Facial Image Pre-Processing and Emotion Classification: A Deep Learning Approach

Alramzana Nujum Navaz*, Serhani Mohamed Adel †, and Sujith Samuel Mathew‡
*†College of Information Technology (CIT), UAE University, UAE
‡College of Technological Innovations (CTI), Zayed University, UAE
Email: *201570182@uaeu.ac.ae, †serhanim@uaeu.ac.ae, ‡sujith.mathew@zu.ac.ae

***Abstract*—Facial emotion detection and expressions are vital for applications that require credibility assessment, evaluating truthfulness, and detection of deception. However, most of the research reveal low accuracy in emotion detection mainly due to the low quality of images under consideration. Conducting intensive pre-processing activities and using artificial intelligence especially deep learning techniques are increasing accuracy in computational predictions. Our research focuses on emotion detection using deep learning techniques and combined pre-processing activities. We propose a solution that applies and compares four deep learning models for image pre-processing with the main objective to improve emotion recognition accuracy. Our methodology includes three major stages in the data value chain, pre-processing, deep learning and post-processing. We evaluate the proposed scheme on a real facial data set, namely Facial Image Data of Indian Film Stars for our study. The experimentation compares the performance of various deep learning techniques on the facial image data and confirms that our approach enhanced significantly the image quality using intensive pre-processing and deep-learning, improves accuracy in emotion prediction.**

Keywords—Emotion Detection, Facial Emotion, Deep Learning, Deep Neural Network, Image Pre-Processing, Image Enhancement, Accuracy Improvement.

I. Introduction

Emotion detection has great potential in video surveillance, monitoring infants, disabled and elderly people. Human facial expressions are natural and direct means to communicate the emotions and intentions, especially in non-verbal communications. With improved and advanced image detection and processing technologies, accurate emotion detection seems fairly achievable. In reality, images taken with different backgrounds, color settings and poses are challenging to the existing emotion detection algorithms. One other challenge, include images hidden with multiple emotions. However, deep learning technologies have made great progress in the field of emotion detection. Few mobile applications are also available that uses deep learning in emotion detection [1].

The performance of various facial emotion recognition techniques are compared based on the number of expressions recognized and the complexity of the algorithms. Six basic emotions are universally experienced in all human cultures. A smile on the human face reveals happiness, and the lips and eyes show a curved shape. In sadness, the face expresses looseness with rising skewed eyebrows and frown. The anger on the human face is expressed with squeezed eyebrows, slender and stretched eyelids. The disgust expressions are expressed with pull down eyebrows and creased nose while surprise is expressed with eye-widening and mouth gaping [2].

Many of the topmost countries, specially UAE is keen on promoting happiness as a national agenda [3]. To monitor and make progress in the achievement of a happier society, it is essential to use computer vision and facial emotion detection technologies. Although, this might have privacy issues, monitoring the workplace on a continuous basis and detecting sadness and depression can lead to early intervention and promote happiness. In addition, the application of emotion detection and classification in day-to-day aspects of the health-care sector can improve patient happiness and satisfaction. The application of deep learning techniques is used heavily in medical fields, especially psychology [4] and human behavior. Facial recognition applications are now widely used in mobile phones, smart cars and in the field of social robotics. In the future, the human happiness formula is likely to be linked with Artificial Intelligence and machine learning.

The rest of this paper is organized as follows: we will first review the most relevant literature in Section II, and then explain our methodology in detail in Section III. Extensive experiments are conducted in Section IV, to validate our approach on the chosen emotion dataset. We discuss the results and move on to Section V, where we draw the conclusions and mention future work.

II. Related Work

Developments within the Deep Learning Architecture has led to numerous advancements in the fields of image analysis and emotion detection in particular. Most of the works categorize facial expressions into six or seven such as anger, disgust, fear, joy, sadness, surprise, and contempt [1]. Research [5] discusses how an off the shelf Convolutional Neural Network (CNN) can be used as a tool for image processing and feature selection and to detect emotions. Two simple, yet effective deep learning-based methods are proposed here for image emotion analysis. The first method used off-the-shelf CNN features directly for classification, while the second method used a fine-tune CNN that is pre-trained on a large dataset and by experimentation, both the deep learning-based methods outperform traditional methods. A CNN based approach for the extraction of features pertaining to emotions from photographs and painting with advanced classification methods using a binary network was suggested by [6]. The

978-1-7281-5052-9/19/$31.00 ©2019 IEEE

proposed model consists of two parts: a binary positive-or-negative emotion classification network and a deep network for a specific emotion recognition. During the network training, an assisted learning strategy is introduced to boost the recognition performance.

