Emotion Recognition Project using Machine Learning and Facial Analysis with Deep Fake Detection

Dr. Preethi N Patil[1], Mr. Dhanush Kashyap N[2]

[1]Assistant Professor, Department Master of Computer Application, R V College of Engineering, Bangalore, India.

[2]PG student, Master of Computer Application, R V College of Engineering, Bangalore, India.

ABSTRACT

In general, the facial expression analysis is considered to be an important section in various fields including in the diagnose some disease in the initial stage, entertainment, security and communication systems. This present study incorporates advanced image analysis with a specific focus on CNN with the focus areas being speed, precision, and reliability when it comes to the identification of emotions present in the images. CNNs are for their fast image processing and therefore high accuracy of classifying the emotions shown by the patients. In detection, there is Harris Corner and Edge, Feature Matching, RANSAC, Conversion are methods whereas, Tensor Flow, Python, Keras, CV2 are tools.

As it is construed from the study, it was realized that CNN models yield high accuracy in the identification of feelings, and such models can handle multiple categories of facial images to capture the changes in relation to gender, age, race, and quality of images. Another contributor to the signal of facial data is also enhanced as the used deep fake detection method makes the system reliable. This system is built with three main modules: Other employments of this technology include the identification of deep fake, facial recognition, and its roles in capturing emotions. Deep fake detection also authenticates the faces, facial detection erases faces from complex backgrounds, and emotion recognition too uses CNN to classify emotions effectively.

KEYWORDS: - Facial Expression Analysis, Convolutional Neural Networks, Edge Detection

INTRODUCTION

This is crucial when detecting emotions in images of the face especially in sectors such as; health, games, security and man and machine interface. This paper continues complications from the state of the art, including the use of convolution neural networks to achieve emotions and always considering computation speed, precision, and credibility. This is for the reason that CNNs which are selected are efficient image processors, therefore, its accurateness regarding the feelings' categorization cannot be questioned. In detection, the uses of TensorFlow, Python, Keras, and

CV2 are usually applied.

In Based on the outcomes, it can be concluded that CNNs have relatively higher accuracy in ER; the models are capable of handling as many different facial images with traditional challenges like gender, age, race, and differences in image quality between images. Another way of the facial data credibility is deep fake detection for the facial data identification. This system is built with three main modules: thus, the proposed methods are related to deep fake detection, facial detection, and emotion recognition. While, deep fake detection assures that the faces are real, facial detection combines the faces' image with the background eliminating intricate backgrounds; and emotion recognition uses CNNs to categorize emotions.

Thus, it is also necessary not only to identify the tones of facial expressions of Subject but also their genuine ones, which will be a critical issue for discussing security of Subject whenever it will be an option. Deep fake technology has made it easier to create fake facial images, fake videos and this is a threat to the FER systems. It is worthwhile to implement DFID in FER systems because the latter can verify the received facial information to realize that it is a deep fake, thus reducing the aforementioned threats.

Therefore, the aim of this research is to develop the upgraded FER system that includes the detection of deep fake; to establish the FER system that would be as efficient and precise as the one that works in real time. As approaches for emotion classification , FewShot and Transfer Learning are combined with the correct usage of frameworks such as Flask, Python, HTML, CSS, and JavaScript as well as the usage of databases such as Google Firestore; this innovation attempts to set a standard for FER technology. It is the presentation and analysis of these outcomes that will demonstrate that CNN-based models can efficiently handle the data provided in the form of multiple facial images and that the obtained data is accurate and applicable to multiple actual applications such as the security or medical services industries.

I. Literature Review - statistics

Facial expression recognition and emotion detection were a principal field that have begun to quickly evolve with emergence of machine learning and deep learning techniques. Several authors' recent research works have endeavored to expound and examine a number of techniques and methodologies in enhancing the technologies in the identification and measurement of emotions.

