A Modified Deep Learning Based Convolutional Neural Network Model for Facial Emotion Prediction

Praveen Kumar
Department Of CSE
Vignan's Foundation For Science, Technology and Research (Deemed to be University)
Guntur, India

Prashanth Upadhyay
Department Of CSE
Vignan's Foundation For Science, Technology and Research (Deemed to be University)
Guntur, India

***Abstract*— Facial Emotion acts as a difficult understanding for a human being or machine to process the real meaning. If we didn't approach the way it, is we will face greater consequences in this AI world. A system needs to have better understanding of a person in terms of his feelings expressed over face. The importance of expressing and classifying based on facial expression gives larger tactic in real world applications. Through this way a greater reliable data is produced for the satisfactory of the recommendations of the consumer. Latest Deep Learning based Convolutional Neural Network adapts the facial expression and predicts the state of emotion in the user. We trained with a huge dataset in order to find the accurate labelling with all kinds of emotions, such as Sad, Happy, Angry, Neutral, Disgust and Surprise. Here we used fully processed to train our model with Keras deep Learning algorithm. Which therefore accompanied with real time dataset served as input for our model. Convolutional Neural Network provided accurate results with higher precision and greater sentiment capturing in classifying the emotions with respective to user. Our research methods follow 1) collecting the datasets from real time world 2) process them towards labelling based on each different emotions 3) refining the labels and constructively divide and prepare the training dataset 4) Flow towards the deep learning algorithms CNN, ResNet-50, VGG-16, and MobileNet. 5) classification and analysing each combined proposed algorithm across its precision and accuracy values 6) determining the best algorithm in competence of deep learning methods. Although, each algorithm performed in its own precision and accuracy values. The final way to determine a novel approach in identifying the facial emotion is using VGG16 + MOBILENET + RESNET50 where each and every layer goes with labelling and composed with a number of convolutional, pooling, flattening and dense ones. It recurrently analyses the input soon after labelling the facial expression, to enhance the model efficiency we also utilized cross-entropy loss function and Adam Optimizer. Across all the methods accuracy is the key role to differentiate the algorithms. Final Validation of accuracy of 80% obtained of our proposed hybrid model to incorporate the new data and analysing its nature of acceptance.**

Keywords— Deep Learning, Convolutional Neural Networks (CNN), Dataset Preprocessing, Keras Deep Learning, Categorial cross-entropy function, Adam Optimizer.

I. INTRODUCTION

In recent trends of the image processing the Deep Learning techniques follow a greater use case, since these algorithms are having a wide range of data models. Understanding and uttering the image expression is quite challenging effort for a machine to give the recommendations based on it. Soon, the user's expression is captured it moves to next steps of recommendation on the reliable data. In current industry, the deep learning algorithms such as Convolutional Neural Networks given highest accuracy, as it is the one of the largest algorithms even if we have big data.

Over the years in research field deep learning techniques give best results to produce the identification from facial expressions. For visual representation of the data from the image produced by the user is classified used convolutional neural network. Due to this algorithm, it provides wide scope in feeding the data from training datasets and trained over many layers. Layering in current model provide the amid results by learning from feature extraction [15][14] or engineering. These algorithms process the spontaneity of the data and its architecture further in an enhanced way.

Prompt and feature absorption from the images in a real-time world is highly challenging. This is because the images we capture doesn't have the resolution and, in the end, it creates a problem for the machine to understand the image and producing the facial expression. So, considering all the fundamental key problems, we came up with a rigours approach to create an algorithm for capturing the highest accurate results. Over the largest impact on deep learning methods to work over the precision, unless it is necessary to obtain correct and accurate results can be used as best algorithm. In our case ResNet-50, VGG-16, and MobileNet everything produced greater results but not much accurate results than using convolutional neural network [16][17[18].

Every frequency from the image have continued as certain amount of data, to feed the algorithm. Unless it is not a appropriate image with an expression containing on its no face picture, it will not able to understand and classify in our labelling. Since, these labelling can be only done when it is an appropriate image with vivid emotions [47]. If it isn't an image with a face then by default the system will labels it as Neutral expression. This is one of the use cases in our research to cope with it. Considering all the use cases it gives greater results with accurate accuracy and precision values.

Understanding any kind of algorithm can only be produced in the amount of training the dataset with a proper labelling's. In context, with labelling, the dataset can be evaluated in order to train the dataset with real time face expressions [19][20][21]. The usage of algorithms can impact in a larger direction in the source of truth by various call out features. For the whole process of creating or generating the output from the algorithm can be captured after removing all the noises from the image and understanding the facial emotion of the user.

If any expression understanded as angry or surprise, few of the expressions can be wide open of the user's mouth. At time algorithm will understand based on the user expression on eyebrows and the way the chin is lifted up if it is angry. All the expressions can be having similar way of expressions but it differs on the way you or machine understands it. The way you want to present your intentions can be captured as facial emotions, a part from the feelings you have.

Not every expression is angry or sad or surprise or disgust, each have their own impact of analysing the input. Some might have already known user interpretations from the trained dataset [42][44]. On this contrary, our research shows all the limitations can be removed in approaching the training dataset to be more and train through many layers. Convolutional neural networks have been understanded as greater impacting methods in usage of Machine learning [43]. This helps the user to understand how his expression is impacting the greater world.