A novel and advanced deployment of deep learning for the automatic detection of facial expressions from images was suggested by [7]. Their study was mainly focused on two aspects, firstly applying a classification based multi-label softmax loss approach on CNN's with various datasets, and secondly, it explored the effect of mismatch within the domain during cross-dataset training and testing. The CNN approach to learn a shared representation between multiple Action Units(AUs) directly from the input image was achieved. However, the data was imbalanced due to the sparsity of the AU occurrences. The work [8], also developed a new technique for weighing different classifiers for the training of a convolutional neural network unlike other conventional methods in the field of computer vision and image recognition which use averaging as the standard weighing option. Apart from using a standard convolutional neural network approach, a supervised committee of 72 individual CNN's were trained. The main advantage of the model, using different CNN's for emotion recognition is the fact that some of these may complement one another. However, they have mentioned that disgust and fear emotion were never classified correctly. The computational cost for testing is stated as 69.65 seconds per image. [9] proposed a Recurrent Regression Neural Network (RRNN) architecture which works on two different and conventional aspects of face regression i.e. based on both still images and videos. Use of RRNN enables the model to also take into consideration sequential images rather than single images.

A multi-cue fusion emotion recognition (MCFER) framework was developed by [10]. The model implemented a CNN and a bidirectional RNN in a cascaded fashion and resulted in modelling human emotion using three complementary aspects, i.e. texture, landmark, and audio signals. A hybrid CNN-RNN design combining the diverse modalities for emotion forecast was proposed by [11].

The authors in [12] proposed an emotion recognition system using auditory and visual modalities. They utilized a CNN network to extract features from the speech, while for the visual modality a deep residual network of 50 layers is used. They claim to have significantly better performance on the test set in comparison to other models using multiple modalities. Two different deep networks are used in [13], such as deep neural network and convolutional neural network and CNN has a better performance. The paper mentions the prediction of disgust label as poor and indicates that it could be due to fewer images in the database representing the emotion. The proposed network in the work of [14], inspired by GoogLeNet and AlexNet architectures consisted of two convolutional layers, each followed by max pooling and then four Inception layers. The Inception layers increase the depth and width of the network while keeping the computational budget constant. However, due to the use of data from all of the studied databases to train the deep architecture, which included images that do not conform to the database setting such as pose and lighting, accuracy's were affected.

TABLE I. IMAGE EMOTION CLASSIFICATION IN EXISTING WORKS NETWORK ARCHITECTURE/TOOLS DATASET COMPARISON

#	Reference	Network Architecture\Tools	Dataset
1	[6]	Deep learning framework - Caffe	IAPS- Subset, ArtPhoto Abstract Paintings
2	[5]	AlexNet	ImageNet, ArtPhoto. Own database FlickrEmotion
3	[8]	Supervised Committee of CNN	SFEW2.0 database EmotiW2015 3rd Emotion Recognition In the Wild
4	[12]	CNN	RECOLA database AVEC 2016 Emotion Recognition challenge
5	[13]	CNN Caffe and CudaConvnet2	CK+ database
6	[9]	Recurrent Regression Neural Network (RRNN) Caffe and CudaConvnet2	MultiPIE database
7	[7]	CNN Caffe	Extended Cohn-Kanade dataset (CK+) DISFA-Spontaneous dataset BP4D-Spontaneous dataset

A different direction in detecting emotional stress from EEG Signal using deep learning technologies is suggested in [15]. They utilized deep learning as the learning method to predict users stress feeling while listening to music.

The network architecture/tools and dataset used in existing works are tabulated in Table I. Appreciable accuracies were obtained by most methods and algorithms that used deep learning approaches. The multi-cue fusion model [10] achieved an overall accuracy of 49.92% which was a significant improvement of the baseline for the same. The work by [7] gave an average accuracy across all datasets of about 75% which was very close to state-of-the-art methods. It was shown that the multi-scale CNN outperformed other conventional CNN techniques and the ones which had shallow feature sets across all the different emotions, judging from the average true positive rates [5]. The assisted learning model which used a convolutional neural network approach [6] gave an overall accuracy of 64.6% on a random test set, which was significantly more than the benchmark accuracy of 58.3% as shown by them. The Recurrent Regression Neural Network model by [9] achieved an accuracy in the range of 92-95% in the classification of sequential images and videos. However, the hybrid approach developed by [11] was one of the best performers and had an accuracy of over 93%.

Although there are a lot of advancements in using deep learning, for image emotion detection, few research works addresses the image pre-processing challenges. To the best of our knowledge and considering the above-mentioned literature review, there are no research works that consider image pre-processing to improve prediction accuracy in facial emotion detection. Our work focuses on a methodical approach which includes enhancing the images using deep-learning for better accuracy in emotion prediction.

