNs on emotion recognition is highlighted, emphasizing their role in the significant transformation of detecting emotions from facial expressions. These models, trained on diverse datasets, enable automatic emotion detection, offering valuable insights into human emotional expression. The exploration extends to various domains, including collaborated computing, entertainment, and psychological research, showcasing the flexibility and applications of CNNs in understanding and responding to human emotions.

In present face emotion detection CNN models include bias due to limited diversity in training data, overfitting, challenges in recognizing complex emotions, real-time processing demands, sensitivity to environmental factors, and ethical concerns regarding privacy. We implemented ResNetB0 Synergy in associated with CNN, which enhances the ability of human-computer interaction in efficient manner.

The significance of human interaction and communication through facial expressions highlighted by Vaishnavi Hosur et al [1], emphasizing that a substantial portion of communication relies on nonverbal cues, particularly emotions conveyed through facial expressions. The study aims to use artificial intelligence to read facial expressions accurately.

The iCV MEFED (Multi-Emotion Facial Expression Dossier-set) utilized by Sabrina Begaj et al [2] as major dossier-set and discusses various methods and approaches in the range of facial expression recognition. The paper evaluates different techniques, including traditional approaches and deep learning-based methods, and provides insights into their performance and limitations.

The mechanized anterior asseveration tumble growing field discussed by D Y Liliana [3] in computer eyesight and intelligent retrieval, with applications in entertainment, erudition, e-commerce, health, and salvation. The paper focuses on detecting AUs using deep CNNs, with a specific emphasis on reducing overfitting using dropout.

A deep learning-based fore melancholy revelation system using CNNs by Akriti et al presents [4]. The contemplated model is designed for emotion detection from facial enunciations and is evaluated on two dossier-sets: Fore Vehemence Avowal Ultimatum and Japanese Muliebrious Anterior Remorse. Three main steps of the emotion detection process discussed in the paper are face unearthing, feature uprooting, and ardor classification.

The application of deep learning architectures discussed by Hari et al proposed [5] for remorse detection in human faces. The study focuses on three key facets: pre-depuration, extracting visages, and echelon knacks. The authors propose a collating of deep learning architectures (VGG-16, ResNet152V2, InceptionV3, and Xception) using relocation erudition from pre-trained models.

A robust real-time emotion detection system discussedby Shruti et al [6] adopting a Convolutional Neural Network (CNNs) architecture. Motivation behind research is the growing need for robots to understand human emotions for effective human-robot interaction, especially in fields like elderly care, treating autistic patients, and child therapeutics. The authors emphasize the importance of emotion understanding for personal robots to drudge in a more customized manner and gain the confidence of individuals. The employment of disturbances in human-to-human communication is essential, as a discourse can derive inspiration from upright emotions.

Fore Tremor a non-verbatim contours revelation led by Prince et al [7], Used a cultural collection in an unchangeable conversation.Vogue's identical quick for examination source cut powerful work. Copy aware human regret sharpened steps illustration. A group speculates splendidly, jokingly determined an exoneration. Hiding bravely, the team starts with consciences, taking a bold step, refining the final work.

A detailed illustration and examination of emotional disturbance patterns, archetypes, and data sets discussed by Anvita et al. are presented in [8]. The content includes different methods like rugged set theory combined with SVM, analysis of depression detection methods using speech signals, stationary wavelet transformation, and statistical features combined with various approaches.

The use of machine learning algorithms, particularly deep learning and K-means clustering, for emotion detection in images. It outlines the steps involved in training the public face database, extracting practicalities for each enclosure of front, aggregating single-hem plausibility, and classifying images using a support vector machine. The document also emphasizes the need for a capacious quantity of test data and a high-performance computer for accurate results.

Facial expression recognition techniques relating to Ninad [10], focusing on distinguishing expressions using explicit classifiers and characterizing them based on extracted facial features. It also details the hardware and software used for the experiments, along with the results and limitations of the proposed method.