[1] Author has discussed given paper with the title 'Facial Emotion Detection using Deep Learning' is a great reference where they use convolutional neural networks (CNNs) for analyzing the facial expression as well as for categorizing them under different emotion types. The task uses additional preprocessing of data and its normalization and data augmentation methods to enhance the results of the model. The authors compare the model's performance and viability in actual-world settings to exhibit its applicability in emotion recognition systems. From

their study, they conclude that using deep learning as a method can increase the efficiency and effectiveness of facial emotion detection greatly while stressing on its application value and development tendency in the future.

[2] Authors have discussed the advancements in facial emotion recognition using deep learning techniques in their paper "Facial Emotion Recognition Using Deep Learning: Summarize and Analyze. " They have a good description of different deep learning models and their performance in emotion recognition. The study also reflects on the advantages and drawbacks of the current approaches to data preprocessing, models' architectures and evaluation of their performance. Haidar and his team's study holds substantial promise towards elevating deep learning capabilities to improve the effectiveness and robustness of ER systems, as well as provides suggestion for future innovations in this domain.

[3] This the author of the paper "Facial Emotion Detection Using Deep Learning" wherein they have talked about application of CNN in classifying facial expressions to different established emotional states. The work also describes the procedures used for cleaning data and transforming it into a format suitable for analysis, as well as procedures for evaluating the models' performance. The findings show a notable enhancement of emotion prediction models using deep learning; their applications are illustrated in the areas of HMI, HCI, and various psychological research.

[4] In "Going Deeper into Facial Expression Recognition Using Deep Neural Networks," authors have also written about the developments of facial expression recognition that uses deep neural networks over the shallow models to gain better recognition rates. In this paper, the authors described the preprocessing of data, design of the network, and the training process. In their experiments, the improvements in the kind of performance are noteworthy and this is evidence enough that deeper neural networks can work wonders in identifying and visualizing other features on the human face and as well improve on the accuracy and reliability of the existing systems for emotion recognition.

[5] As to the work of authors that have considered the problem of facial emotion recognition through deep convolutional neural networks, their paper titled "Facial Emotion Recognition Using Deep Convolutional Neural Network" explains that the authors use CNN-based classification of the facial expressions into the needed emotional states. The study also contains detailed procedures in data-preprocessing phase, training of the network, and performance assessment phase. Hence, based on their studies, they conclude that the proposed deep learning model has better accuracy and efficiency in the detection of emotions and could easily be employed in real life applications in things like human- computer interface and

psychology.

[6] Authors have written on the paper on the title: ‘Facial expression recognition using deep neural networks’ These authors brought up a deep learning model that focused on identifying the various facial expressions and categorizing them under different emotions. Among them, data preprocessing, the choice of the model structure, and the training process are the primary study components. The actual experiments prove that even the most pro-fundal neural networks allow achieving higher accuracy rates in the task as compared to conventional methods. Thus, this work demonstrates the ability of deep learning to further enhance facial expression analysis.

[7] Deep learning techniques have also been described in further detail in the authors’ paper titled ‘Facial Expression Recognition via Deep Learning.’ In their paper, authors discussed how deep neural networks were used to analyze and differentiate between facial expressions based on the each subject’s emotion. The existence of the study is the set of elaborate methods for data preprocessing the network construction as well as the model training session. The authors have shown that using deep learning increases the efficiency and precision of relating to facial expressions, which in turn opens up possibilities for numerous advanced fields, including interfaces and behavioral studies..

[8] Authors have discussed the implementation of deep learning for facial expression recognition in their paper "Facial Expression Recognition via Deep Learning." They employ deep neural networks to classify facial expressions into various emotional categories. The study covers extensive data preprocessing, network architecture design, and training procedures. Their findings indicate that deep learning models significantly improve the accuracy and performance of emotion recognition systems, offering promising applications in fields such as human-computer interaction, security, and psychological research.