Whenever there is a change in exploring the facial context, one can normal understand through the way the expressions are shown in the face of the user. One of the used methods in our research shows we have taken ResNet-50, which contains more than 50 convolutional neural networks over the millions of images data. The larger it is the greater impact it will be, in obtaining the accurate results. It is designed to be captured as a powerful method to obtain the results with more accurate way.

Basically, human expressions can be classified in two ways central and automatic [46][49]. The automatic expressions are considered to be mentally present the users' emotions on face, whenever there is a sudden change in the system or behaviourally impacting environment. It makes the user to reflex his emotions based on the neural system activity. These can be used and impacted to capture for understanding the human expression and act accordingly for the system in greater extent. [34] This way we are going to make the machine to understand the emotions from the human and act accordingly on the repurpose of this method over and over again.

In the end, we are going to discuss the methods and results of our research principles and the working methods in order to make it more novel. The way of translating an image to understanding the facial expression for identifying its emotion is quite to be a larger topic to explore.

II. Ease of Use

The usage of Deep Learning methods such as convolutional neural network gives a wide ease for using the product in a seamless way. In the end implementation can be utilized in any kind of forms inputting the data into the system for performing the high-end algorithms. This derives the data to be predicted and classified across simplified methods. An example of such tools was implemented using existing libraries and frameworks such as Tensorflow, keras, PyTorch and many more [45][47].

In the end every library and method can ease the workflow in order to classify the user emotions based on neural networks, due to this pre constructed mechanism of processing the data, architecture modelling and hyper parameters, so this ensures people who are working on building this product can mainly work on logical building of the data. Since, accuracy is the key for every method we are using in implementing an application, our research eases the time for consuming and interpreting the data as a result

III. Related Work

Visualization of images and knowing the internal meaning for image in order to understand its context is difficult. This can be achieved with lot of methods in Deep learning. We have taken existing methods into consideration in developing our model. [1] It states to differentiate the algorithmics across the model to understand the binary central compounding the facial representation. Compounding of the facial expressions in automatic feel of neural organs can be achieved with spontaneity of the workflow. Although each algorithm has its own significant approach [2] [3][4] have vast usage of convolutional neural networks into consideration for understanding the facial emotions if any misplace of system happens. The algorithmically approach of each data interpretation of libraries, its mathematical proven technologies have vivid usage in our research.

Automatic Facial expressions [5][40][41] based on vivid faces effective in robots with a multigranularity regional representation in affectively on robots. Continuous progressing on the macro expressions is a method to understand in a facial expression, where it helps to produce the wide range of accuracy results [6]. Understanding of the facial expression and producing a survey across all the proposed techniques was classified [8] and continuously reinforcement learning from the data taken from the facial expression of the user and producing a great workflow [9].

Visually [7] producing an attribute in learning the facial expression is a thing to be considered while validating and preprocessing an image.[10] Multimodal emotion identification is a challenging approach, but usage to propose the labelling on facial expressions based on speech and EEG. [11] Although deep learning methods can be utilized and

proposed in a huge number of products, we identified already proposed surveys and produced to propose our solution in identifying highest accuracy for Facial Emotion Recognition. Multimodal Emotion Recognition based on Facial Expressions, Speech, and EEG.
Using attributes and Deep learning methods the techniques followed in [12] to produce the behavioral facial expressions from the children is highly challenging but, the system accurately predicted the emotions from the images, with highly enhanced methods. Including the cloud methods in the model to predict the depression levels from the user [13] based on the facial expressions and training the dataset usage recurrent neural network and feature extractions.

IV. Methodology

A. *Importing all the packages:*

The utilization of deep learning methods can be simplified with providing the existing packages to be used for our model. In our case we have rigorous used TensorFlow and Keras libraries and Data Interpretation libraries. This made the facial identification of emotions develop as an application. Preprocessing with the larger datasets can be utilized in identifying the labelling in the images and produce faces with correct results. I have worked with the Kaggle dataset, which include 73249 images in dataset with 6 types of emotions [39].

In Data Preprocessing we prepared images to be of same size and quality. Standard preprocessing steps typically include resizing the images to a consistent size, converting them to grayscale, and normalizing pixel values.

Model Architecture Design of our CNN's architecture. The architecture may be a combination of convolutional layers [38], max-pooling layers, flattening and then fully connected layers then output layer, this derives a greater accuracy in training our model. While training [48], the network learns how to match facial features to corresponding emotions by adjusting the internal parameters(weights and biases) [33].

However, the capability of CNNs in facial emotion identification is impressive though it relies on dataset quality, network architecture, and training parameters among others. Moreover, CNNs can call for high computational resources on the training side, e.g. in case of dealing with large datasets or complex architectures.

B. *Convolutional Neural Network:*

There of course we have number of methods in deep learning, but we choose to CNN for better accurate results while performing the huge dataset classification across the emotions on faces. Although CNN performs as per the feature extraction from image and labeling the dataset with different faces across the emotions.

