Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 113–123

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Smart Facial Emotion Recognition With Gender and Age Factor Estimation

Surya Teja Chavali[a, *], Charan Tej Kandavalli[b], Sugash T M[c], Subramani R[d]

[a-c]*Department of Computer Science and Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, Bengaluru – 56003, India*

[d]*Department of Mathematics, CHRIST (Deemed to be University), Bengaluru 560029, India*

Abstract

Human-Computer Interaction (HCI) in an intelligent way, which aims at creating scalable and flexible solutions. Big tech firms and businesses believe in the success of HCI as it allows them to profit from on-demand technology and infrastructure for information-centric applications without having to use public clouds. Because of its capacity to imitate human coding abilities, facial expression recognition and software-based facial expression identification systems are crucial. This paper proposes a system of recognizing the emotional condition of humans, given a facial expression, and conveys two methods of predicting the age and gender factors from human faces. This research also aims in understanding the influences posed by gender and age of humans on their facial expressions. The model can currently detect 7 emotions based on the facial data of a person - (Anger, Disgust, Happy, Fear, Sad, Surprise, and Neutral state). The proposed system is divided into three segments: a.) Gender Detection b.) Age Detection c.) Emotion Recognition. The initial model is created using 2 algorithms - KNN, and SVM. We have also utilized the architectures of some of the deep learning models such as CNN and VGG - 16 pre-trained models (Transfer Learning). The evaluation metrics show the model performance regarding the accuracy of the Recognition system. Future enhancements of this work can include the deployment of the DL and ML model onto an android or a wearable device such as a smartphone or a watch for a real-time use case.

Keywords: *HCI; Emotion recognition; Gender and Age recognition; SVM; KNN; CNN; VGG - 16*

1. Introduction

Face recognition and Face differentiation have always been innate capacities in humans. Computers can now perform the same thing. This brings up a slew of human computers. Face detection and recognition can be used to improve access and security, process payments without actual cards, identify criminals, and provide individualized healthcare and other services. Face recognition is undoubtedly the most debated topic over extensive research in recent decades, owing to rising security needs and its potential in commercial and law enforcement applications.

* Corresponding author. Tel.: +91-8639439871
E-mail address: bl.en.u4aie19014@bl.students.amrita.edu

1877-0509
10.1016/j.procs.2022.12.407

Face recognition has a wide range of use cases, including human-computer interaction, content-driven coding of image video, Biometric Analysis, and so on. It has the potential to be a non-intrusive biometric identification method that does not require human collaboration. In this paper, well-known machine learning algorithms such as K-nearest neighbors (KNN), and Support Vector Machine (SVM) have been used, and also popular deep learning techniques like CNN and VGG-16 with transfer learning have also been used to perform emotion recognition along with the estimation of gender and age of a person from the facial image.

Age Detection provides additional data about an individual's identification. Being able to distinguish people based on their age might be beneficial in product sales. You might personalize adverts in stores based on who is watching or exiting the store. It could also be used to reduce the number of elements studied in a database if they are first filtered by age, cutting a search time from minutes to seconds or even minutes. In the forensic sector, where a suspect's description is frequently all that is given, an age range is usually the only data that is always accessible, which would be very valuable in narrowing down the list of candidates to be compared with the sketch to get more exact answers. The same deep learning and machine learning algorithmic techniques that were used for facial recognition have also been used for age and gender identification.

Many studies on human-machine emotional communication have been identified as one of the most critical issues in the interaction between computers and humans. Image classification is frequently researched upon [17, 1] and machine learning techniques are employed over to build robust applications to help solve some challenging real-world scenarios. Though many studies previously on Statistical models have been considered for multi-class problem statements such as the two – dimensional Hidden Markov models [9, 11], modern deep learning classification algorithms outperform in a larger bias. As part of human-computer interactions, the emotional relationship between humans and machines has gotten a lot of attention. And, among humans, emotional expressions are the most potent and natural form of communication. Significant head motion, temporal patterns, and partial occlusions are common features of spontaneous interactions' facial expressions. However, utilizing facial expression to characterize human emotional states is problematic since these facial expression traits are sensitive to lighting circumstances, dynamical head motion, and other factors [2]. It's also difficult to categorize a specific feeling because they're so diverse. People can develop simulations of facial emotions that are characteristic of each unique emotion in reaction to similar facial expressions that occur in response to certain emotional- circumstances. For emotion recognition also, previously considered machine learning and deep learning models were analyzed.

