Facial Emotions and Behavior Monitoring System using DNN

Bindu Bhargava Reddy Chintam[1], Maram Venkata Naga SaiTeja[2], M Sumanth[3], Lokireddy Sai Siddhardha Reddy[4], Prof. Raghavendra Reddy[5]

[1,2,3,4]School of Computer Science, REVA University, Bangalore – 560064
[5]Assistant Professor, Dept. of Computer Science, REVA University, Bangalore – 560064
Author Mail Id: bhargavchintam@gmail.com
Author Mobile No.: +91 9849936686

ABSTRACT

In this paper, Deep-Neural-Networks (DNN) and Machine Learning Algorithms are implemented to identify social distance, face masks, drowsiness detection, age-gender detection, and emotion detection. While dealing with social distancing initially we need to detect humans which are done by using COCO (Common Objects in Context) datasets and later on polygon-shaped ROI (Rectangular-region of Interest) is warped with a rectangle which helps to find the distance from each centroid(person). Similarly, we predict the face mask, age-gender, emotion, and drowsiness altogether using frontal-face detection and eye-detection via haarcascade dataset loaded into CNN for training and testing the models on color mapped images. The proposed model uses various machine learning techniques consists of classifiers like Linear discriminant Analysis (LDA), Independent Component Analysis (ICA), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). The accuracy of proposed system depends on timing (i.e., 88.2%, 89.7%, 95.1% and 98.3% in 0~0.2s, 0.2~0.6s, 0.6~1s, >1s time windows respectively). The accuracy even depends upon the distance away from the camera (i.e., 60.4%, 73.9%, 89.3%, 95.2%, and 62.2% in >15, 15~10, 10~6, 6~0.5, <0.5 meters respectively). The resultant average accuracy of all the models is about 96.3% which is capable to predict various tasks as said above. This complete model is made accessible to users via a standalone software/Desktop GUI. The proposed approach is promising for performing all the tasks and activities more accurately and efficiently.

KEYWORDS

Convolutional Neural Networks, K-Nearest Neighbor, Support Vector Machine, Linear Discriminant Analysis, Independent Component Analysis, Principal Component Analysis, Logistic Regression, Hidden Markov Model.

INTRODUCTION

As per the current situation, the COVID-19 is still prevailing all over the world. As per the WHO report, 136 million people got affected till march 2021 and out of which 2.9 million people died. Many people all around the globe are still facing challenges. As COVID-19 is a mutated virus, vaccines are not up to the mark as expected. The only solution is maintaining social distancing and face masks but it cannot be monitored continuously. This becomes a real challenge to monitor people in public areas like parks, shops, malls, and transport. In Similar terms, drowsiness of driver while driving causes lots of accidents in and around the world. As per the FARS (Fatality Analysis Reporting System) survey, 63% of accidents were due to drowning while driving for the past 2 years, out of which 12% involved collision with another motor vehicle and the remaining are crash scenarios. Face and Body are major parts of the human body to identify a person and their attributes as well. Emotion is a psychological state connected with the nervous system with perceptions, feelings, behavioral responses, and a degree of satisfaction or annoyance. Emotion detection has great latent in video surveillance, monitoring infants, disabled and elderly people. Human facial expressions are usual and straight means to interact emotions and intentions, especially in non-verbal communications. To solve these issues, we do need a system that helps to prevent losses and gain features from resources.

To develop such a system, we need to implement machine learning algorithms and Neural Networks. With the development of image processing techniques, human face recognition techniques have been applied. For emotion detection, we need to classify between multiple classes, we need to implement ML Techniques like LDA, KNN, and Naïve Bayes. Some of the gender features like hairs, beard/mustaches, lips, and skin. As there are 2 genders to identify as they belong to the same class. Hence, we can use binary classification methods likewise ICA, PCA, SVM, and Logistic Regression. For Age