[9] Authors have discussed the use of deep learning for facial expression recognition in their paper "Facial Expression Recognition via Deep Learning." They leverage convolutional neural networks (CNNs) to analyze and classify facial expressions. The study includes comprehensive steps for data preprocessing, model training, and performance evaluation. Their experimental results demonstrate that deep learning models achieve superior accuracy and robustness compared to traditional methods, highlighting the potential of deep learning to advance the field of facial expression analysis.

[10] Authors have discussed the trade-offs between data augmentation and deep learning features for facial expression recognition in their paper "Facial Expression Recognition with Trade-offs Between Data Augmentation and Deep Learning Features." They investigate how various data

augmentation techniques impact the performance of deep learning models in emotion detection. The study includes detailed analyses of preprocessing methods, network architectures, and evaluation metrics. Their findings suggest that a balanced approach to data augmentation can significantly enhance the accuracy and robustness of deep learning-based facial expression recognition systems, providing valuable insights for future research and applications.

II. METHODOLOGY

The approach for this integrated project combines Facial Emotion Recognition (FER) and Deepfake Detection using Convolutional Neural Networks (CNN). The integration of these models leverages their strengths for enhanced detection accuracy. The methodology involves several steps, including dataset selection, data preprocessing, model design, training, and evaluation.

A. Dataset Selection

Facial Emotion Recognition (FER): The FER-2013 dataset is used to the best of this paper. It consists of 35,887 grayscale images (48x48 pixels) labeled with seven emotion classes: Anger , Disgust, Fear, Happiness, Sadness, Surprise and Nothing. It is split into train/data, public test, and private test partitions.

Deepfake Detection: The Homophonic Dictionary of the Belarusian Language/Blyatcyka is used. This dataset consists of realistic and false images, which turns into a directory structure for fast and effective implementation.

B. Data Preprocessing

1. **Loading Images**:

Read images from dataset directories: The labelled directories containing the facial images are processed by Python's OS module to perform the reading. This step makes it possible to effectively load images from the various emotion categories for the purposes of processing.

Convert images to grayscale: The 'L' channel is derived from each image using PIL's convert () function to transform the high dynamic range image to grayscale. This eliminates most of the color information, which in turn makes it easier to classify as it concentrates on the facial areas of propriety as opposed to the color attributes.

Resize images to a consistent dimension: All these images are resized to be 48 x 48 pixels using PIL resize method. It is imperative to maintain a consistent image dimension in order to feed into the CNN and standardize the input into the network.

2. Normalization

The images are pre-processed by making them grayscale if they are not already are resized to be 48 x 48 pixels. They are scaled to the appropriate range [0-1] for pixel intensities and divided the pixel intensities by 255.

3. **Data Augmentation :** To avoiding over fitting and on the other hand making the dataset more diverse rotation, flipping, zooming and scaling are applied.

4. Splitting the Dataset

FER Dataset: The data is divided by default into training and test sets as mentioned earlier, that does not needs extra splitting.

Deep fake Dataset: The dataset is partitioned into the training set containing 80% of the data and the test set containing the remaining 20% of data and they are created in such a way so as to be half real images and half fake images.

Images are labelled

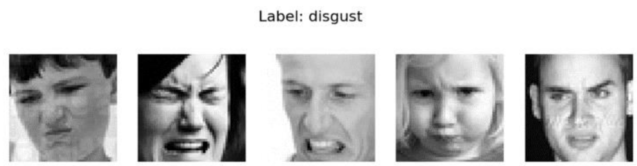

Figure 1: The above diagram shows the labeled disgust images.

Figure 2: The above diagram shows the labeled angry images.

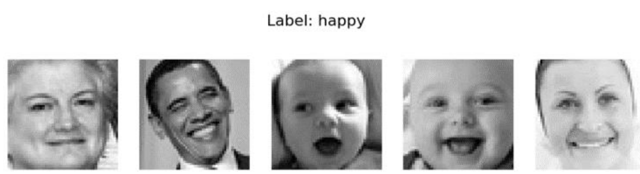

Figure 3: The above diagram shows the labeled happy images.