There are a variety of real-life applications for Emotion recognition, some of them are:

- Threat–proof systems of security
- Interactive simulations and designs of modern computers
- Psychology and Computer Vision
- Driver Fatigue Monitoring
- Video Games

This paper's organization is as follows – In the next section, we analyze the existing works and compare the models with our development. In section 3, we introduce the FER and UTKFace datasets used for building the models in this work. In sections 4 and 5, we detail all the implementation steps and the proposed methodologies, along with the analysis of the experimental results obtained thereafter. We list some important conclusions derived and discuss the future scope of the work portrayed, in the final section.

2. Related Works

Several different works on robust face detection and recognition, involving OpenCV tools have been developed in recent years, to eliminate the effects of privacy threats and potential theft of confidential data. In [12, 13], the authors have used a 3-step approach to explain facial recognition algorithms, involving - face detection, face extraction, and face recognition. The authors in [13] have presented a generic approach to finding faces in 3 different spaces of finding faces in – color, motion, and limited area of pixels in a frame of a live video feed, and developed and implemented their model using MATLAB software.

Researchers in [5] have proposed two distinct architectures of deep CNN models and aimed at utilizing distributions as the training tasks can take advantage of the uncertainty brought on by manual labeling to create a stronger model, instead of using ages as the goal of the research. In this work, we have also detailed the approaches using classical machine learning techniques using Support Vector Classifier (SVC) and K – Nearest Neighbor (KNN). The authors have used a competition dataset called the ChaLearn dataset [6]. On the contrary, we have made use of the FER and the UTKFace datasets, which are well-studied datasets before making them publicly accessible and are more complicated datasets, with human accuracy only reaching 65±5%, and the best-reported efforts getting a test accuracy of 75.2% (as mentioned in one of the Stanford University's winter project reports for the year 2020 in the CS230 Deep Learning course).

In [7], the work portrays the implementation of SVM, Deep Boltzmann Machine (DBM), and Fusion Methods [8]. The authors achieved an average F1 score of 0.23 and 0.03 using SVC and DBM classification techniques respectively. The models developed in our work achieved an overall weighted average F1 score of 0.41 using the SVM approach and 0.33 using the KNN technique.

3. Datasets

3.1. Facial Emotion Dataset

There are multiple international datasets utilized in different research works on the topic of emotion recognition. We have used one such dataset known as the FER – 2013 dataset which can be found on the Kaggle repository (https://www.kaggle.com/datasets/msambare/fer2013). The images in this dataset have been clustered into 7 different directories, which are:

- Happy
- Disgust
- Natural
- Angry
- Fear
- Sad
- Surprise

Fig. 1. A glimpse of the emotion dataset.

Each of these categories consists of nearly 1000 images, each being 48 (x) 48 size and formatted in. ".png" format. These faces were automatically recorded such that they are almost cantered in each picture and take up around the same amount of area.

This facial data has a total of 40 facial attributes (as shown in Fig. 2) ranging from the type of eyebrows to ornaments and wearables worn over the faces. Fig. 2 illustrates the frequency (standardized between 0 and 1) of all the 40 attributes of the faces in the FER dataset, where the images with “No_beard” and “Bald” are the highest and least in number respectively.