III. MULTI-DEEP LEARNING MODELS FOR IMAGE PRE-PROCESSING, EMOTION CLASSIFICATION AND PREDICTION

Deep learning is a machine learning technique that teaches computers to do what comes naturally to humans i.e. learn

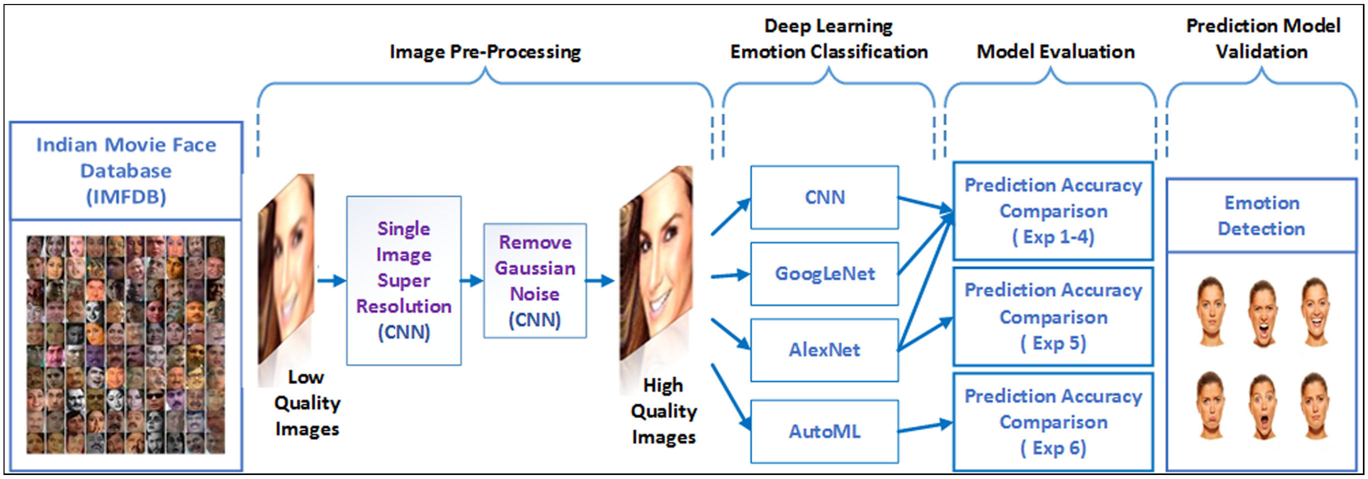

Fig. 1. Multi-Deep Learning Models For Image Pre-Processing, Emotion Classification and Prediction

by examples. To build deep learning models by skipping the annoying configuration and setup commands, we can use pre-configured deep learning configurations. Pre-trained image classification networks available have been trained on over a million images and can classify images into more than 1000 object categories. These networks have learned rich feature representations from a wide range of images. It takes an image as input and then outputs a label for the object with the probabilities for each of the object categories. In our work we have used, transfer learning. It can take a pre-trained network and use it as a starting point to learn a new task. Fine-tuning a network with transfer learning is usually much cheaper, easier and faster than training a network from scratch [16]. In a deep learning process, the aim is to have the accuracy of the model increase over time. As the network trains, the progress plot appears.

In this research, we have used a methodological deep learning approach following the data value chain as illustrated in Figure 1, which gives an overall view of the processes involved and the experimental scenarios. The detailed steps are described in the following sub sections.

A. Pre-processing

Pre-processing is a common name for the initial data handling operations and the processing to enhance images, when dealing with image data.

1) Image Data Handling:

Once the dataset of IMFDB is downloaded the first step is to do the pre-processing, to enable the image processing for emotion detection. The image files are placed inside the folder names (with actor names) which have the movie names. Each sub-folder has annotation file with the labeled features of each image. These include actor name, age, labelled emotion, gender, etc. Initially, a Perl file was written to segregate the images according to the emotion labelled in to a folder with the given emotion name. Later as part of fine-tuning the results, we had to manually segregate the images as some of them were wrongly labelled and the neural network algorithm got confused and took a long time for the training. Further, to improve the accuracy, we removed images with less resolution which was undertaken again using another Perl script.

2) Image Processing for Quality Improvement:

For the preprocessing of images, we followed several steps. Firstly, we utilized the Single Image Super-Resolution Using Deep Learning (SISR) [17] to create high-resolution images from the low-resolution images in our dataset. SISR is challenging because high-frequency image content typically cannot be recovered from the low-resolution image. Without high-frequency information, the quality of the high-resolution image is limited. It uses Very-Deep Super-Resolution(VDSR), a convolutional neural network architecture which is a 20-layer deep network, designed to perform single image super-resolution. In this process, we use a high-resolution reference image. A standard way to increase image resolution without deep learning is to use bicubic interpolation and the images in our database is upscaled using bicubic interpolation to match the size of the reference image. A residual image contains information about the high-frequency details of an image. The VDSR network detects the residual image from the luminance of a color image. The luminance channel of an image, Y, represents the brightness of each pixel through a linear combination of the red, green, and blue pixel values. In contrast, the two chrominance channels of an image, Cb and Cr, are different linear combinations of the red, green, and blue pixel values that represent color-difference information. VDSR is trained using only the luminance channel because human perception is more sensitive to changes in brightness than to changes in color. The steps involved are:-

1) Convert the low-resolution image from the RGB color space to luminance and chrominance channels.
2) Upscale the luminance and two chrominance channels using bicubic interpolation.