Convolutional Neural Network performs with the need of values as input and it goes and looks for gradient which looks for minimum loss.

$$G[m,n] = (f * h)[m,n] = \sum_j \sum_k h[j,k] f[m-j, n-k]$$

-----------------(1)

Here in this kernel Convolutional Equation, it is not only mentioned the key components of many other deep learning methods. It is a process where we take everything into a small matrix of numbers called as kernel, these goes to weights of the image pixels and therefore labelling of the images gets simplified with the behavior of the convolutional neural network [25][26][27].

In the above formulae, we denoted the input image as f and the kernel by h. The indexes of rows and all the columns of the output matrix are marked by m and n.

Typically, after every activation of the ReLU it is considered for non-linear neural activation with a vivid complex calculation.

ReLU (x) = max (0, ---(2)

The CNN architecture is specialized for the facial emotion detection with the main building blocks. It begins with the convolutional layers each of which uses filters to extract the image's features [34][35]. Batch normalization layers normalize the activations, which in turn contributes to the introduction of non-linearity via the ReLU activation functions for learning complex patterns. Max pooling layers bin feature maps, leading to reduction of spatial dimensions, and dropout layers are responsible for preventing overfitting by randomly dropping input units during training.

Flattening converts feature maps into vector for input by fully- connected layers [37], which use this vector to learn high level representations. The compact and last deep layer, using soft max activation as a function, outputs probabilities for seven facial emotions. Such an architectural design provides an efficient way of detecting and analyzing facial characteristics to grant reliable emotion classification.

C. *Working Principle:*

Step 1: Start.
Step 2: Select the dataset of images included the labelling
Step 3: Extraction of the face from the input image and process.
Step 4: If face is found in the image crop the face and do preprocess. Else end the step of preprocessing that particular picture and move to next.
Step 5: If face is found remove the noises in image which acts as obstacles in identification [36].
Step 6: Face resizing and preprocessing for the algorithm
Step 7: Face extraction from the image and go through all the test data labelling and perform CNN on the image
Step 8: Emotion is captured on the image of face and it results better accurate using CNN algorithm.

Step 9: Iterate the steps.
Step 10: End.

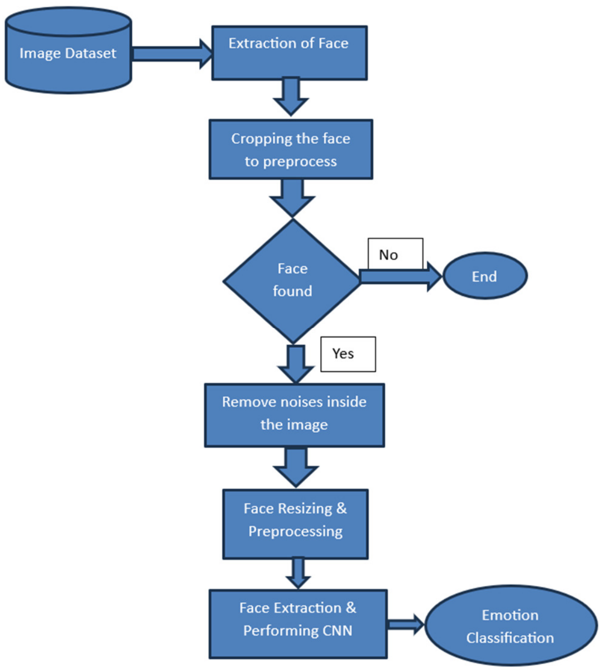

Fig.1: Flowchart of Proposed model.

The flow of our working model starts with inputting the image dataset with around 73 thousand images with corresponding labelling of the emotions. The preprocessing of our algorithm used TensorFlow and kernel libraries to extract more and utilize existing libraries [29][30]. Soon, after inputting the image, the model will extract the face from image and crop the face if a face is found in the image. If it didn't find the face it will end the process. For suppose if any image contains more than one face, then also our model processes the image and produce each significant emotions according to the facial expressions [28]. Each step of extraction and preprocessing happens once all the noises from the image has been removed perfectly and process the rest of the image into Convolutional Neural Network Algorithm. At the final step of our process model trained to classify any new image is inputted based on that it will extract and classify the emotions from image.

Final step of our model is model accuracy, which can be achieved after vast research over algorithms and found better results in the Convolutional Neural Network. Soon after any proposed algorithm only highest accuracy will be more effective than the remaining other algorithms. Details on Accuracy of the algorithms has been discussed further in next sections.

V. Experimental Results

A. Performance Evaluation Metrics

In general, efficiency can be evaluated by referring to quantitative measures of performance metrics, which ensures the success of a particular process, system, or activity. Widely depending on the context, these metrics can vary but typically involve assessing various factors such as productivity, quality, timeliness, cost-effectiveness, customer satisfaction, and overall performance standards.

TABLE I
ACCURACY OF EXISTING MODELS

Algorithms	Accuracy
VGG16 [49]	0.70
VGG19 [48]	0.69
Densenet121 [19]	0.70
MobileNet [35]	0.68
ResNet50 [39]	0.69

Table I summarizes the accuracy of five classifiers across different numbers of selected features. Across these VGG16 Densenet121 shows consistent performance maintaining the accuracy of 0.70.