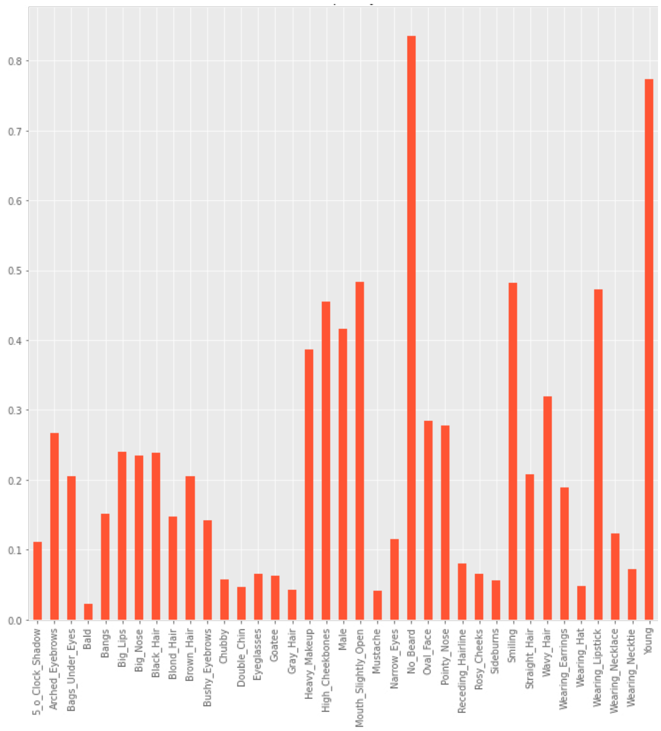

Fig. 2. Frequency of Facial Attributes across the entire FER – 2013 dataset

3.2. Age and Gender Dataset

For the age and gender identification part of the model, we have considered the UTKFace dataset which can be accessed through this web link – (https://susanqq.github.io/UTKFace/). It is a huge substantial facial dataset with a wide age stretching from 0 to 116 age groups of human faces. The compilation of image data consists of more than 20,000 facial photographs with classifications for age groups, ethnic groups, and gender. The images feature a variety of stances, facial gestures, and expressions, conditions illumination, backlighting, clarity, etc. This collection could be used for a variety of tasks, such as face detection and recognition, age estimation, age progress monitoring, etc.

Fig. 3. A sample of the UTKFace dataset.

4. Proposed Methodology

4.1. Support Vector Machine (SVM)

SVM belongs to the group of non–linear classifiers, that uses a kernel transformation matrix to convert the input feature vectors into a generally higher dimensional feature space. The dividing plane seen between borders of 2 categories is optimally placed to achieve maximum classification. The support vectors cover the plane, reducing the number of references. Input vectors from SVM are mapped to a higher-dimensional vector space, where a maximal separation hyperplane is built [15]. On either side of the hyperplane that divides the data, two parallel hyperplanes are built. The hyperplane that optimizes the separation between the two is known as the decision boundary or the separating hyperplane.

For example, we have a set of points $x_1,\ x_2,\ x_3, \dots \dots, x_n$, where $x_i\ \in R^n$. Our objective is to find the equation and model a hyperplane that separates the set points on the 3D plane. If our data points were linearly separable, then there exists a relationship:

$$a_i\ [(p.x_i) + b] \geq 1 \quad (1)$$

Therefore, $(\boldsymbol{p}.\boldsymbol{x_i})$ will form our hyperplane. This plane is referred to as the diving hyperplane, which is then formulated using an optimal objective function. In the implementation part of this work, we have incorporated the linear kernel in the SVM classifier while we train the model on our data points. The formulation shown above is[10]. This work explores the different kernels in SVC.

4.2. K – Nearest Neighbors (KNN)

The k-NN method is a basic and intuitive classification approach used by researchers to distinguish feature vectors. This classifier relates a newly labelled sample (validation) to the training data and makes a decision based on the results. Class labels are included in the training data set itself and the algorithm discovers the k i.e., the nearest neighborhood in the learning data set for a given value from the data set.

The algorithm then allocates a class that is frequently seen in the area. The algorithm can be summarized as follows:

- A training data set, which is a combination of input and class variables, is used to accomplish the k-nearest neighbour classification.
- Then assess test data that only comprises input variables to the reference set.
- By analyzing closest neighbour points, K-NN operates on k patterns, where the proximity of unknown k decides its category.
- A majority voting mechanism is employed, with each class receiving one point for each instance in the surrounding samples.

In this work, the value of k has been considered as 8 in the case of Gender and Age estimation and 5 in the case of Emotion recognition and the optimal value of k has been decided based on the performance of the algorithm.