3) Pass the upscaled luminance component, through the trained VDSR network which results in the desired residual image.
4) Add the residual image to the upscaled luminance component to get the high-resolution VDSR luminance component.
5) Concatenate the high-resolution VDSR luminance component with the upscaled color components.
6) Convert the image to the RGB color space which results in the final high-resolution color image using VDSR.

Secondly, we removed image noise and created high-resolution images from low-resolution images, using convolutional neural networks. For our work, we used a pre-trained neural network to identify and remove Gaussian noise from images [18].

Finally, we adjusted image intensity values or color-map of the already enhanced image from the previous step, specifying the contrast limits. However, this step in image pre-processing seemed to improve the results only marginally.

B. Deep Learning for Emotion Classification

We performed our machine learning step by using deep learning networks, using the pre-trained neural networks which have been trained on over a million images. In our work, we use transfer learning which uses a pre-trained network (Alexnet, GoogLeNet and CNN), and use it as a starting point to learn the task of emotion detection. The idea is that this pre-trained model will act as a feature extractor and we remove the last layer of the network and replace it with our own classifier [19]. Fine-tuning a network with transfer learning is usually much faster and easier than training a network from scratch with randomly initialized weights.

AlexNet is a pre-trained CNN that has been trained on approximately 1.2 million images and the model has 23 layers. GoogLeNet [20], a 22 layers deep network, was used in our experiments, the quality of which is assessed in the context of classification and detection. The name is referred to the particular incarnation of the Inception architecture used in the model. A CNN passes an image through the network layers and outputs a final class. The network can have many layers, with each layer learning to detect different features. Filters are applied to each training image at different resolutions, and the output of each convolved image is used as the input to the next layer [21].

To further verify, if the image pre-processing gave better results, we tested our dataset with Google's Cloud Vision [22]. Cloud Vision offers the ability to build custom models using AutoML. After uploading and labeling images, Google's AutoML Vision, train the model and offers relatively high model accuracy in a shorter timespan.

C. Post-processing - Model Evaluation

This section refers to the evaluation of the identified classes following the deep learning exercise. In the field of machine-learning, a confusion matrix, a specific table layout that allows visualization of the performance of an algorithm is used. For an individual class C_i, the assessment is defined by tp_i; fn_i; tn_i; fp_i. $Accuracy_i$; $Precision_i$; $Recall_i$ are calculated from the counts for C_i where tp_i are true positive for C_i, and fp_i false positive, fn_i false negative, and tn_i true negative counts respectively. Quality of the overall classification in a multi-class problem is usually assessed in two ways:- Macro-averaging is a measure of the average of the same measures calculated for C_1-C_n and Micro-averaging, the sum of counts obtained by cumulative tp; fn; tn; fp and then calculating the performance measure. Macro-averaging treats all classes equally while micro-averaging favors bigger classes [23]. The accuracy and sensitivity indices correlated with the Emotion Perception is defined in [24]. Accuracy is positively related to Emotion Perception (r = .38, p b .01), whereas sensitivity had a negative correlation with Emotion Perception (r=.32, p b .05).

In our experiments, we have used Matlab to plot the Confusion Matrix, for each of the experiments conducted. The column on the far right of the plot shows the percentages of all the examples predicted that belong to each class that are correctly and incorrectly classified. These metrics are often called the precision (or positive predictive value) and false discovery rate, respectively. The row at the bottom of the plot shows the percentages of all the examples belonging to each class that are correctly and incorrectly classified. These metrics are often called the recall (or true positive rate) and false negative rate, respectively. The cell in the bottom right of the plot shows the overall accuracy [25].

IV. Experimentation and Results Discussion

In this section, we mention the experimental setup and tools, briefly describe the dataset and detail the experiments followed by the discussion of results.

A. Experimental Tools

Matlab provides a Deep Learning Toolbox (formerly Neural Network Toolbox), a framework for designing and implementing deep neural networks [26]. It allows importing pre-trained models including AlexNet, GoogLeNet, VGG-16, VGG-19, ResNet-101, Inception-v3, and SqueezeNet. The tool helps in visualizing the network topology and view details such as training parameters and activations.

B. Dataset

Indian Movie Face database (IMFDB) is a large unconstrained face database consisting of images of 100 famous Indian actors collected from more than hundred videos. The images are manually chosen and cropped from various video frames of released films resulting in high variation with respect to scale, pose, expression, illumination, age, resolution, occlusion, and makeup. IMFDB claims to be the first face database that provides a detailed annotation for every image. To ensure diversity movies are selected from 5 Indian languages namely, Hindi, Telugu, Kannada, Malayalam, and Bengali [27].