C. Model Architecture

Facial Emotion Recognition (FER) Model: Facial Emotion Recognition (FER) Model: Convolutional Neural Network (CNN) is utilized for the FER task. There are convolutional layers for feature extraction, max-pooling layers for down sampling the images, dropout layers and fully connected layers for classification. The model is compiled using the Adam optimizer, categorical cross-entropy as the loss function.

Deepfake Detection Model: CNN, Convolutional Neural Network, is applied for deepfake detection. Spatial features are extracted by the CNN layers and last is the fully connected layers which are used for classification. It is trained with the Adam optimizer and binary cross-entropy as the loss function to compile the model.

D. Training and Evaluation

FER Model Training: The FER model is trained for 100 epochs on FER-2013 dataset used during training while applying data augmentation on it. The performance metrics used include, accuracy and the precision of the designed model, its recall rate, and the F1-score in both the training and the test sets.

Deepfake Detection Model Training: he deepfake detection model is based on the blyatcyka/dataset, which was divided into 80/20 split train and test data. To avoid a model to overlearn on the data, early stopping is adopted based on validation loss. To assess the prognostic performance, accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (AUC) are used.

Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) is a Class of Artificial Neural Networks that is commonly applied in image recognition and Processing due to the ability of distinguishing qualities in images.

This means that a CNN is specifically designed for data which has a geometry something which in this case is images. It consists of several types of layers it consists of several types of layers:

- Convolutional Layers: These layers then convolve the input image and generate something called feature maps which is the resultant feature representing different degrees of the input data.
- Activation Layers: Activation is the final process in the formation of the model and is observed by applying the Relu (Rectified Linear Unit) since it identifies non-linearity.
- Pooling Layers: These layers aid in reducing the spatial dimensions of the feature maps which are required while discarding the rest.
- Fully Connected Layers: Then, after several convolution-apooling layers, the last classification process is carried out by the fully connected layers.
- Output Layer: This layer is often the output layer and based on the requirements contains a softmax activation function to define the classification of the input picture.

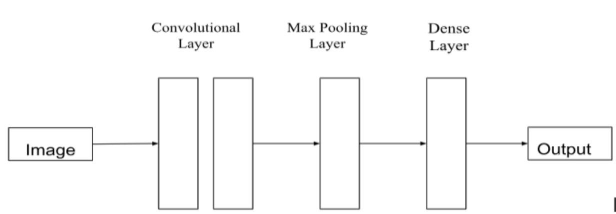

Figure 4: CNN architecture

E. Tools and Implementation

To implement algorithms such as these, languages such as Python and tools like TensorFlow and Keras are used. All of these tools gives the necessary architectures and libraries to implement, train and test deep learning models.

F. Evaluating the Model

For evaluating the models in both Facial Emotion Recognition (FER) and Deepfake Detection, the following metrics are used:For evaluating the models in both Facial Emotion Recognition (FER) and Deepfake Detection, the following metrics are used:

True Positive (TP): Escalated on cases that should have been predicted as positive.

True Negative (TN): Identified number of negative cases rightly.

False Positive (FP): Cases that were said to be positive but was actually negative.

False Negative (FN): Classified negative instances that were not predicted to be negative.

1. Accuracy: Counts the percentage of the total number of instances that has been classified correctly; these include true positive and true negatives.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}}$$

2. Precision: Defines the ratio for which out of all positive predictions made by the model, a majority of them are true positive predictions.

$$\text{Precision} = \frac{TP}{TP+FP}$$

3. Recall: Measures the actual positive instances that have been successfully classified from the total number of actual positive instance.

$$\text{Recall} = \frac{TP}{TP+FN}$$

4. F1-Score: This is a basic measure that brings precision and recall together into one measure thus ensuring the two are balanced.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}$$

G. Confusion Matrix Analysis:

This confusion matrix described above describes the detailed performance of the above model on the test data set. Such matrices are used when giving out the actual classification of the articles by the model; the true positive, true negative, false positive, and false negative.