4.3. Convolutional Neural Network (CNN)

Convolutional neural networks (CNNs) are a type of Artificial neural network (ANNs) and are supervised learning algorithms, used in deep learning to evaluate visual input data. The foundation of CNNs (Shift Invariant or Space Invariant Artificial Neural Networks) is a shared-weight structure of convolution kernels or filters that travel along data instances and produce feature mappings, which are translation-equivariant results. Interestingly, rather than being invariant under translation, most convolutional neural networks are merely increasingly adaptable and the equivalent [4]. Only a few of the use cases include image and video classification and identification, content marketing (systems for recommendations), picture recognition, edge detection, medical image processing, natural language processing, APIs for brain-computer communications, and financial time series. The architecture of CNNs is comprised of three layers – input, hidden, and output. Because the activation function and very last convolution cover their inputs and outputs, any intermediate layers in a feed-forward neural net are referred to as hidden. Convolutional layers are among the artificial neural network's hidden layers. The number of parameters is calculated in the following way:

$$(\text{No. of inputs}) * (\text{Input Height}) * (\text{Input Width}) * (\text{No . of inputs channels}) \quad (2)$$

To understand this, let us consider we give the input to the Conv2D layer a facial image. During the initial phases, the grayscale version of this image is considered for the layover of a suitable filter, which then processes the pixel intensity values of this image based on the given stride value. A portion of this scalable image, where the area extracted is only of an eye, hair, mouth, etc., is then passed onto a layer of max pooling for further size reduction, extracting only the important information out of those pixel values.

A convolutional operation is usually carried out by laying the flipped filter over the image and then finding the sum of dot products of the corresponding elements of the filter and the image. A feature map is built of out the above process and the Relu activation function is used to discard all of the negative values. In max pooling, the highest values of each local group of neurons in the feature map are used, whereas, in average pooling, the average value is used. Each neuron in a neural network calculates a result by implementing a function to the input parameters acquired from the previous layer's input patch. The functional technique that is applied to the input values is determined by a vector of weights and a bias (typically real numbers). Learning is all about iteratively adjusting these values of biases and weights [17, 18].

4.4. VGG – 16 (Transfer Learning)

VGG16 is a CNN, which is universally acknowledged as one of the most advanced computer vision models currently available. The creators of this model analyzed the networks and enhanced the layer depth using a framework with narrow (3, 3) convolution kernels, which outperformed previous-art setups significantly. They extended the weight layer depth to 16–19, yielding 138 trainable parameters.

The 16 in VGG16 stands for 16 weighted layers. VGG16 has twenty-one layers in total, including thirteen convolutional layers, five max-pooling, three classification (dense) layers, and only sixteen weight layers, or trainable parameter layers. VGG16 has three RGB channels and a 224, 244 input tensor. The biggest distinguishing aspect of VGG – 16 is that instead of having a large number of hyper-parameters, they concentrated on having 3x3 kernel convolution layers with stride 1 and only used the identical padding and max pool layer of 2 x 2 kernel with a stride value of 2. Throughout the architecture, the convolution and maximum pool layers are grouped in a similar fashion. There are 64 filters in the Conv-1 Layer, 128 filters in the Conv-2 Layer, 256 filters in the Conv-3 Layer, and 512 filters in the Conv 4 and Conv 5 Layers. Following a stack of convolutional layers, three fully-connected (FC) layers are added: the first two have 4096 channels apiece, while the third performs 1000-way ILSVRC classification and so has 1000 channels (one for each class) and the soft-max layer is the final layer.

However, since the model has been trained on the dataset known as 'ImageNet', and the last layer where the classification has been done has many different neurons because of the number of classes in the ImageNet dataset contains. Therefore, we have removed the last dense layer (ANN) part and added extra hidden layers for our dataset, and lastly added a dense layer with 4 neurons for our classification task. The above process is referred to as Transfer Learning. A piece of detailed information on the architecture of VGG – 16 can be found in [19, 20].

5. Experimental Results and Analyses

All the machine and deep learning models were developed in Jupyter piece of detailed information Notebook installed on the machine running on Intel core i7 processor and a memory capacity of 16 GB. Eight different Python notebooks were developed using the necessary modules and libraries [3], for Gender and Age recognition as one set and Facial Emotion recognition as another set, consisting of all 4 algorithms in each set.