The provided accuracy comparison table showcases the performance metrics of five different models: Densenet, ResNet-50, VGG-16, VGG-19 and MobileNet, with their respective accuracies standing at 0.70, 0.69, 0.70, 0.69 and 0.68. Notably, the Densenet and VGG16 model presented in the study exhibits the highest accuracy of 0.70, surpassing its competitors by a significant margin. This observation underscores the efficacy of the suggested Densenet architecture in achieving superior performance in facial emotion recognition tasks and so VGG16.

In comparison to ResNet-50, VGG-19, and MobileNet, which demonstrate accuracy levels ranging from 0.68 to 0.70, both the VGG16 and Densenet model offers a marked improvement in accuracy. This signifies the robustness and effectiveness of the proposed VGG16 and Densenet121 architecture in capturing intricate facial features and nuances associated with different emotions. The higher accuracy attained by Densenet and VGG16 model indicates its capability to discern and classify facial expressions with greater precision, thus enhancing the overall performance of facial emotion recognition systems.

The results highlight the importance of architectural design and feature extraction mechanisms in determining the performance of deep learning models for facial emotion recognition. While ResNet-50, VGG-19, and MobileNet are well-established architectures, the superior accuracy achieved by the presented Densenet and VGG16 model underscores the significance of tailored model architectures optimized for specific tasks. The success of the CNN model in outperforming its counterparts underscores its potential for real-world applications, where accurate emotion recognition plays a pivotal role in various domains such as human-computer interaction, healthcare, and social robotics.

Table II Proposed Hybrid Model

Proposed Hybrid Model	Accuracy
VGG16 + VGG19	0.62
VGG16 + DENSENET121	0.50
VGG16 + RESNET50	0.55
VGG16 + MOBILENET	0.60
VGG16 + VGG19 + DENSENET121	0.63
VGG16 + VGG19 + MOBILENET	0.65
VGG16 + VGG19 + RESNET50	0.55
VGG16 + DENSENET121 + MOBILENET	0.53
VGG16 + DENSENET121 + RESNET50	0.67
VGG16 + MOBILENET + RESNET50	0.80

Overall Table II, the comparative analysis underscores the significance of continual refinement and innovation in deep learning architectures for improving the accuracy and efficacy of facial emotion recognition systems. The demonstrated superiority of Densenet and VGG16 model underscores its potential as a valuable tool in advancing the state-of-the-art in facial emotion recognition technology, paving the way for enhanced human-machine interaction and emotional intelligence in artificial systems.

Fine tuning the algorithm makes best suitable to enhance the model to adapt all the changes in given dataset. In the proposed model, adjust the loss function 'categorical_crossentropy' loss in appropriate multi class classification problems, while binary classification is suitable since we are using binary labels (0 and 1). Always adjust the model architecture by increasing or modifying the layers of neural networks for potentially better learning. Hyperparameter tuning is required in adjusting the rates and batch sizes and other hyperparameters.

B. Performance Visualization

Evaluating facial emotions by visualizing performance metrics is crucial. This includes emotion detection models, aiding in trend analysis, comparison, and optimization through tools like accuracy plots, loss plots and AUC plots.

Figure 2 illustrates the Training and Testing accuracy graphs that we obtained from our model's experiments on neural networks for classification.

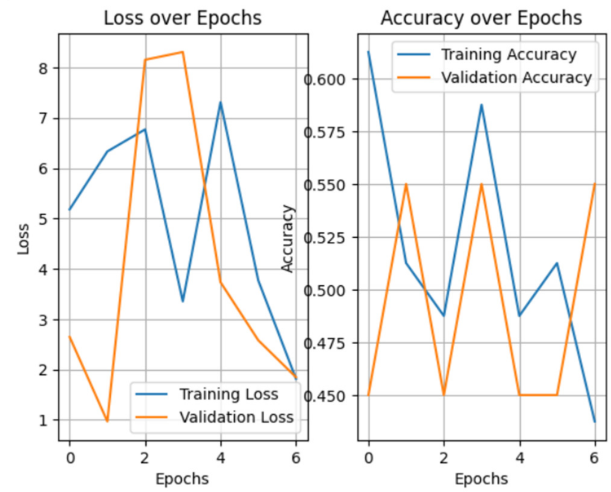

Fig.2. Metrics Evaluation Plots for VGG16 + MOBILENET + RestNet50

It illustrates the Training and testing accuracy graphs, that we obtained from our model's experiments on neural networks for classification.

C. Heatmap

With the graphical representation of data, Heatmap helps us understand the values by colors. Heatmaps are commonly used to visualize the intensity of data at different spatial locations. While the colors are measured by the pixels Of the image which gives the values that helps to format the image and analyze through the performance factors. In the context of deep learning and computer vision, heatmaps are often used to visualize the areas of an image.