C. Experimental Scenarios

In this section, we identified and performed few experiment scenarios on our chosen dataset IMF database and evaluated the overall accuracy of the models derived. We divided the data into 70% training data and 30% validation data in all the below experiments.

1) Experiment 1. Classification of Emotions (6 Classes): The initial experiment was to classify the image emotions in to six categories namely Anger, Disgust, Happiness, Neutral, Sadness and Surprise. The overall accuracy, of emotion detection, is depicted in Figure 2, where Alexnet performed better than the other two network configurations. The confusion matrix plotted for Alexnet, the chosen deep learning pre-trained network, is shown in Figure 3.

Prediction Accuracy Comparison

Alexnet	GoogLeNet	CNN
43.7 %	44.72 %	23.5 %

Fig. 2. Overall Accuracy Comparison- Emotion Detection 6 Classes

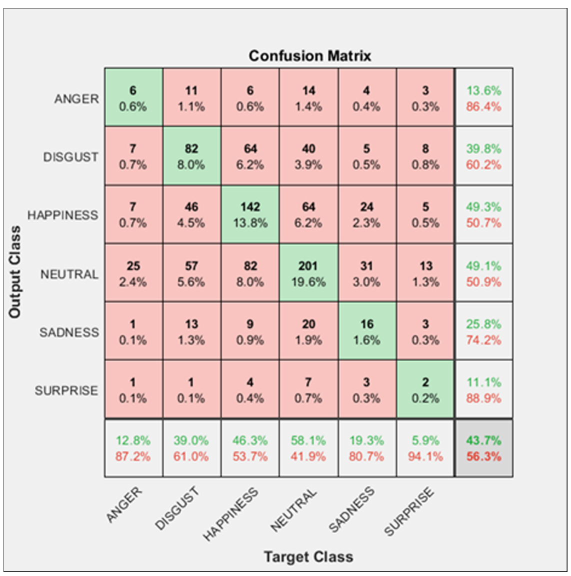

Fig. 3. The Plot of Confusion Matrix of Alexnet Deep Learning Network - Emotion Detection 6 classes

2) Experiment 2. Classification of Emotions (6 Classes) After manual labelling of images that were wrongly classified: We conducted the second experiment by manually segregating the images by visually verifying the image emotion. At this stage, we also eliminated low-resolution images using a Perl script. The experimental accuracy comparison on three different pre-trained networks is shown in Figure 4. The results shows a slight improvement in accuracy in all the three configurations, when we labelled accurately the emotion. This indicates, the importance of labeling the classes correctly, for accurate predictions.

Prediction Accuracy Comparison

Alexnet	GoogLeNet	CNN
57.9 %	47.14 %	29.5 %

Fig. 4. Overall Accuracy Improvement Compared to Exp.1 On Manual Labelling Of Images

3) Experiment 3. Classification of Emotions (4 Classes): From the previous experiments, we found that the emotions Disgust and Neutral were confusing within themselves, even for human eyes. So the experiments were conducted with these two emotions removed and the accuracy is depicted in Figure 5. Again the prediction accuracy has improved, indicating that narrowing the classes i.e removing classes that are unclear is essential in improving classification accuracy.

Prediction Accuracy Comparison

Alexnet	GoogLeNet	CNN
60 %	51.2 %	33.9 %

Fig. 5. Overall Accuracy Comparison - Emotion Detection 4 classes

4) Experiment 4. Classification of Emotions (4 Classes) After Image Enhancement: Here experiment 3 is repeated with image enhancement referred in Image Pre-processing. The accuracy obtained with the same settings in previous experiment, but with image enhancement is shown in Figure 6. The accuracy has improved significantly with image enhancement in all the neural network configurations.

Prediction Accuracy Comparison

Alexnet	GoogLeNet	CNN
67.6 %	62.5 %	59.42 %

Fig. 6. Overall accuracy improvement on Image Enhancement

5) Experiment 5. Classification of Emotions (4 Classes) Alexnet Varying Layers: With the enhanced images, this experiment was conducted with Alexnet, as it gave reasonably better results in our previous experiments. Here Weight Learn Rate Factor and Bias Learn Rate Factor are set to 10. The network constructs a hierarchical representation of the input images and deeper layers contain higher-level features. To get the feature representations of the training and test images, we used activation's on the layers 23, 8 and 14 and compared the accuracy's as shown in Figure 7, where layer 14 had the highest accuracy.

Prediction Accuracy Comparison

Layer 23	Layer 8	Layer 14
23 'fc8' Fully Connected 1000 fully connected layer	8 'norm2' Cross Channel Normalization cross channel normalization with 5 channels per element	14 'conv5' Convolution 256 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
65 %	**67.6%**	**70.6 %**

Fig. 7. Alexnet - Activations on Varying Layers

6) Experiment 6. Classification of Emotions - Verification of Exp. (1-4) using Google Cloud Vision:
Our experiments were repeated in Google's Cloud Vision AutoML, and verified for improvement of accuracy following our methodical approach. The prediction accuracies of the experiments conducted in AutoML are compared in Figure 8. With image enhancement, we find an approximate accuracy improvement of 10%, while image deblurring exercise only had a marginal improvement in emotion detection.