prediction, we use multi-label classification and along with multi-class classification techniques like PCA, Random forests, and Naïve Bayes. Age ranges are fixed as 0-6, 7-12, 12-18, 18-25, 25-32, 32-42, 42-50, 60+ where each category loaded with model images of both male and female. For drowsiness detection, we consider eyes that are completely open by excluding all closed and partially closed eyes. Therefore, we found it deals with a single class (i.e., eyes opened completely) as we implement binary classification techniques likely LDA, SVM, and KNN. Finally, we process them using an R-CNN (Region-based CNN) image processing for object detection implemented as Standalone Software (GUI Based Application). The major contributions towards the project application, where we focused on accurate results and high-scale feature extraction from the images are taken into considerations. CNNs are trained with the latest dataset and own datasets with much-needed test cases. High scale image processing algorithms like R-CNN and YOLO were implemented and needful techniques like SVM, LDA, PCA, ICA, and logistic regression approaches are used. We even design a frame that can grasp multiple faces in a single shot.

The paper is organized as follows: Section II guidelines the related works in the area of analysis for detecting activities. The overall view of the proposed method is explained in Section III. Implementation details are described in Section IV followed by the conclusion and future works in Section V.

LITERATURE SURVEY

Because of deep learning, a lot of progress has been made in machine vision. As a result, stronger and more reliable algorithms have emerged. In the field, certain algorithms were used to solve a wide variety of complicated computer vision problems. This requires a more accurate assessment of the face's main points.

Christopher Streiffer et. al. [1] proposed that DarNet, a multi-modular info accretion and study framework intended to recognize and classify unfocussed driving conduct. Driver distraction has been identified as a significant contributor to motor vehicle crashes and injuries. DarNet is a convolutional and sporadic neural system that can deconstruct a driving picture. To find out how to identify up to 6 groups of driving activities with increased accuracy using IMU sensor data.

S. Zhang et. al. [2] proposed a multi-layer neural network model with the FERN-2013 dataset loaded and significantly used to capture the face from an image and video track using face frontal structure. Generating results even for background bright conditions. Where they used 2 groups of face mask activities are loaded into the system (i.e., with and without mask) and increased the accuracy by implementing logistic regression and SVM.

S.Alizadeh et. al. [3] proposed the architecture of Hidden Markov models (HMMs) for classifying expressions from video and images as well. He classified emotions into 6 categories and loaded CNN with 20,000 images of each category to grab back good accuracy in the model.

S. Yang et. al. [4] proposed that trained for body detection and performance results are reported by using the learned model. They implemented using a-DNN model along with the SVM technique. This helped to figure the person even in the good crowd as well using a wide-angle lens camera with 170 degrees.

B Subarna et. al. [5] proposed that predict human emotions (Frames by Frames) using deep Convolution Neural Network (CNN). They categorized emotions into 7 types and developed a dataset as well. They developed CNN to grasp emotions from multiple persons in a single frame.

Ref No.	Approach	Advantages	Limitations	Accuracy
[1] Christopher Streiffer et. al.	DarNet – Multi-modular info accumulator CNN.	Implementation of Intermitted CNN.	Fails in night mode, direct focused light, and partially closed eyes.	~93.2%
[2] S. Zhang et. al.	Uses Open-CV and SVM techniques.	Able to generate results for bright backgrounds.	Fails while wearing sunglasses and lighting conditions.	~89.2%
[3] S.Alizadeh et. al.	Implemented 3D-CNN with HMM.	Able to detect emotions with images and video.	Fails to work under multiple faces in the background.	~91.5%
[4] S. Yang et. al.	Trained body detection using R-CNN.	Able to figure out humans in a good crowd.	Fails in bad lighting conditions.	~95.2%
[5] B Subarna et. al.	Using Convolution Neural Network (CNN) and LDA.	Able to work in multiple face backgrounds.	Fails to bring accuracy in the prediction of emotions.	~85.3%

As there are few limitations from existing papers in summary says that prediction at night mode is much difficult. Face identification while wearing sunglasses is too unconditional for prediction. Multiple face prediction is complex in identifying in a single frame. Few models even failed to perform predictions with good accuracies. Limitations towards the datasets even prevent getting good results quickly and precisely.