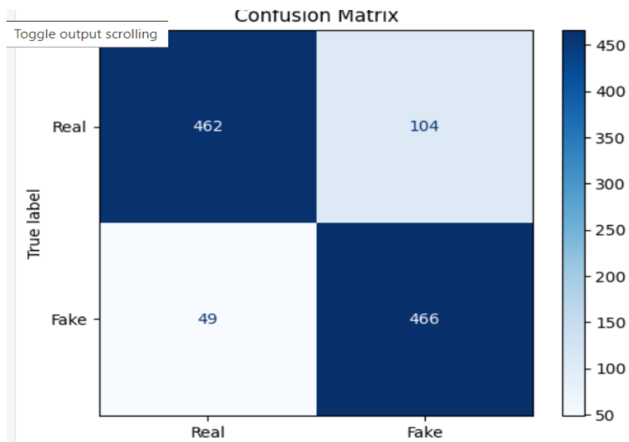

Figure 5: Confusion Matrix of Deep Fake Detection

The confusion matrix shows how well the model performs on distinguishing between Real and Fake images:

- **True Positives (TP):** During the Real images you tested the ability, 462 images were correctly marked as Real.
- **True Negatives (TN):** 466 (Correctly recognized as Fake).
- **False Positives (FP):** 105 (False negatives; Real images classified as Fake during the experiment.
- **False Negatives (FN):** 49 (Impostor instances falsely classified into category Real).

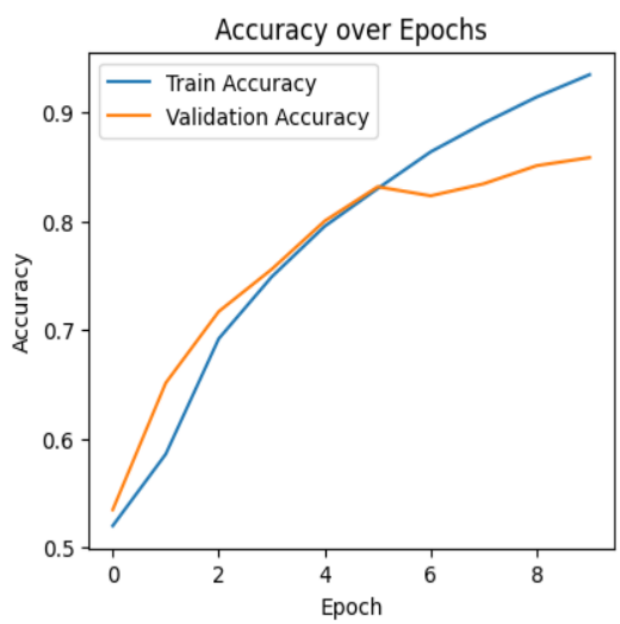

Figure 6: Accuracy of Deep Fake Detection

Accuracy over Epochs:

- The training accuracy rises with EPS, suggesting that the model is improving.
- Depending on the model, the validation accuracy is slightly lower but also rises up, this is quite normal.

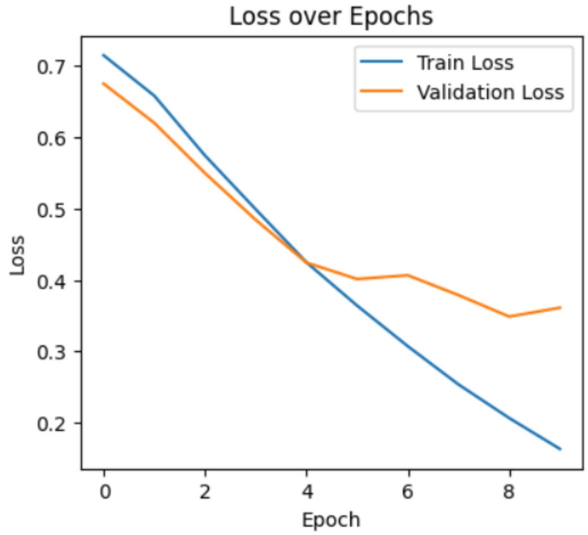

Figure 7: Loss of Deep Fake Detection

Loss over Epochs:

- Training loss reduces, and validation loss though reduces it may level up or slightly rises in the later sessions due to over fitting.