Prediction Accuracy Comparison

AutoML	AutoML With Manual Labelling	AutoML Narrowing Classes	AutoML With Image Enhancement	AutoML With Image Enhancement & Deblur
54.8 %	**59.5**	**71.5 %**	**80 %**	**81.2 %**
Precision 64.024%	Precision 67.647%	Precision 75.0%	Precision 80.392%	Precision 72.917%
Recall 30.435%	Recall 38.983%	Recall 64.286%	Recall 65.079%	Recall 62.5%

Fig. 8. Comparison Of Model Detection Accuracy in Google Cloud Vision

D. Evaluation and Validation of Prediction Model (Google Cloud Vision)

After training the model, AutoML Vision uses images from the TEST set to evaluate the quality and accuracy of the new model. It provides the confusion matrix, which is an aggregate set of evaluation metrics indicating how well the model performs overall, as well as evaluation metrics for each category label, indicating how well the model performs for that label [28].

The confusion matrix plotted is shown in Figure 9, and represents the percentage of times each emotion was predicted for each emotion in the training set during evaluation. The confusion matrix indicates that the model has less accuracy (33.3%) in detecting the emotion anger, while happiness (89.3%) was the best accurately detected emotion.

We validated the accuracy of our prediction model, by using the predict method with online images of famous actors depicting facial emotions. The predict method applies labels to the given image based on the primary object of the image that the model predicts. The prediction results are depicted in Figure 10, Figure 11, Figure 12 and Figure 13. This again illustrates that predictions of the model in detecting the emotion anger Figure 13 (57.5%), is less compared to the predictions in happiness Figure 10 (97.4%), sadness (93.5%), and surprise (94%).

True label \ Predicted label	sadness	anger	surprise	happiness
sadness	64.3%	21.4%	-	14.3%
anger	16.7%	33.3%	33.3%	16.7%
surprise	-	-	80.0%	20.0%
happiness	3.6%	-	7.1%	89.3%

Fig. 9. Confusion Matrix Of The Final Google Vision AutoML Model. Blue indicates how often the model classified each emotion correctly, while the orange label shows how often it was confused.

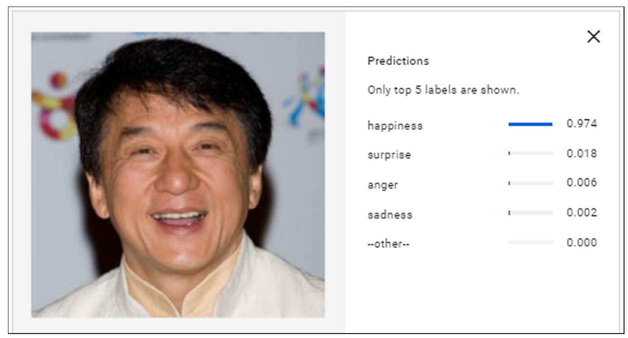

Fig. 10. Testing Model Accuracy In Predicting Happiness

Fig. 11. Testing Model Accuracy In Predicting Sadness

E. Discussion

In this study, we aimed to increase facial emotion recognition accuracy. In our experiments (1-4) we used, a systematic approach to increase accuracy in three of the neural network configurations namely Alexnet, GoogLeNet, and CNN. The steps followed and the accuracy improvements achieved are:-

1. Manual labeling of Images (Alexnet - 43.7 to 57.9 % $(\approx 14\% increase)$).
2. Removal of confusing emotions, thereby narrowing the classes (GoogLeNet - 47.1 to 51.2 % $(\approx 4\% increase)$).
3. Image Enhancement using deep learning techniques