Commercial facial emotion recognition (FER) systems developed by Eugenia et al [11] show age bias, more accurately detecting facial expressions in images of younger individuals than in images of older ones. Bent lamp 4 commercial FER systems evaluated in the study, and was consistent across both 2019 and 2020.

A method proposed by Allen et al [12] for foremost emotion detection using a mutated eye map–mouth map algorithm on an enhanced idol and classification with TensorFlow.

Moutan et al [13] suggested a technique to assess a learner's level of awareness in an online learning system using upfront enthusiasm detection. Author argues that this can help to improve the learning experience by allowing teachers to adapt their teaching strategies in real time to the needs of the learners.

Syed Aley Fatima et al [14] introduced a technique for real-time detection of human sadness using the Mini-Xception framework. The Mini-Xception model is a lighter version of the Xception architecture, a deep convolutional neural network (CNNs) design.

The model introduced by Dasharath et al [15] combined with algorithms will work more efficiently then a single algorithm alone.

The paper is orderly in as follows:
Firstly, the introduction to the topic then followed by second the literature survey, next the methodology followed by experimental analysis and finally ending with the result and conclusion.

II. LITERATURE SURVEY

The Hosur, V., & Desai, A. R. paper [1] conducts a survey to provide insights into previous work and developments related to facial emotion recognition. They reference various studies and techniques, including the work of Paul Ekman and Friesen in defining six basic emotions, the importance of emotion recognition in fields like biomedical engineering and psychology, and the relevance of emotion detection for security applications. The survey establishes the foundation for their research.

The paper [2] discusses different approaches to Facial Expression Recognition (FER) over the past two decades. These approaches are broadly categorized into two main adjustments. The choice of approach depends on the implements and the data being processed. Conventional approaches involve manual feature engineering, which requires preprocessing the image, feature extraction, and expression classification. Deep Learning approaches combine these steps, making it suitable for challenging tasks. Common datasets for FER include CK+, MMI, MPI, JAFFE, and KDEF. Data augmentation techniques are applied to improve recognition accuracy.

Two approaches were discussed from author D Y Liliana paper [3] for facial expression cognizance are AU detection and facial point unearthing. FACS quantifies facial expressions by observing changes in facial muscles, specifically 44 AUs. Deep learning fashions have been contemplated to address ventral semantic ingredient recognition stints. Feature extraction is a crucial step in anterior expression recognition, affecting the quality of results. The paper highlights the use of Convolutional Neural Networks (CNNs) whereas feature evocation and classification. Various facial expression recognition methods use congruous-feature based, actualization-feature based, texture-based, or crossbreed features. CNNs, has gained popularity in facial expression recognition, and this research introduces dropout as a regularization mechanism.

The paper [4] surveys and provides insights into different algorithms, overcoming legacy methods by incorporating new approaches in algorithmic techniques. In CNN architecture, they extracted guidance image datasets, then integrated guidance cues into the convolutional mantle around the input data, ultimately enhancing the accuracy and overall similarity of two similar models during the learning process.

The paper [5] sheds light on the comparison of deep learning frameworks using Keras enthusiasm detection, utilizing deep facial features similarity through transfer learning from well-prepared images on VGG 16, Res Net 152V2, Inception V3, and Xception. This process generates characteristic features for input images.

The paper [6] introduces a network based on convolutional neural network principles, incorporating 909 criteria from Vanilla CNNs and 509 from recent cutting-edge research, surpassing the current standards. The network is thoroughly tested on eight diverse datasets, including Fer 2013, CK and CK+, Chicago Face Database, JAFFE Scoop-set, FEI face score-set, IMF DB, TFE ID abstracts-set, and a practical lab dataset, covering various angles, perspectives, backgrounds, and age groups.