Table 1. Accuracies achieved when different kernels have been employed under SVC for emotion recognition

Kernel	Accuracy of the model (in %)
Linear	36.074766
Polynomial (degree)	38.878504
Radial Basis Function (RBF)	45.607476
Sigmoid	40.74098

Table 2. Accuracies achieved when different kernels have been employed under SVC for gender and age identification

Kernel	Accuracy of the model (in %)
Linear	82.455
Polynomial (degree)	83.125
Radial Basis Function (RBF)	86.579
Sigmoid	83.125

We observe that the RBF kernel's performance is the best due to its capacity to handle non–linear data. Though polynomial kernel could process such type of data at higher degrees, it could easily lead to overfitting of the model due to the creation of a more flexible decision boundary. As a reason, the default value of the degree in that kernel is prescribed to be as low as '3'. Due to the high resemblance of the RBF Kernel to the K-Nearest Neighborhood Algorithm, it is widely used over all other kernels in the Support Vector Classifier (SVC). Since RBF Kernel Support Vector Machines only need to store the support vectors throughout the training and not the complete dataset, it has the advantages of K-NN and solves the space complexity problem.

The F1 score plays a vital role in determining the model performance against classifying data, as it is the harmonic mean of both precision and recall. Since both the datasets are imbalanced in terms of the number of images (indirectly an imbalance in the number of vector spaces while preparing the data for model training, the Macro Average of the F1 score has been taken into account. Also, since all the classes contribute equally to our output, it is viable to consider the Macro Avg. F1 score for analyzing the performance of the models. Tables 3 and 4 show the classification analyses for both the problems including all the important performance metrics when the RBF kernel in SVC was incorporated.

Table 3. Classification results of emotion recognition when RBF kernel was employed

Emotion Type	Precision	Recall	F1 score	Support
Anger	0.33	0.50	0.40	82
Disgust	1.00	0.05	0.10	40
Fear	0.40	0.14	0.21	58
Happy	0.51	0.81	0.63	141
Neutral	0.38	0.09	0.14	57
Sad	0.42	0.41	0.42	73
Surprise	0.53	0.52	0.53	84
Accuracy			0.46	535
Macro Avg	0.51	0.36	0.34	535
Weighted Avg	0.49	0.46	0.41	535

In developing the deep CNN architecture for simultaneous classification of age and gender features in an image, a total of five 2D convolutional layers, along with 3 max-pooling layers, a global – average pooling layer, 4 dropout layers, and 10 dense layers for the final ANN part (classification phase) of the model have been incorporated. Tables 3 – 6 show the respective results of the models obtained when the best kernel out of the four kernels used in SVC.

Table 4. Classification results of gender identification when RBF kernel was employed

Emotion Type	Precision	Recall	F1 score	Support
Male	0.90	0.80	0.85	93
Female	0.76	0.88	0.81	67
Accuracy			0.83	160
Macro Avg	0.83	0.84	0.83	160
Weighted Avg	0.84	0.83	0.83	160

Table 5. Classification results of emotion recognition using the KNN algorithm

Emotion Type	Precision	Recall	F1 score	Support
Anger	0.21	0.41	0.28	82
Disgust	0.24	0.25	0.24	40
Fear	0.27	0.10	0.15	58
Happy	0.46	0.56	0.51	141
Neutral	0.20	0.14	0.16	57
Sad	0.44	0.23	0.30	73
Surprise	0.45	0.31	0.37	84
Accuracy			0.34	535
Macro Avg	0.32	0.29	0.29	535
Weighted Avg	0.35	0.34	0.33	535

The similarity between results obtained from Tables 3 and 4 is due to the relation between KNN and the RBF kernel since both the techniques are classified as non-parametric techniques for calculating the density function or the probability distribution of our data.

Table 6. Classification results of gender identification when RBF kernel was employed

Emotion Type	Precision	Recall	F1 score	Support
Male	0.70	0.86	0.78	72
Female	0.86	0.70	0.78	88
Accuracy			0.78	160
Macro Avg	0.78	0.78	0.78	160
Weighted Avg	0.79	0.78	0.78	160

Table 7. Accuracies achieved by the models when different algorithms were used.