METHODOLOGY

Proposed System:

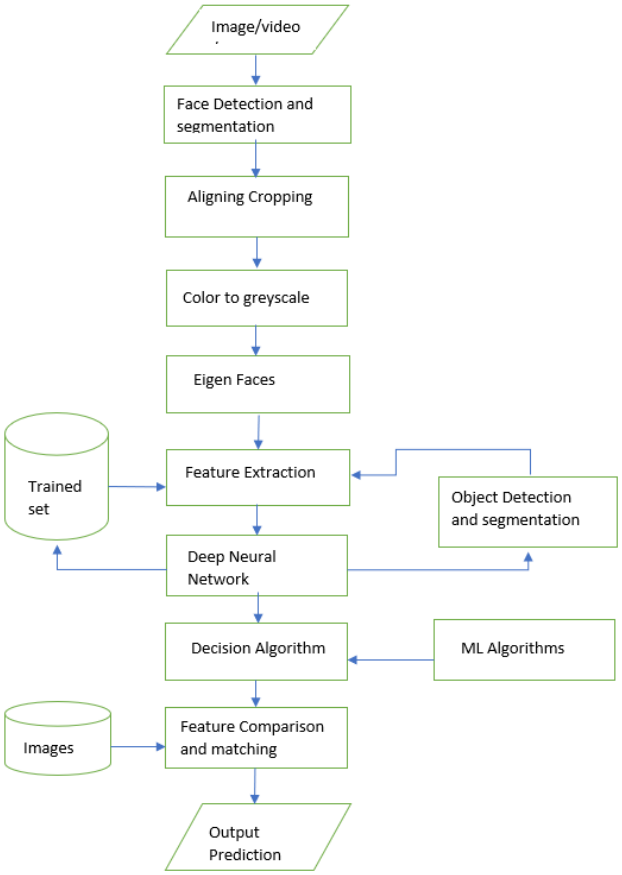
Face and Body are a major part of the human body to identify a person and their attributes as well. Emotion is a psychological state connected with the nervous system with perceptions, feelings, behavioral responses, and a degree of satisfaction or annoyance. Emotion detection has great latent in video surveillance, monitoring infants, disabled and elderly people. Human facial expressions are usual and straight means to interact emotions and intentions, especially in non-verbal communications. With better and progressive image detection and processing technologies, precise emotion detection seems equally attainable. The performance of various facial emotion recognition techniques is related based on the number of expressions recognized and the complication of the algorithms. Six basic emotions are universally experienced in all human cultures. A smile on the human face discloses happiness, and the lips and eyes show a curled shape. In sadness, the face expresses bagginess with rising tilted eyebrows and frowns. The anger on the human face is showcased with cuddled eyebrows, slim and strained eyelids. The disgust expressions are articulated with pull-down eyebrows and wrinkled nose while surprise is expressed with eye-widening and mouth huge. With the development of image processing techniques, human face recognition techniques have been applied. Here, as we need to classify between multiple classes, we need the implementation of LDA, KNN, and Naïve Bayes.

Meanwhile, the gender identification technique is based on the gender dataset from Yolo and COCO. Gender identification of human face images means that computers process human face images extracts the features of images, then identifies the gender by using classification. Some of the gender features like hairs, beard/mustaches, lips, and skin. As there are 2 genders to identify as they belong to the same class. Hence, we can use binary classification methods likewise ICA, PCA, SVM, and Logistic Regression.

Age prediction deals with various features like skin wrinkles, eyes, bread, nose shape, cheeks, Hairstyle, and color. These features are extracted from images and processed to predict the age from the age range table. As we need to predict from multiple labels here, we use multi-label classification and along with multi-class classification techniques like PCA, Random forests, and Naïve Bayes. Age ranges are fixed as 0-6, 7-12, 12-18, 18-25, 25-32, 32-42, 42-50, 60+ where each category loaded with model images of both male and female.