This paper [7] clearly explores contemporary deep learning structures and algorithms for facial emotion detection and recognition. The paper highlights the dominance of CNN architecture over others, such as RNNs and SVMs, emphasizing legacy, exemplary performance, and challenges in modernization. It also identifies potential opportunities and ongoing debates for future FER research. The paper recognizes the importance of assessing existing resources and acknowledges the need for extensive facial emotion datasets to drive progress.

This paper [9] incorporates the Viola-Jones algorithm to detect the features and contours relevant to facial expressions. Additionally, it utilizes Machine Learning techniques, Deep Learning models, and Neural network algorithms for emotion recognition. The paper identifies emotions based on facial features, capturing nuances of expression and viewpoint. It specifically recognizes seven emotions: embarrassment, malice, nausea, abhorrence, happiness, sadness, and surprise, by analyzing the facial expressions of individuals.

The paper [10] outlines a FERC model based on a two-part convolutional neural network. The first part focuses on capturing the essence for facial engraving, while the second part delves into the extraction of facial feature lines. Utilizing an explanatory vector in the FERC model, it identifies five different types of consistent facial expressions. The model was tested on a comprehensive database of 10,000 images from 154 individuals, achieving a notable 96% accuracy in correctly classifying emotions, with an EV (Error Vector) of 24 values.

The paper [11] highlights that FER reveals both gender and racial biases, unlike other prominent analysis methods. Consequently, research efforts in emotion recognition have primarily concentrated on addressing gender and racial disparities, while considerations related to age and other mathematical biases have received less attention. This study aims to examine the performance of modern FER technology on diverse images of men and women from three distinct groups. It employed four different FER algorithms in an opaque box fashion to accurately identify six emotions—joy, disgust, fear, excitement, neutrality, and sadness.

In the study of [12], facial expression recognition concerning changes in facial features is presented. Initially, the model enhances image quality through isolated curls and soft variety. The geometry is adjusted, creating a visual map and an input map with landmarks. Finally, spatial angles form triangles, demonstrating the neural network's effectiveness using tensor flow in an intermediate processing stage. The results suggest that the proposed method is suitable for detecting facial expressions.

The paper [13] discloses that, in a practical scenario, a simulated set of four fundamental human emotions was created for novice learning. Various emotions, problem continuum in facial expressions, and numerous faces were combined. CNNs were designed to identify basic emotions and locate challenges faced by beginners. The causes were verified through algebraic methods.

The paper [14] analyzed that by validating a paradigmatic real-time framework, it efficiently accomplishes tasks such as facial unmasking and emotion categorization simultaneously in a unified step, employing the provided mini-Xception architecture. Through a discriminative approach, the study utilized the FER-2013 dataset, successfully handling challenges like identification and classification of seven different emotions using the Mini-Xception model, achieving an overall accuracy of 95.60%.

The paper by Bhadangkar, D., Pujari, J., & Yakkundimath, R. [15] reveals that by accurately detecting facial features, the subconscious well-being of individuals can be monitored and improved, thereby preventing impulsive actions. They employ machine learning methods for sentiment extraction and classification to perform emotion detection solely based on facial expressions. In this study, they achieve precision levels up to 0.9042 using a combination of Latent Dirichlet Allocation and Support Vector Machines.

III. METHODOLOGY

A. Proposed System

The strategy implemented in the project involves training the CNN model with a variety of images that express certain emotions. The primary emotions considered include joy, sadness, anger, disgust, surprise, and neutrality. The dataset comprises images of facial expressions corresponding to these emotions, which are fed into the CNN model for training. The process utilizes integrated data processing modules such as OpenCV, Keras, and Tensor Flow. Figure 1 illustrates the architectural layout of the implementation. The specific emotion recognition model employs Convolutional Neural Networks (CNNs) to distinguish emotions from facial features. It encompasses the training of CNN on a diverse dataset containing labeled facial portraits depicting various emotions.