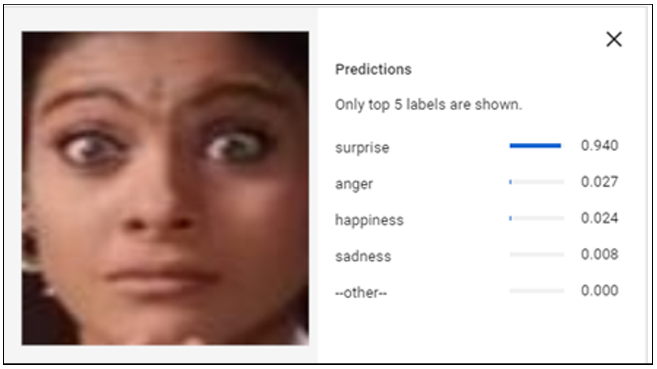

Fig. 12. Testing Model Accuracy In Predicting Surprise

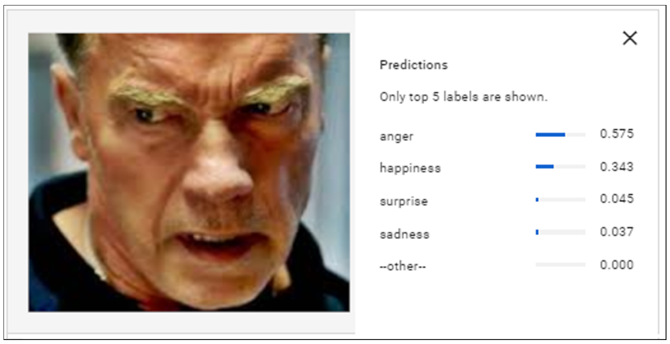

Fig. 13. Testing Model Accuracy In Predicting Anger

(CNN 33.9 to 59.42 % ($\approx 26\% increase$)).

Image enhancement followed at the pre-processing stage using SISR and the removal of Gaussian noise had a greater impact on accuracy improvement, in all the three selected network configurations.

The maximum overall accuracy improvement from experiments (1-4) was identified in CNN, from 23.5 to 59.4 % ($\approx 36\% increase$).

Experiment 5, highlighted the prediction accuracy improvement by varying the layer activations. An overall ($\approx 5\% increase$) is experienced by altering from Layer 23 to Layer 14 activation. Ideally, the neural network configuration can only be fixed on a trial and error basis. A more deeper or fully connected network (Layer 23 in our experiment), may not give the optimum prediction as seen from our accuracy results on varying layers in Figure 7.

In experiment 6, we followed the same steps as in experiments (1-4), but using the latest technology in cloud, Google's Cloud Vision AutoML. The accuracy improved from (54.8 % to 81.2 % ($\approx 26\% increase$)), which reconfirms that our methodical approach can indeed improve emotion detection accuracy.

Figure 14 depicts the summary of the experimental accuracy improvements achieved. There is a steady improvement in accuracy from exp.1 to exp.4 in all the three chosen neural network configurations. Exp.4 which included image enhancement, improved and made the accuracies of GoogleNet and CNN, in par with Alexnet. Similarly, the AutoML experiments, which followed the experiments(1-4) attained a steady improvement in prediction accuracy as depicted by the green trend line in the graph.

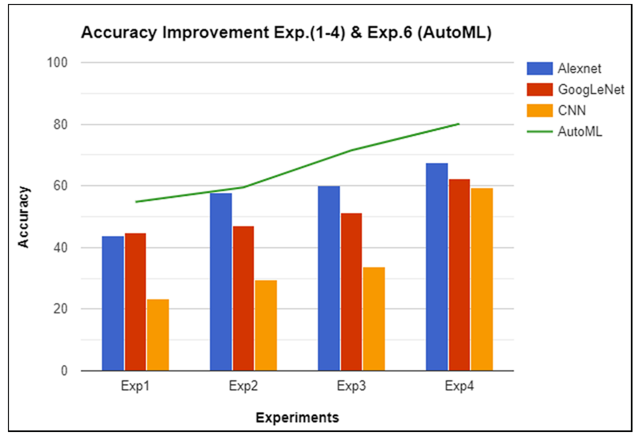

Fig. 14. Overview of Prediction Accuracy Improvements Attained

Our experiments also reveal that designing neural networks is extremely time consuming. The usage of pre-trained networks is better, but again the selection of the network configuration to optimize prediction accuracy is tedious. One of the best approaches is AutoML, where it is possible for the neural network to design the best possible neural network, according to the data. We have experimentally found that the AutoML prediction accuracy ($\approx 20\%$), better than the pre-trained neural networks.

One of the challenges in dealing with emotion detection from images is that, training a deep learning network on images requires more computation time as compared to textual and conventional databases. As we used images from a database which is extracted from videos, it has added blurriness to the images. Also, the images varied in pose and lighting and therefore required manual intervention to discard images that do not comply with the quality constraint. However, the database chosen was apt for our study, as the aim is to improve facial emotion detection accuracy.

V. Conclusion and Future Work

In reality, Artificial Intelligence can never detect emotion with 100% accuracy. This work presents Deep neural network architectures for image pre-processing and emotion detection. We have experimentally proven that image pre-processing can improve emotion detection accuracy. In addition, improving accuracy of emotion detection by fine-tuning the pre-trained neural networks was achieved. During the initial runs, the deep learning network training was time-consuming, which means that the network is unable to converge on a solution. However, the image pre-processing techniques fastened the deep training.

Only a small portion of the IMF database was considered for this study and hence would like to extend our work to include the full dataset. Facial emotion recognition pertaining to real-time videos would be another interesting challenge to explore. One more research direction is in combining emotion detection with EEG brain waves which might serve the medical field. We would also be interested to explore TensorFlow, an open-source software library which is trending in machine learning and develop mobile apps for emotion detection.