Driver Drowsiness is one of the top causes of motor vehicular accidents. For these motives, a risk alert system for drivers using a detector that can determine drowsiness is extremely suggested. The alert system can wake the drowsy driver or hand over the switch to the autonomous vehicle. Various techniques have been implemented to measure driver drowsiness. Here, we pick the driver who asleep for 3 seconds and alert him with a high-pitched alarming sound. Here, we capture the image of the driver. Later on, we catch up with the eyes of the driver by extracting features of the eyes. Here, we consider eyes that are completely open by excluding all closed and partially closed eyes. Therefore, we found it deals with a single class as we implement binary classification techniques likely LDA, SVM, and KNN.

The Architecture of Proposed System:



The features used in the proposed model need to be pre-processed and divide into various stages. Each stage needs input feature data. Each stage has multiple neurons trained with various conditions to be performed.

1. Preprocessing

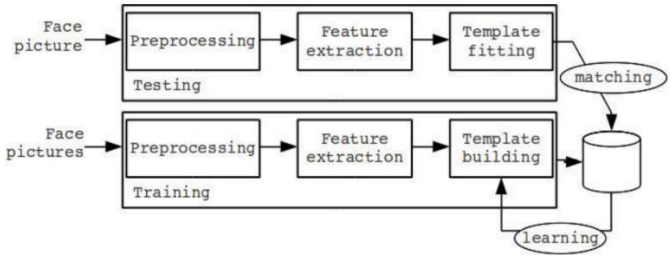
Lots of feature extraction to be done initially to capture various features from a bunch of images. Image max face area to be captured and crop the face using the harscascade frontal face and eye detection model dataset. The frame size should be set based on the model dataset. Background blurring, brightness control, aperture, and sharpness to be monitored based on image collection in real-time.

2. Training Data

Before training, preprocessing using database images on the dataset and validation sample images for training. Few image samples fail in the detection task because of the model dataset. So, fail case images to be processed using excessive-performance hardware like GPU to prevent forbidden detections.

Large quantities of training data are required for neural networks, especially deep neural networks. Furthermore, the photographs used for the training model for a significant portion of the final model's output. It entails the need for high-quality, quantitative data collection. Several datasets, varying from a few hundred high-resolution pictures to tens of thousands of smaller photographs, are available for study into emotion recognition.

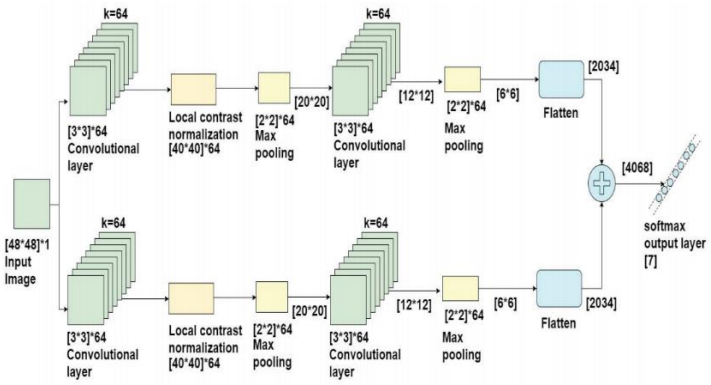
We train the network using GPU for 20 epochs to ensure that the precision converges to the optimum. The network will be trained on a larger set than the one previously described in an attempt to improve the model even more. Training will take place with 20,000 pictures from the dataset instead of a lesser no. of pictures.



3. Neural Networks

In recent years, neural networks have proved to be inspiring in terms of computer vision tasks and performance as well. Whereas, neural networks require high-performance GPU for rendering and computing. Here, in this project, we implemented a neural network with 9 main layers.

CNN Implementation Diagram:



Layer 0: (Input Layer) Input frame size [48*48] and with one color frequency is taken here. Preprocessed images are given as input to this layer.

Layer 1: (CNN Layer) Calculates the output of all neurons that are connected with the input layer, outputs the dot product of their weights.

Layer 2: (RELU Layer) In this layer, each element works on function *max(0,x)* zero. Here, we find batch normalization is performed.

Layer 3: (Max Pooling Layer) It performs a downsampling process along through the dimensions. Pooling is done on each layer specifically.

Layer 4: Convolution with 64 filters and restores original size.