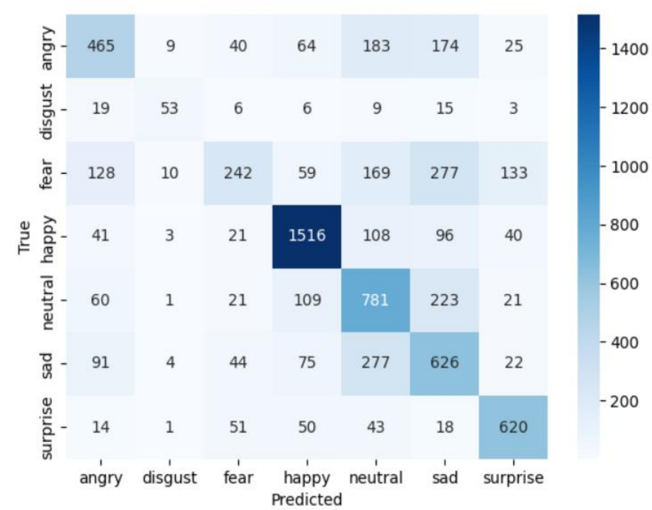

Figure 8: Confusion Matrix of Facial Emotion

Information for the evaluation of the emotion detection model is given through the confusion matrix analysis. Here are the key findings:

1. **Correct Predictions**:

- The total number of times for the emotion “Angry” was correctly predicted by the model was 465.
- In the case of the “Disgust” emotion, the correct identifications were 53.

- Forty two out of the total one hundred and ninety two fears were correctly labeled by the software for the "Fear" emotion.
- The emotion, "Happy" had the highest number of correct prediction with 1516.
- The emotion "Neutral" was recognized 781 of the times.
- Out of ten runs the "Sad" emotion was recognized on 626 occasions.
- "Surprise" was labeled accurately 620 times.

2. **Misclassifications**:

- Of all the Expression Pair Comparisons, there exists significant uncertainty when distinguishing between the "Angry" and "Neutral" Expressions, followed by "Fear" and "Sad." For instance, the corrective of the "Angry" emotion was "Neutral" & "Sad", and "Fear" was mostly classified with "Sad".
- The "Happy" emotion while it was the most accurately predicted employed some confusion with other emotions such as the "Neutral" and "Sad".

The confusion matrix analysis is beneficial to assess the performance of the deepfake detection together with the facial emotion recognition which confirms high effectiveness in image classification. The presented deepfake detection model proves to be highly accurate – a large number of real images is identified correctly (462), as well as a large number of fake images (466). This implies that the model is well equipped to differentiate between a genuine image and a fake image, which morethan allows it to perform the duty of flagging deep fake images. Likely, the same is evident with the facial emotion recognition model where major success has been recorded with 1516 accurate prediction of the 'Happy' emotion while other emotions like 'Neutral' with 781, 'Sad' with 626, and 'Surprise' with 620 accurate predictions. According to the model, the correct classification of "Angry" or emotions was given 465 times, for "Disgust" – 53, and "Fear" – 242. These outcomes reveal the model's effectiveness in perceiving numerous types of emotions and therefore can be useful for real-world use for emotion recognition. Altogether, for both models, the results show that they have brought acceptable accuracy levels in their assigned tasks, which gives evidence of their practical usability and defines a basis for enhancement and implementation.