Model	Emotion (in %)	Gender (in %)	Age (in %)
SVM	36.074766	82.5	66.789
KNN	34.018691	76.25	60.086
CNN	84.63	90.000	91.17
VGG – 16	97.834	98.246	95.31

Upon close observation of the two algorithms, we notice that both of them use different approaches to the same issue: KNN establishes the number of points, 'k', and then identifies the region in space that contains those points, while kernel techniques establish the neighborhood size (h) and then compute K. We observe the robustness of deep learning models over the traditional machine learning models in terms of their performance. Although the performance shown by VGG – 16 is the best amongst all, SVM and KNN have shown good results, as they are one of the basic classifiers available. Furthermore, feature extraction performed is highly accurate in the case of deep learning-based models due to data compression as a result of the convolution operation, thus the error-susceptibility is also high in the case of traditional machine learning approaches. Figures 4 and 5 show the training (vs) validation parameters across all epochs

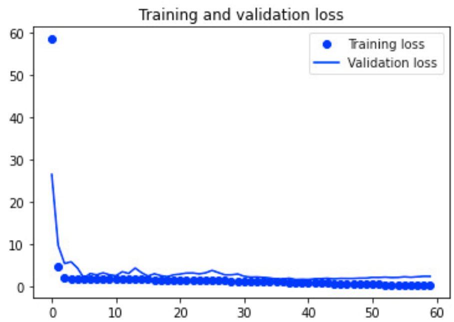

Fig. 4. Loss values of the CNN model for Emotion Recognition

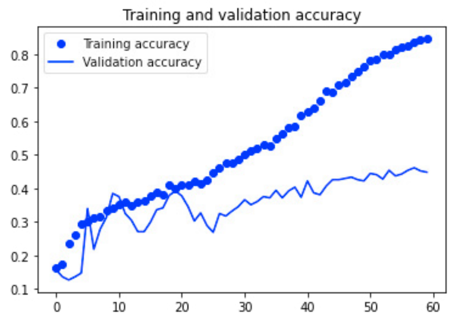

Fig. 5. Accuracies values of the CNN model for Emotion Recognition

for the CNN model for Emotion Recognition. As we observe closely, although training and validation accuracies improve as the epochs keep increasing, the loss values for training and validation decrease and then remain constant for the rest of the training.

Table 8. Score values achieved when different models were validated on the datasets

Model	Recognition Type	Epochs	Training Loss	Validation Loss	Training Accuracy (in %)	Validation Accuracy (in %)
CNN	Emotion	60	0.1578	0.4131	88.63	81.86
VGG – 16	Emotion	12	0.0739	0.2002	95.31	93.91
CNN	Age & Gender	40 (early stopping at 32)	0.0242	0.3487	84.71	86.50
VGG – 16	Age & Gender	5	0.0722	0.2614	98.24	95.31

The drastic decrease in the number of epochs in the case of training the VGG – 16 models can be understood by the reason that the default trainable weight parameter is set to the imported ImageNet weights, on which the original model was built. Moreover, the neural network's weights are modified more frequently as the number of epochs rises, resulting in unwanted curve shifts from being underfitting to optimal curve, and finally overfitting the model. As stated in [16], Emotional expressions have been shown to differ significantly by gender in a small proportion, with women expressing feelings more openly, particularly when they are joyful, and internalizing negative emotions like grief or sadness. The results obtained from validation also show that gender has a slight influence over emotion capturing as visualized in Figure 6, which could be due to better feature extraction by the models built by CNN and VGG – 16, in which they primarily excel.