Layer 5: Convolution with 128 filters and again restores the original size.

Layer 6: Convolution with 128 filters and size of output 64 and depths of 128 filters.

Layer 7: Fully connected with neurons. Weights are determined by backpropagation.

Layer 8: (Fully Connected Layer) calculates the class scores, resultant volume size.

Layer 9: (Max Layer) with neurons to predict the output.

Algorithms:

Algorithm of face detection from an image/video/stream.

Step 1: Input Face Images, Videos, stream
Step 2: Preprocessing the Face Images
- 2.1 Face Detection and Segmentation
- 2.2 Aligning & Cropping
- 2.3 Color to Grayscale
- 2.4 Eigen Faces
- 2.5 Capturing Features

Step 3: Feature Extraction from Images

3.1 Feature-based /Appearance-based approach
3.2 Object detection and template building (Testing)
3.3 Matching with trained dataset utilizing CPU (GPU, if available)
3.4 improving trained dataset (Training)
Step 4: Classification Methods
4.1 Binary Classification or Multi-label classification
4.2 Decision Algorithm
Step 5: Final Prediction with an output frame

Pseudo Code:

Pseudo code for detecting face from Image.

```
# Command line image inputs
imagePath = sys.argv[1]
cascPath = sys.argv[2]

# Create the haarcascade
faceCascade = cv2.CascadeClassifier(cascPath)

# Read the image
image = cv2.imread(imagePath)

# Coverting to Gray Scale Image
gray = cv2.cvtColor(image,
cv2.COLOR_BGR2GRAY)

# Detect faces in the image
Faces = faceCascade.detectMultiScale(gray,
            scaleFactor=1.1,
            minNeighbors=5,
            minSize = (30, 30),
            flags =
        cv2.cv.CV_HAAR_SCALE_IMAGE)
print "Found {0} faces!".format(len(faces))

# Draw a rectangle around the faces
for (x, y, w, h) in faces:
    cv2.rectangle(image, (x, y), (x+w, y+h), (0, 255,
    0), 2)

# Displaying the output image
cv2.imshow("Faces found", image)
cv2.waitKey(0)
```

Formulas/Functions:

The features also were converted to a convolution layer then passed via a deep network with 128 neurons, followed by a 50% slacker even before the output protective layer the model from overloading.

RELU Activation Function:

The Relu activation function was used in the hidden layers. R(z) represents the ReLu function's output, while z represents the weighted sum within each neuron's input variables. If the weight vector is less than zero, the method contains zero, and if it is greater or equal to zero, the method returns z.

$R(z) = \max(0, z)$

R(z) shows the output ReLu function.

Sigmoid Activation Function:

For the output units, a sigmoid activation mechanism was used. The sigmoid output system is described by (z), and the weighted average of input parameters is represented by z in this formula. Within the layer, plots of activation functions are seen.

$\sigma(z) = 1 / (1+ e^{-z})$

σ(z) shows the sigmoid output function.

Binary Cross-Entropy (Minimized):

The loss measure is binary cross-entropy, which is reduced during preparation. N denotes the total number of training instances, and y denotes the classified output (in binary classification, 0 and 1), while y denotes the output provided by the model throughout training.

$BCE = -1/N (E) y_i * \log (y_i^o) + (1 - y_i) * \log (1 - y_i^o)$

N shows the total number of training examples, and yi shows the labeled output.

RESULTS

The Requirement for this project:

Hardware requirements, we need two hardware devices such as Webcam/GoProCam and Laptop/Desktop. Software requirements we use python and its modules/packages such as OpenCV, Keras, Matplotlib, SciPy, TensorFlow, NumPy.

Images and Videos are used to send through the software to detect required results based on the need.

OpenCV is a cross-platform library that mainly focuses on image processing, video capture and used for face and object detection.

Keras is a Python library for developing and evaluating deep learning models.

Matplotlib is used to represent the data in the graphical model using the python programming language.

SciPy and **NumPy** are used for scientific and mathematical problems.

TensorFlow is used to train the model with a different test case to increase the accuracy of the model.