REFERENCES

[1] H.-J. Lee and K.-S. Hong, "A study on emotion recognition method and its application using face image," *2017 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 370–372, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/8191005/

[2] I. M. Revina and W. R. Emmanuel, "A Survey on Human Face Expression Recognition Techniques," 2018.

[3] "UAE Government of Future Happiness," 2018, https://government.ae/en/about-the-uae/the-uae-government/government-of-future/happiness.

[4] A. A. Moustafa, T. M. O. Diallo, N. Amoroso, and N. Zaki, "Applying Big Data Methods to Understanding Human Behavior and Health," *Frontiers in Computational Neuroscience*, vol. 12, no. October, pp. 10–13, 2018.

[5] M. Chen, L. Zhang, and J. P. Allebach, "Learning deep features for image emotion classification," *Image Processing (ICIP), 2015 IEEE International Conference on*, pp. 4491–4495, 2015.

[6] X. He and W. Zhang, "Emotion recognition by assisted learning with convolutional neural networks," *Neurocomputing*, vol. 291, pp. 187–194, 2018.

[7] S. Ghosh, E. Laksana, S. Scherer, and L. P. Morency, "A multi-label convolutional neural network approach to cross-domain action unit detection," *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, pp. 609–615, 2015.

[8] G. Pons and D. Masip, "Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis," *IEEE Transactions on Affective Computing*, vol. 9, no. 3, pp. 343–350, 2017.

[9] Y. Li, W. Zheng, Z. Cui, and T. Zhang, "Face recognition based on recurrent regression neural network," pp. 50–58, 2018.

[10] J. Yan, W. Zheng, Z. Cui, C. Tang, T. Zhang, and Y. Zong, "Multi-cue fusion for emotion recognition in the wild," pp. 27–35, 2018.

[11] M. Z. b. Neha Jain a , , Shishir Kumar a , Amit Kumar a , Pourya Shamsolmoali b , c, "Hybrid deep neural networks for face emotion recognition," *Computers and Fluids*, vol. 156, pp. 579–589, 2017.

[12] P. Tzirakis, G. Trigeorgis, M. A. Nicolaou, B. W. Schuller, and S. Zafeiriou, "End-to-End Multimodal Emotion Recognition Using Deep Neural Networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 8, pp. 1301–1309, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/8070966/

[13] H. Jung, S. Lee, S. Park, B. Kim, J. Kim, I. Lee, and C. Ahn, "Development of deep learning-based facial expression recognition system," *2015 Frontiers of Computer Vision, FCV 2015*, pp. 2–5, 2015.

[14] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–10, 2016. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7477450

[15] C.-y. Liao, R.-c. Chen, and S.-k. Tai, "Emotion stress detection using EEG signal and deep learning technologies - IEEE Conference Publication," *2018 IEEE International Conference on Applied System Invention (ICASI)*, no. 2, pp. 90–93, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8394414/

[16] Matlab. (2018). [Online]. Available: https : //in.mathworks.com/help/deeplearning/examples/train-deep-learning-network-to-classify-new-images.html

[17] "Single image super-resolution using deep learning- matlab & simulink," 2018. [Online]. Available: https://www.mathworks.com/help/images/single-image-super-resolution-using-deep-learning.html

[18] (2018). [Online]. Available: https://in.mathworks.com/help/images/deep-learning.html

[19] A. Deshpande, "A beginner's guide to understanding convolutional neural networks part 2," 2018.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Computer Vision and Pattern Recognition (CVPR)*, 2015. [Online]. Available: http://arxiv.org/abs/1409.4842

[21] Mathworks, *Practical Deep Learning Examples with MATLAB*. The MathWorks Inc, 2018.

[22] "Cloud vision api," 2018. [Online]. Available: https://cloud.google.com/vision/

[23] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management*, vol. 45, no. 4, pp. 427–437, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.ipm.2009.03.002

[24] D. Lyusin and V. Ovsyannikova, "Measuring two aspects of emotion recognition ability: Accuracy vs. sensitivity," *Learning and Individual Differences*, vol. 52, pp. 129–136, 2016.

[25] "Plot classification confusion matrix - matlab," 2018. [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/plotconfusion.html

[26] "Matlab - Deep Learning Toolbox," 2018. [Online]. Available: https://www.mathworks.com/products/deep-learning.html

[27] P. B. J. G. M. K. R. V. V. H. J. C. K. R. R. R. V. K. Shankar Setty, Moula Husain and C. V. Jawahar, "Indian Movie Face Database: A Benchmark for Face Recognition Under Wide Variations," in *National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Dec 2013.

[28] "Google Cloud. (2019). Evaluating models — Cloud AutoML Vision." [Online]. Available: https://cloud.google.com/vision/automl/docs/evaluate