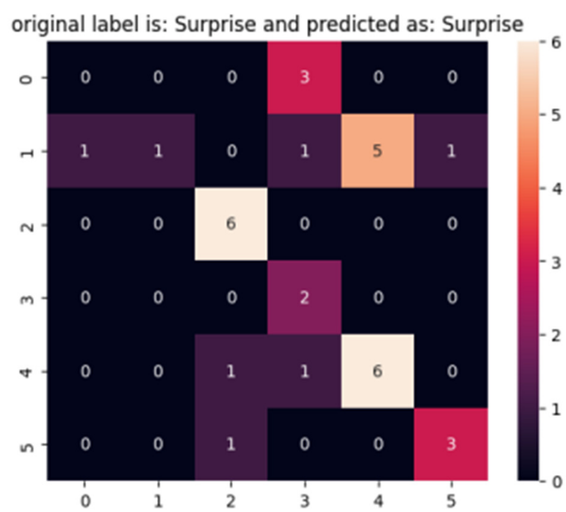

Fig.3 .Heatmap

Figure 3 following the completion of model training using both Training and testing datasets, the heatmap is given comparing the predicted outputs with the actual data.

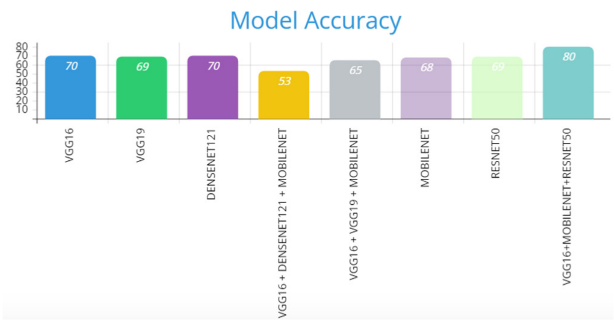

Fig.4.Model Accuracy Comparision

Figure 4, the image provided displays a bar graph comparing the accuracy of various image classification models. In the graph X-axis shows VGG16, VGG19, Densenet, Mobilenet and Resnet50. The Y-axis shows the accuracy, ranging from 0.0 to 1.0.

The graph highlights the performance factors on a scale which shows the change in the accuracy across different conditions. This visual helps identify how variations impact model accuracy and overall effectiveness.

Note: that model accuracy can vary depending on the dataset and task used.

D. Preditcion Based On Visualization

The input picture and its emotion prediction are demonstrated in Figure 5, Result-1 where the isolated emotion is highlighted. Once the analysis is done, the page title automatically changes with the emotion that the prediction algorithm believes is the most likely. This type of feedback happens live as a result of which the user gets an immediate idea and insight about which emotion the invisible system detected.

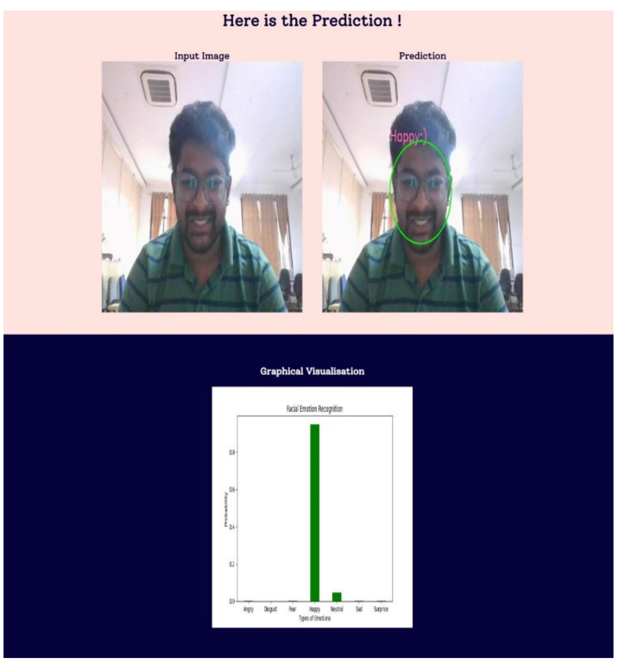

Fig.5 Result-1

The bar plot graph showing the detected emotions which originated from the input image and the process of this determination is as follows: the image is viewed and the Haar cascade classifier is used to detect faces [31] [32]. By this process, it slices the detected face and resizes the cropped area based on the input requirement of the model, and makes an emotion prediction. Upon prediction, it labels the image with the predicted emotion as well as some specific symbols. Furthermore, it displays a bar char to represent the probability distribution of various emotions and keeps both the annotated image and the chart. Lastly, it produces a list containing the original image name, bar plot name, the annotated image name, and the predicted emotion. As we can see the X-axis refers to the types of emotions and Y-axis refers to the probability of emotions i.e., percentage of emotion that is detected.

The main advantage of this research is it can also able to detect multi faces and can predict based on their appearance in the input image. The overall plotting of the algorithm in visual representation leads to better results in understanding our advantages of the algorithm and its necessity. Although the accuracy and precision are very close to each other to maintain the novel research of this model. Facial emotions in different faces can be changed as quick so in order to capture every face expression and its emotion is quite challenging, although it resulted better results. This happened due to training the dataset, in a volume and this makes more data to be predicted with correct labelling of that particular expression or emotion.

Fig.6: Result-2

If we have single face, it can also predict that our input image contains single face and it gives the correct dataset result as per the Results. These images tell the difference how we are populating the input image to make the system understand its significance and result.