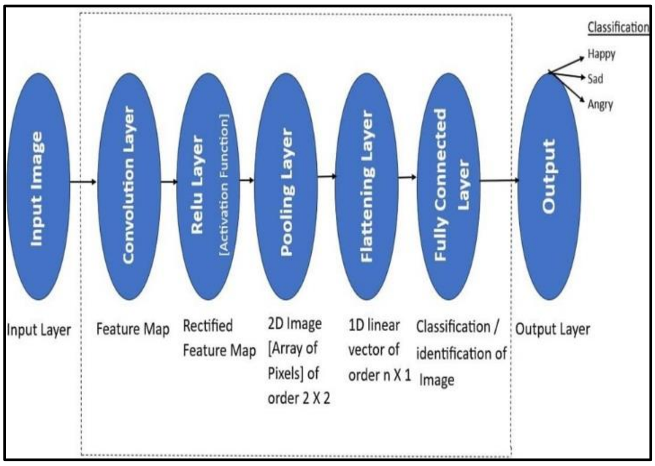

Fig. 1. Model builiding for the processed model

A Convolutional Neural Network (CNN) comprises several key layers as shown in Fig 1 which collectively enable effective image recognition and classification. The journey begins with the Input Layer, which receives raw data, typically pixel values for images. Subsequently, Convolutional Layers play a pivotal role by applying filters to detect intricate patterns such as edges and textures. Following this, the Rectified Linear Unit (ReLU) Layer introduces non-linearity to the model, enhancing its capacity to capture complex relationships within the data which provides a rectified feature map. Pooling Layers come into play to reduce dimensionality of image pixel, employing operations like max pooling to distill essential features. The Flattening Layer will convert the processed 2D matrix data into linear vector, preparing it for subsequent analysis. Fully Connected Layers establish comprehensive connections between neurons, often serving as the final stage for classification tasks such as happy, sad, angry as shown in Fig 1. The Output Layer produces the network's final output, its neurons aligning with the classes in classification scenarios.

B. Advantages of Proposed model

1. Leveraging CNNs allows for accurate detection of a wide range of emotions from facial expressions.
2. Real-time processing enables instant emotional feedback, enhancing human-computer interaction.
3. CNN-based models can adapt to varying lighting conditions, expressions, and cultural nuances.
4. Enables non-intrusive collection of emotional responses for market research and user experience testing.
5. Supports systems that adjust responses based on detected emotions, enhancing user satisfaction.
6. Fosters a more empathetic and emotionally aware interaction between humans and technology.

IV. EXPERIMENTAL ANALYSIS

In experimental analysis, ensuing steps:

- Data Collection
- Data Preprocessing
- Model building
- Testing and training of model
- Generating output

Data Collection- For the face emotion detection model involved curating a diverse dataset of facial images, encompassing a wide range of emotions. Rigorous quality control ensured consistency and minimized bias. Data augmentation techniques like cropping, rotation, and flipping expanded the dataset. Special attention was given to balance the classes to prevent bias towards prevalent emotions. The collected data provided a comprehensive foundation for training the Convolutional Neural Network (CNN), enabling the archetypal to imbibe robust and generalized facet for emotion recognition.

Data Preprocessing- Keras is an open-source colossal-akin deep learning athenaeum authored in Python. It provides a user-friendly collocation for building, training, and deploying tactile networks, making it easier for researchers and practitioners to work with complex deep learning architectures. Keras was designed with user friendliness, modularity, and extensibility in mind.

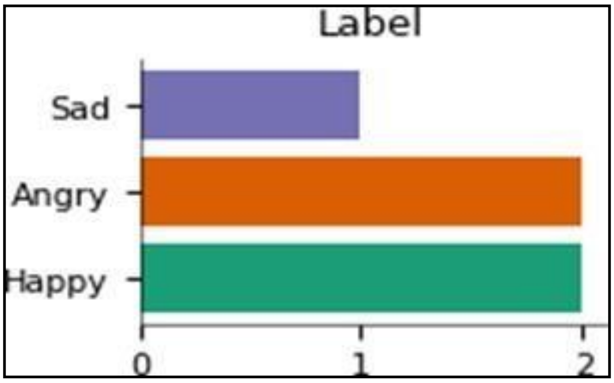

Fig. 2. Labels for the processed model.