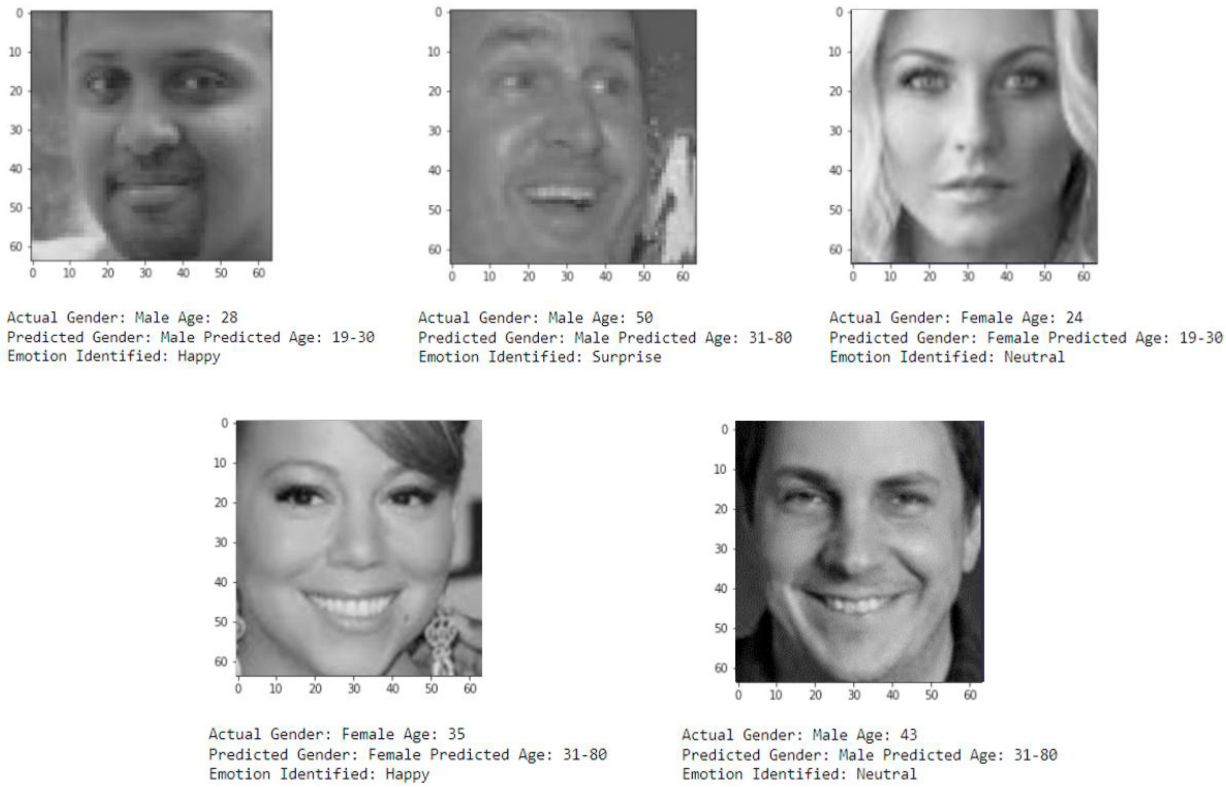

Fig. 6. Complete Facial Analysis of a few samples selected from the UTKFace dataset

6. Conclusion

Determining the age and gender of people we meet regularly plays a significant function in our social lives as well. Many intelligent devices, such as visual surveillance, clinical diagnosis, and product marketing, now require persons to be assessed using their facial photos. This research successfully implemented facial emotion recognition along with gender and age estimation, further improved it, and achieved excellent accuracy for VGG – 16 model. Comparing with the existing predictive systems, an overall best result has been shown by the SVM model with an F1 score of 0.83 for age detection and 0.46 for facial emotion recognition, and with less computation involved the current VGG model achieved an accuracy of 95.31% in the validation phase. To the best of our knowledge and considering the resources gathered and studied, in this paper, we have detailed the disparity in the usage of traditional machine learning and deep learning algorithms, along with the influence shown by gender on facial emotion.

The current work can still be extended and improved, i.e., anyone who would like to collaborate can extend this work of research by implementing several different pre-trained models such as AlexNet, ResNet - 50, GoogleNet, etc., to understand the effects of gender and age on determining the emotional expression of a human face on a higher scale and deploy them on to a specific platform to for developing a real-time use case.

References

[1] X. Zhao, S. Zhang and B. Lei, “Facial expression recognition based on local binary patterns and local fisher discriminant analysis,” WSEAS Transactions on Signal Processing, vol. 8, no. 1. pp.21-31, 2012

[2] M. Pourebadi and M. Pourebadi, “MLP neural network-based approach for facial expression analysis,” 2016.

[3] Bradski G., “The OpenCV Library,” Dr. Dobb’s Journal of Software Tools, 2010.

[4] Boureau, Y-Lan & Ponce, J & Lecun, Yann, “A Theoretical Analysis of Feature Pooling in Visual Recognition,” ICML 2010 - Proceedings, 27th International Conference on Machine Learning, 111-118, 2010.