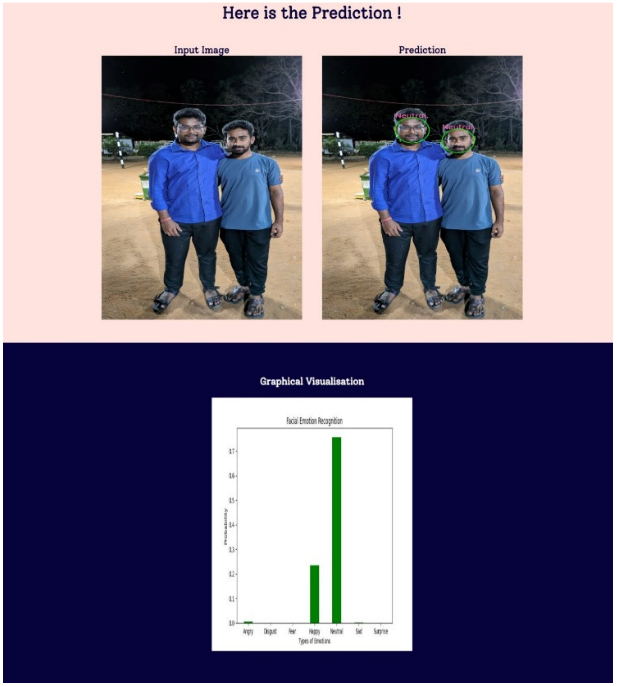

Fig.7 Result-3

. If you look upon Figure 8, you will see that more than one face is detected since immediately the faces are detected the title is fixed to Multi faces detected [22][23][24], and based on their appearance in the input image the prediction performed and highlights the emotions on the faces. There are sad and fear emotions are predicted in these multi-faces image.

By all means of the graphical visualization of the data produces its significance in providing the wide range in outputting the image expression on the face and producing what emotion that user holds off. This tells us even when we input an image with one face or multiple faces our system is capable of doing the classification across the images.

VI. Conclusion

In this research work we have successfully implemented an effective novel approach in order to identify the real-time facial expression prediction system, by utilizing the techniques of deep learning methods. Among all these algorithms, which is not only a recording but also predicted facial expressions as being integrated with the emotions. This given a highly accurate prediction using Convolutional Neural Network. In fact, the model we use does not just distinguish between sadness and happiness but it manages to identify many other emotions precisely as well. Graphic visualization is meant to be an opening-uptool as it exposes particular level of confidence to the modelwhich predicts emotions; that is why the comprehending of the process is facilitated.

Concisely, the deep learning approach of us in facial emotion detection system appear to get success in sensing and understanding human emotion bymeans of facial expression that is quite useful in AI, psychology and healthcare applications. The plan, for the future work, also contains increasing model accuracy, expanding the ability to detect emotions, develop the model to deploy it in the real world, optimizing user experience, and also considering the ethical consequences. Theseprograms will unwrap a new realm of emotion recognition and introduce solutions that can be easily deployed for real- time applications including speech emotion recognition.

In conclusion, the study depicts that the steps needed for using several deep learning models such as VGG16, VGG19, MobileNet, Densenet121, RESNET50. Among these models, VGG16, MobileNet and RESNET50 got the highest accuracy, with an incredible 0.80. These findings highlight the usefulness of VGG16, MobileNet and ResNet50 in our classification challenge and point to the potential for additional optimization and research of ensemble technique to improve model performance.

VII. Future Scope

Further implementation, can be extended in live so that it can be provided to users as software which helps to know the real time facial emotions of the people in Health [14], Army, Border and Airport security checking.

Although these methods can help vividly in understanding the product consumer experience, which helps user of this product to maintain security and get alert for any kind of harm that's going to take place.

VIII. Acknoldgments

I would like to express my sincere gratitude to everyone involved in developing the deep face emotion recognition system and appreciate the researchers in deep learning, my project guide Prashant sir, coordinator Deva sir and HOD of CSE Dr.K.V.Krishna Kishore sir for their valuable support.. My heartfelt thanks to family and friends for unwavering encouragement. This project wouldn't be possible without this talented and supportive community.

References

[1] R. Dong and K. -M. Lam, "Bi-Center Loss for Compound Facial Expression Recognition," in IEEE Signal Processing Letters, vol. 31, pp. 641-645, 2024, DOI: 10.1109/LSP.2024.3364055.

[2] M. C. Gursesli, S. Lombardi, M. Duradoni, L. Bocchi, A. Guazzini and A. Lanata, "Facial Emotion Recognition (FER) Through Custom Lightweight CNN Model: Performance Evaluation in Public Datasets," in IEEE Access, vol. 12, pp. 45543-45559, 2024, DOI: 10.1109/ACCESS.2024.3380847.

[3] R. Dong and K. -M. Lam, "Bi-Center Loss for Compound Facial Expression Recognition," in IEEE Signal Processing Letters, vol. 31, pp. 641-645, 2024, DOI: 10.1109/LSP.2024.3364055.

[4] C. Yu, D. Zhang, W. Zou and M. Li, "Joint Training on Multiple Datasets With Inconsistent Labeling Criteria for Facial Expression Recognition," in IEEE Transactions on Affective Computing, DOI: 10.1109/TAFFC.2024.3382618.