We have used CNN for sadness, anger, and happiness as shown in Fig 2 for classification of facial expressions through Convolutional Layers capturing emotion-specific patterns. The network employs ReLU activation for non-linearity and concludes with a Softmax Output Layer, efficiently categorizing the input into one of the three emotions. This streamlined architecture excels in discerning emotional nuances through convolution, activation, and classification stages.

Model building-

Here are the main algorithmic elements involved in a face emotion detection model using CNNs:

1. Convolutional Layers: These seams use convolutional filters to scour the input image and decoction of local features, helping the model detect patterns like edges, textures, and shapes that are important for recognizing facial expressions.
2. Fully Connected (Dense) Layers: These stratum are often used at the point of the CNN design to perform classification. They connect each one neuron from the antecedent layer to all neuron in the current mantle, culminating in the final output couch representing emotion classes.
3. Activation Functions: Activation functions like ReLU (Rectified Linear Activation) introduce non-linearity, enabling the network to capture complex affinity in the data and expressions.
4. Dropout: Dropout layers help prevent overfitting by haphazardly dropping out a fragment of neurons midst tutelage, permeating the nexus to learn more stout and epitomized savors.
5. Loss Functions: Emphatic cross-entropy is commonly pre-owned as the loss employment for codification tasks, including emotion detection. It quantifies the difference between predicted and actual emotion labels.
6. Optimization Algorithms: Algorithms like Adam, RMSProp, or SGD are used to update the weights of the network during training, minimizing the loss function.

7. Data Augmentation: Although not strictly an algorithm, data accession techniques like gyration, ascent, and reveling are applied to artificially expand the drilling dataset and embroider the model's ability to generalize.
8. Preprocessing Steps: Grayscale conversion, histogram equalization, and face detection algorithms like Haar cascades are often applied to preprocess the input images before feeding them into the CNN. While these are the foundational algorithmic elements, the effectiveness of a face emotion detection model also depends on the architecture design, dataset quality, hyperparameter tuning, and the overall training process.

Testing and training of model -

The algorithm learns to map input data to consonanting output labels based on the provided examples. In the context of CNNs for face emotion detection, a labeled dataset of facial effigies along with their associated emotion epithets (e.g., happy, sad, angry) is used to train the network. During training, the CNN adjusts its internal criterions (adiposity and penchants) to trivialize the antithesis interpolated its previsions and the true emotion labels. Once trained, the CNN can then predict the emotion label of new, unseen facial images.

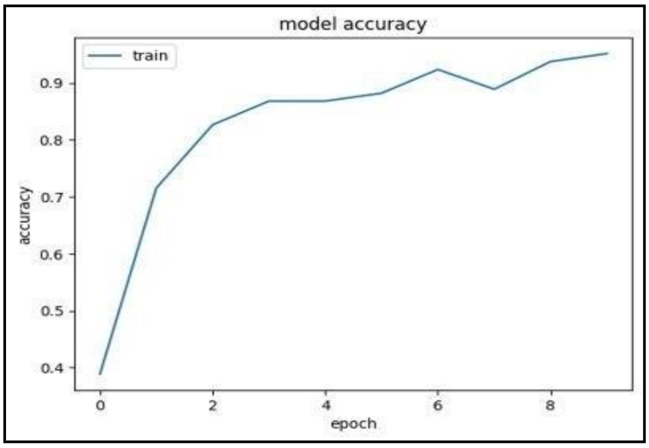

Fig. 3. Model accuracy generated by the processed model.

The model's accuracy starts to be upwards as shown in Fig 3, as soon as training is infused and the drop seems to be there at around epoch 7 and the upward continuum follows.