[5] X. Yang et al., "Deep Label Distribution Learning for Apparent Age Estimation," 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), 2015, pp. 344-350, doi: 10.1109/ICCVW.2015.53.

[6] S. Escalera et al., "ChaLearn Looking at People 2015: Apparent Age and Cultural Event Recognition Datasets and Results," 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), 2015, pp. 243-251, doi: 10.1109/ICCVW.2015.40.

[7] M. Xiaoxi, L. Weisi, H. Dongyan, D. Minghui and H. Li, "Facial emotion recognition," 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP), 2017, pp. 77-81, doi: 10.1109/SIPROCESS.2017.8124509.

[8] D.-Y. Huang, Z. Zhang, and S. S. Ge, Speaker state classification based on fusion of asymmetric simple partial least squares (SIMPLS) and support vector machines, Special Issue of Computer and Speech Language on “Broadening the View on Speaker Analysis” 28(2), pp. 392-491, 2014

[9] S. T. Chavali, C. T. Kandavalli, S. T M and S. R, "Grammar Detection for Sentiment Analysis through Improved Viterbi Algorithm," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2022, pp. 1-6, doi: 10.1109/ACCAI53970.2022.9752551.

[10] S. S. Teja Gontumukkala, Y. S. Varun Godavarthi, B. R. Ravi Teja Gonugunta, R. Subramani and K. Murali, "Analysis of Image Classification using SVM," 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021, pp. 01-06, doi: 10.1109/ICCCNT51525.2021.9579803.

[11] D. Biswas, S. Jayan, S. Nadipalli, S. B and S. R, "Attendance Tracking with Face Recognition Through Hidden Markov Models," 2022 International Conference on Electronics and Renewable Systems (ICEARS), 2022, pp. 1640-1645, doi: 10.1109/ICEARS53579.2022.9751888.

[12] M. Khan, S. Chakraborty, R. Astya and S. Khepra, "Face Detection and Recognition Using OpenCV," 2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), 2019, pp. 116-119, doi: 10.1109/ICCCIS48478.2019.8974493.

[13] K. Goyal, K. Agarwal and R. Kumar, "Face detection and tracking: Using OpenCV," 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), 2017, pp. 474-478, doi: 10.1109/ICECA.2017.8203730.

[14] L. He and J. Pan, "Face Recognition Using Nonlinear Regression," 2011 Sixth International Conference on Image and Graphics, 2011, pp. 548-552, doi: 10.1109/ICIG.2011.182.

[15] A. Virigineni, M. Tanuj, A. Mani and S. R, "Stock Forecasting using HMM and SVR," 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), 2022, pp. 1-7, doi: 10.1109/IC3IOT53935.2022.9767969.

[16] Chaplin TM. Gender and Emotion Expression: A Developmental Contextual Perspective. Emot Rev. 2015 Jan;7(1):14-21. doi: 10.1177/1754073914544408. PMID: 26089983; PMCID: PMC4469291.

[17] K. Duvvuri, H. Kanisettypalli and S. Jayan, "Detection of Brain Tumor Using CNN and CNN-SVM," 2022 3rd International Conference for Emerging Technology (INCET), 2022, pp. 1-7, doi: 10.1109/INCET54531.2022.9824725.

[18] A. H. N. S. Chandana Sarvani, B. Sai Bharath, P. R. Vijaya Bharathi Reddy and S. R, "AI-Driven Medical Imaging Analysis for COVID-19 Detection," 2022 International Conference on Electronics and Renewable Systems (ICEARS), 2022, pp. 1799-1804, doi: 10.1109/ICEARS53579.2022.9751986.

[19] T. Kaur and T. K. Gandhi, "Automated Brain Image Classification Based on VGG-16 and Transfer Learning," 2019 International Conference on Information Technology (ICIT), 2019, pp. 94-98, doi: 10.1109/ICIT48102.2019.00023.

[20] L. T. Thao and N. H. Quang, "Automatic skin lesion analysis towards melanoma detection," 2017 21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES), 2017, pp. 106-111, doi: 10.1109/IESYS.2017.8233570.