[5] Y. Wang et al., "MGR3Net: Multigranularity Region Relation Representation Network for Facial Expression Recognition in Affective Robots," in IEEE Transactions on Industrial Informatics, DOI: 10.1109/TII.2024.3353912.

[6] F. Wang et al., "Progressively Learning from Macro-Expressions for Micro-Expression Recognition," ICASSP2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, Korea, Republic of, 2024, pp. 4390-4394, DOI: 10.1109/ICASSP48485.2024.10446028.

[7] Y. Lv, G. Huang, Y. Yan, J. -H. Xue, S. Chen and H. Wang, "Visual-Textual Attribute Learning for Class-Incremental Facial Expression Recognition," in IEEE Transactions on Multimedia, DOI: 10.1109/TMM.2024.3374573.

[8] M. Karnati, A. Seal, D. Bhattacharjee, A. Yazidi and O. Krejcar, "Understanding Deep Learning Techniques for Recognition of Human Emotions Using Facial Expressions: A Comprehensive Survey," in IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-31, 2023, Art no. 5006631, DOI: 10.1109/TIM.2023.3243661.

[9] C.-S. Jiang, Z.-T. Liu, M. Wu, J. She, and W.-H. Cao, "Efficient facial expression recognition with representation reinforcement network and transfer self-training for human–machine interaction," IEEE Trans. Ind. Inform., vol. 19, no. 9, pp. 9943–9952, Sep. 2023.

[10] J. Pan, W. Fang, Z. Zhang, B. Chen, Z. Zhang and S. Wang, "Multimodal Emotion Recognition based on Facial Expressions, Speech, and EEG," in IEEE Open Journal of Engineering in Medicine and Biology, DOI: 10.1109/OJEMB.2023.3240280.

[11] S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in IEEE Transactions on Affective Computing, vol. 13, no. 3, pp. 1195-1215, 1 July-Sept. 2022, DOI: 10.1109/TAFFC.2020.2981446.

[12] J. Chen, C. Guo, R. Xu, K. Zhang, Z. Yang, and H. Liu, "Toward children's empathy ability analysis: Joint facial expression recognition and intensity estimation using label distribution learning," IEEE Trans. Ind. Inform., vol. 18, no. 1, pp. 16–25, Jan. 2022.

[13] Y. Yu et al., "Cloud-edge collaborative depression detection using negative emotion recognition and cross-scale facial feature analysis," IEEE Trans. Ind. Inform., vol. 19, no. 3, pp. 3088–3098, Mar. 2023.

[14] C. Bisogni, A. Castiglione, S. Hossain, F. Narducci, and S. Umer, "Impact of deep learning approaches on facial expression recognition in healthcare industries," IEEE Trans. Ind. Inform., vol. 18, no. 8, pp. 5619–5627, Aug. 2022.

[15] M. Sun et al., "Attention-rectified and texture-enhanced cross-attention transformer feature fusion network for facial expression recognition," IEEE Trans. Ind. Inform., vol. 19, no. 12, pp. 11823–11832, Dec. 2023.

[16] Y. Wang et al., "A systematic review on affective computing: Emotion models, databases, and recent advances," Inf. Fusion, vol. 83–84, pp. 19–52, 2022.

[17] P. D. M. Fernandez, F. A. Guerrero Pena, T. Ren, and A. Cunha, "FERAtt: Facial expression recognition with attention net," in Proc. IEEE Conf. Comput. Vis. Pattern Recog. Workshop.2019, pp. 837–846,.

[18] E. Pranav, S. Kamal, C. Satheesh Chandran and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2020, pp. 317-320, DOI: 10.1109/ICACCS48705.2020.9074302.

[19] S. Li and W. Deng, "Deep facial expression recognition: A survey," IEEE Trans. Affect. Comput., vol. 13, no. 3, pp. 1195–1215, Jul.–Sep. 2020.

[20] D. Zeng, Z. Lin, X. Yan, Y. Liu, F. Wang, and B. Tang, "Face2exp: Combating data biases for facial expression recognition," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2022, pp. 20291–20300.

[21] K. Zhang, Y. Li, J. Wang, E. Cambria, and X. Li, "Real-time video emotion recognition based on reinforcement learning and domain knowledge," IEEE Trans. Circuits Syst. Video Technol., vol. 32, no. 3, pp. 1034–1047, Mar. 2022.

[22] Y. Wang et al., "FERV39k: A large-scale multi-scene dataset for facial expression recognition in videos," in IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2022, pp. 20922–20931.

[23] Y. Wang et al., "Dpcnet: Dual path multi-excitation collaborative network for facial expression representation learning in videos," in Proc. ACM Int. Conf. Multimedia, 2022, pp. 101–110.

[24] C. Turan, R. Zhao, K. M. Lam, and X. He, "Subspace learning for facial expression recognition: An overview and a new perspective," APSIPA Trans. Signal Inf. Process., vol. 10, pp. 1–22, Jan. 2021.

[25] R. Zhao, T. Liu, J. Xiao, D. P. K. Lun, and K. M. Lam, "Deep multitask learning for facial expression recognition and synthesis based on selective feature sharing," in Proc. Int. Conf. Pattern Recognit., 2020, pp. 4412–4419.