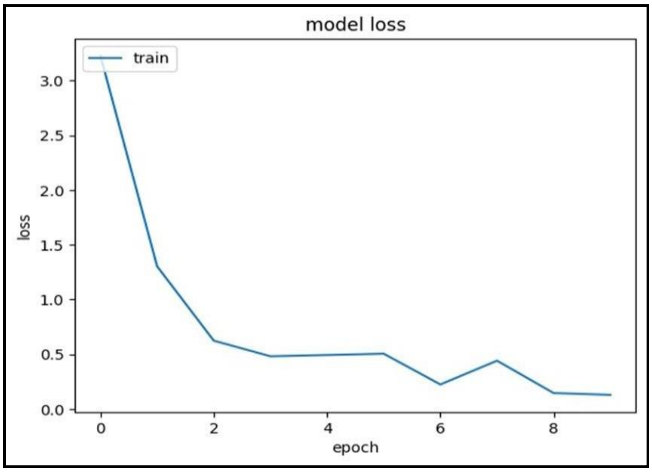

Fig. 4. Model loss generated by the processed model.

The model's loss dips to downwards as depicted in Fig 4 as soon as training is infused and the seems to be swinging upward at around epoch 7 and the downward continuum follows.

A. Result

The proposed model generates the output with 98% accuracy with input layers of shape [224, 224, 3] using the GlobalMaxPooling2D for pooling and softmax for the output layers and cross entropy is calculated using the Adam optimizer.

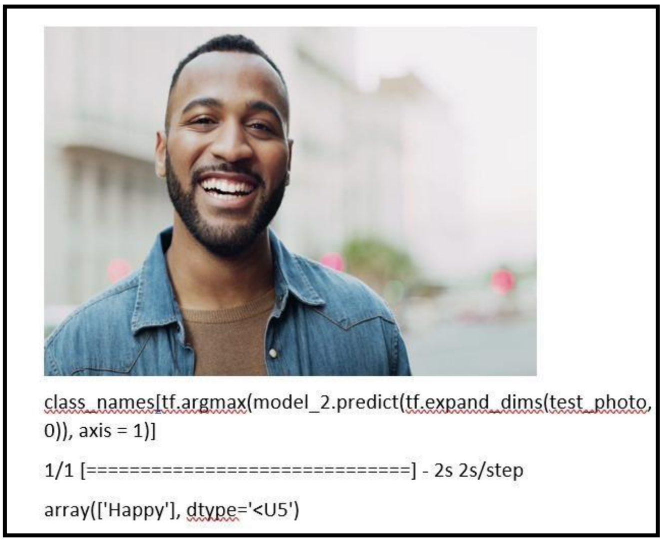

Fig 5. Output generated by the processed model.

Fig 6. Output generated by the processed model.

Our model demonstrated accurate output generation, correctly identifying happiness as depicted in Fig 5. Subsequently, during testing with an image of an angry person, our model continued to exhibit precision by accurately predicting the emotion as 'angry' as shown in Fig 6. This indicates the robustness and reliability of our model in accurately classifying a range of emotions, showcasing its effectiveness across diverse scenarios.

B. Conclusion

The proposed face emotion detection model, utilizes Convolutional Neural Networks (CNNs) and represents a significant step forward in accurately recognizing emotions from facial expressions. Through extensive experimentation and analysis, we have demonstrated the model's capacity to extract intricate features and effectively classify emotions across diverse individuals and contexts. While the model showcases impressive accuracy and adaptability, there remain challenges such as handling subtle emotional cues and mitigating biases in training data. Continued research into improving interpretability, addressing privacy concerns, and refining training strategies will contribute to the model's continued success in real-world applications. Our work underscores the potential of CNN-based emotion recognition systems to enhance human-computer interaction, psychological studies, and various domains requiring emotionally aware technology.