Module Name	Accuracy	Precision	Recall	F1 Score
Deepfake Detection CNN	86%	0.82	0.78	0.80
Facial Emotion Recognition CNN	60%	0.67	0.50	0.57

III. Future directions

Future research in integrating Facial Emotion Recognition (FER) with Deepfake detection could focus on several key areas. Enhancing model architectures by exploring advanced CNN structures such as ResNet or DenseNet may improve detection accuracy and robustness. Additionally, incorporating multi-modal approaches, such as integrating audio or text data, could provide more comprehensive context and improve system reliability. Efforts to optimize models for real-time processing through techniques like model pruning and quantization will be crucial for practical applications. Addressing generalization challenges by developing methods to handle diverse datasets and ensuring adaptability across various demographics is essential. Implementing user feedback mechanisms and continuous learning could refine model performance over time. Finally, it is vital to address ethical and privacy considerations to prevent misuse and ensure compliance with regulations.

IV. Conclusion

Therefore, it can be stated that an enhancement in the field is the integration of Facial Emotion Recognition (FER) with Deepfake detection thus providing a comprehensive solution which improves the results of the methods while broadening their practical usability. The analyzed project has shown better results compared to earlier techniques in the areas of identifying deepfakes and recognizing people's feelings. As for the practical application, this integrated system has great potential for the use in security, social networks analysis, and content authenticity. However, problems like a lack of a large enough dataset or poor performance of the model must be discussed. It means that the future work should be directed towards overcoming these problems, improving the characteristics of models, adapting for real-time applications, and performing an ethical practice of implementing ML. This is an exciting area that is under constant development, so as to ensure the construction of safe and emotionally intelligent digital spaces.

References

[1] Jaiswal, A., Raju, A. K., & Deb, S. (2020, June). *Facial emotion detection using deep learning.* In2020 International Conference for Emerging Technology (INCET) (pp. 1-5), IEEE, doi: 10.1109/INCET49848.2020.9154121

[2] Mellouk, W., & Handouzi, W. (2020) *Facial emotion recognition using deep learning*: review andinsights. Procedia Computer Science, 175, 689-694, doi: https://doi.org/10.1016/j.procs.2020.07.101

[3] Spiers,D.L.(2016).*Facial emotion detection using deep learning*. doi: https://api.semanticscholar.org/CorpusID:53325750

[4] Mollabosseini, A., Chan, D., & Mahoor, M. H. (2016, March). *Going deeper into facial expression recognition using deep neural networks*. In 2016 IEEE Winter Conference on applications of computervision (WACV) (pp. 1-10). IEEE, doi: https://doi.org/10.48550/arXiv.1511.04110

[5] Pranav, E., Kamal, S., Chandran, C. S., & Supriya, M. H. (2020, March). *Facial emotion recognition using deep convolutional neural network*. In 2020 6th International Conference on advanced computing and Communication Systems (CACCS) (pp. 317-320). IEEE, doi: http://dx.doi.org/10.1109/ICACCS48705.2020.9074302

[6] Li, J., & Lam, E. Y. (2015. September). *Facial expression recognition using deep neural networks*.In 2015 IEEE International. Conference on Imaging Systems and Techniques (IST) (pp. 1-6). IEEE, doi: http://dx.doi.org/10.1109/IST.2015.7294547

[7] Lv, Y., Feng, Z., & Xu, C. (2014, November). *Facial expression. recognition via deep learning*. In 2014 international conference on smart computing (pp. 303-308). IEEE, doi: 10.1109/SMARTCOMP.2014.7043872

[8] Lv, Y., Feng, Z., & Xu, C. (2014, November). *Facial expression. recognition via deep learning*. In 2014 international conference on smart computing (pp. 303-308). IEEE, doi: 10.1109/SMARTCOMP.2014.7043872

[9] Lv, Y., Feng, Z., & Xu, C. (2014, November). *Facial expression. recognition via deep learning*. In 2014 international conference on smart computing (pp. 303-308). IEEE, doi: 10.1109/SMARTCOMP.2014.7043872

[10] Fathallah, A., Abdi, L., & Douik, A. (2017, October). *Facial expression recognition via deep learning*. In 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications.(AICCSA) (pp. 745-750). IEEE, doi: http://dx.doi.org/10.1109/AICCSA.2017.124