[26] R. Zhao, T. Liu, Z. Huang, D. P. K. Lun, and K. M. Lam, "Spatialtemporal graphs plus transformers for geometry-guided facial expression recognition," IEEE Trans. Affect. Comput., vol. 14, no. 4, pp. 2751–2767, Oct.-Dec. 2023.

[27] R. Zhao, T. Liu, Z. Huang, D. P. K. Lun, and K. M. Lam, "Geometry-aware facial expression recognition via attentive graph convolutional networks," IEEE Trans. Affect. Comput., vol. 14, no. 2 pp. 1159–1174, Apr.-Jun. 2023.

[28] G. Wen, T. Chang, H. Li, and L. Jiang, "Dynamic objectives learning for facial expression recognition," IEEE Trans. Multimedia, vol. 22, no. 11 pp. 2914–2925, Nov. 2020.

[29] F. Xue, Q. Wang, and G. Guo, "TransFER: Learning relation-aware facial expression representations with transformers," in Proc. IEEE Int. Conf. Comput. Vis., 2021, pp. 3601–3610.

[30] H. Li, N. Wang, X. Yang, X. Wang, and X. Gao, "Towards semi-supervised deep facial expression recognition with an adaptive confidence margin," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2022, pp. 4166–4175.

[31] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK) : A complete dataset for action unit and emotion-specified expression," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2010, pp. 94–101.

[32] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in Proc. IEEE Int. Conf. Multimed. Expo., 2005, pp. 1–5.

[33] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion," J. Pers. Soc. Psychol., vol. 17, no. 2, pp. 124–129, 1971.

[34] R. Plutchik, "A general psychoevolutionary theory of emotion," in Theories of Emotion, 1st ed.Cambridge, MA, USA: Academic Press, 1980, pp. 3–33.

[35] Yahui Nan, Jianguo Ju, Qingyi Hua, Haoming Zhang, Bo Wang, A-MobileNet: An approach of facial expression recognition, Alexandria Engineering Journal, Volume 61, Issue 6, 2022, Pages 4435-4444

[36] S. Li and W. Deng, "Deep facial expression recognition: A survey," IEEE Trans. Affect. Comput., vol. 13, no. 3, pp. 1195–1215, Mar. 2020.

[37] Y. Ju, M. Jian, C. Wang, C. Zhang, J. Dong, and K. M. Lam, “Estimating high-resolution surface normals via low-resolution photometric stereo images,” IEEE Trans. Circuits Syst. Video Technol., early access, Aug. 04, 2023, doi: 10.1109/TCSVT.2023.3301930.

[38] Y. Ju et al., “GR-PSN: Learning to estimate surface normal and reconstruct photometric stereo images,” IEEE Trans. Vis. Comput. Graph., early access, Nov. 03, 2023, doi: 10.1109/TVCG.2023.3329817.

[39] Bin Li, Dimas Lima, Facial expression recognition via ResNet-50, International Journal of Cognitive Computing in Engineering, Volume 2, 2021, Pages 57-64, ISSN 2666-3074

[40] C. Zhang, K. M. Lam, and Q. Wang, “CoF-net: A progressive coarseto-fine framework for object detection in remote-sensing imagery,” IEEE Trans. Geosci. Remote Sens., vol. 61, 2023, Art. no. 5600617.

[41] C. Yang, J. Xiao, Y. Ju, G. Qiu, and K. M. Lam, “Improving robustness of single image super-resolution models with monte carlo method,” in Proc. Int. Conf. Image Process., 2023, pp. 2135–2139.

[42] C. Yang, H. Luo, G. Liao, Z. Lu, F. Zhou, and G. Qiu, “Self-supervised video super-resolution by spatial constraint and temporal fusion,” in Proc. Pattern Recognit. Comput. Vis., 2021, pp. 249–260.

[43] J. Xiao et al., “Online video super-resolution with convolutional kernel bypass grafts,” IEEE Trans. Multimedia, vol. 25, pp. 8972–8987, 2023.

[44] J. Xiao, T. Liu, R. Zhao, and K. M. Lam, “Balanced distortion and perception in single-image super-resolution based on optimal transport in wavelet domain,” Neurocomputing, vol. 464, pp. 408–429, 2021.

[45] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, “A discriminative feature learning approach for deep face recognition,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 499–515.

[46] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “SphereFace: Deep hypersphere embedding for face recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 212–220.

[47] W. Liu, Y. Wen, B. Raj, R. Singh, and A. Weller, “SphereFace revived: Unifying hyperspherical face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 2, pp. 2458–2474, Feb. 2023.

[48] Facial Expression Recognition Method Based on Improved VGG Convolutional Neural Network Shuo Cheng (1Harbin Normal University, Harbin 150000, P. R. China) and Guohui Zhou (1Harbin Normal University, Harbin 150000, P. R. China) International Journal of Pattern Recognition and Artificial Intelligence 2020 34:07

[49] Bewoor, Mrs. Mrunal & Patil, Mrs.Sheetal & Kushwaha, Saumya & Tandon, Sparsh & Trivedi, Siddharth & Pawar, Avinash. (2023). Face recognition using open CV and VGG 16 transfer learning. AIP Conference Proceedings. 020019. 10.1063/5.0157084.