References

[1] Hosur, V., & Desai, A. R. (2022). Facial emotion detection using convolutional neural networks. 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon). https://doi.org/10.1109/mysurucon55714.2022.9972510Sabrina

[2] Begaj, S., Topal, A. O., & Ali, M. (2020). Emotion Recognition Based on Facial Expressions Using Convolutional Neural Network (CNN). IEEE Xplore. https://doi.org/10.1109/contesa50436.2020.9302866

[3] D Y Liliana: "Emotion Recognition from Facial Expression using Deep Convolutional Neural Network", IOP Conf. Series: Journal of Physics: Conf. Series 1193 (2019) 012004, 2019

[4] Jaiswal, A., Raju, A. K., & Deb, S. (2020). Facial Emotion Detection using Deep learning. 2020 International Conference for Emerging Technology (INCET). https://doi.org/10.1109/incet49848.2020.9154121

[5] Kondaveeti, H. K., & Goud, M. V. (2020). Emotion Detection using Deep Facial Features. 2020 IEEE International Conference on Advent Trends inMultidisciplinary Research and Innovation (ICATMRI). https://doi.org/10.1109/icatmri51801.2020.9398439

[6] Jaiswal, S., & Nandi, G. C. (2019). Robust real-time emotion detection system using CNN architecture. Neural Computing and Applications, 32(15), 11253–11262. https://doi.org/10.1007/s00521-019-04564-4

[7] Baffour, P. A., Nunoo - Mensah, H., Keelson, E., & Kommey, B. (2022). A survey on deep learning algorithms in facial Emotion Detection and Recognition. Jurnal Inform, 7(1), 24 – 32. https://doi.org/10.25139/inform.v7i1.4563

[8] Saxena, A., Khanna, A., & Gupta, D. (2020). Emotion Recognition and Detection Methods: A Comprehensive survey. Journal of Artificial Intelligence and Systems, 2(1), 53–79. https://doi.org/10.33969/ais.2020.21005

[9] Forhad Ali, Md., Khatun, M., & Aman Turzo, N. (2020). Facial Emotion Detection Using Neural Network. International Journal of Scientific & Engineering Research, Volume 11(Issue 8), ISSN 2229-5518. https://www.researchgate.net/publication/344331972

[10] Mehendale, N. (2020). Facial emotion recognition using convolutional neural networks (FERC). SN Applied Sciences, 2(3). https://doi.org/10.1007/s42452-020-2234-1

[11] Kim, E., Bryant, D., Srikanth, D., & Howard, A. M. (2021). Age Bias in Emotion Detection: An Analysis of Facial Emotion Recognition Performance on Young, Middle-Aged, and Older Adults. AIES '21, May 19–21, 2021, Virtual Event, USA. https://doi.org/10.1145/3461702.3462609

[12] Joseph, A., & Geetha, P. (2019). Facial emotion detection using modified eyemap–mouthmap algorithm on an enhanced image and classification with tensorflow. The Visual Computer, 36(3), 529–539. https://doi.org/10.1007/s00371-019-01628-3

[13] Mukhopadhyay, M., Pal, S., Nayyar, A., Pramanik, P. K. D., Dasgupta, N., & Choudhury, P. (2020). Facial Emotion Detection to Assess Learner's State of Mind in an Online Learning System. ICIIT '20: Proceedings of the 2020 5th International Conference on Intelligent Information Technology. https://doi.org/10.1145/3385209.3385231

[14] Fatima, S. A., Kumar, A., & Raoof, S. S. (2021). Real time emotion detection of humans using Mini-Xception Algorithm. IOP Conference Series, 1042(1), 012027. https://doi.org/10.1088/1757-899x/1042/1/012027

[15] Bhadangkar, D., Pujari, J., & Yakkundimath, R. (2020). Comparison of tuplet of techniques for facial emotion detection. 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). https://doi.org/10.1109/i-smac49090.2020.9243439