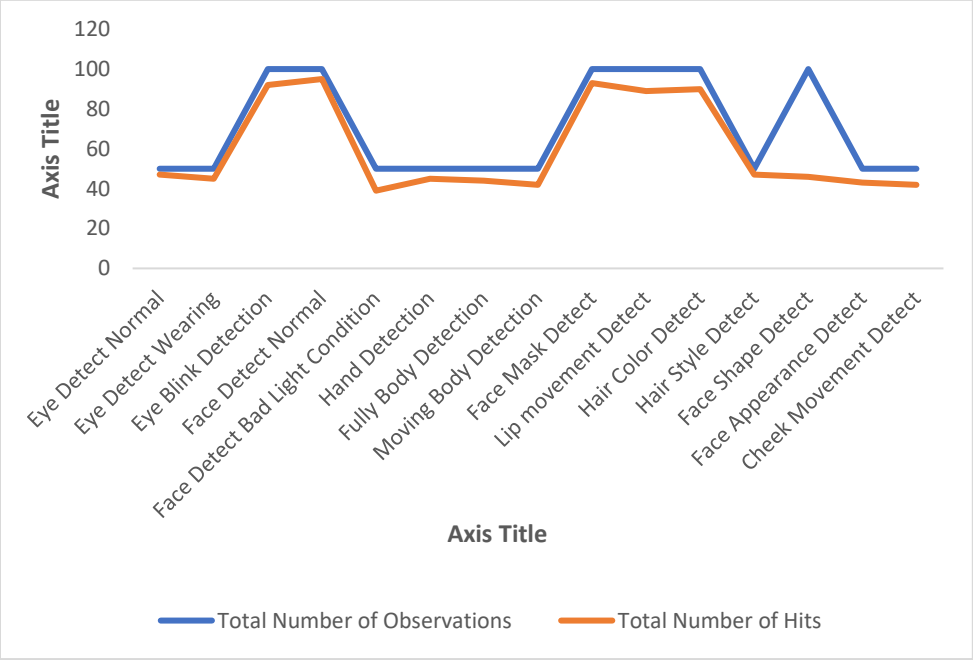
Detecting Tasks and Accuracies:

Test cases

These test cases are performed based on the observations for 50 to 100 cases. Based on these test cases we usually identify the hits or correct results every time. Performance and prediction even depend upon the GPU performance and CPU speed as well.

The below results from the table shows the hits of the parameters considered.

Test Parameters	Total Number of Observations	Total Number of Hits	Percentage of Hits
Eye Detect Normal	50	47	94.0%
Eye Detect Wearing	50	45	90.0%
Eye Blink Detection	100	92	92.0%
Face Detect Normal	100	95	95.0%
Face Detect Bad Light Condition	50	39	78.0%
Hand Detection	50	45	90.0%
Fully Body Detection	50	44	88.0%
Moving Body Detection	50	42	84.0%
Face Mask Detect	100	93	93.0%
Lip movement Detect	100	89	89.0%
Hair Color Detect	100	90	90.0%
Hair Style Detect	50	47	94.0%
Face Shape Detect	100	46	92.0%
Face Appearance Detect	50	43	86.0%
Cheek Movement Detect	50	42	84.0%



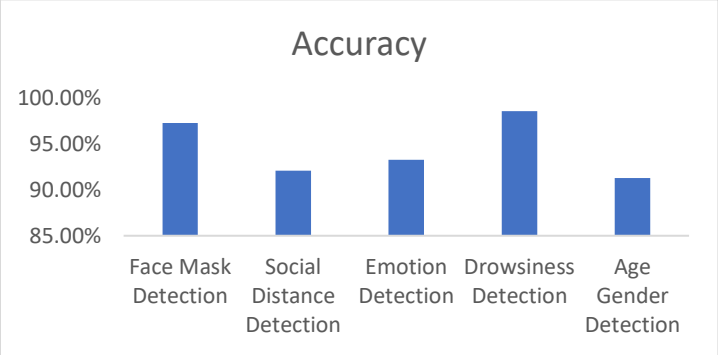
Overall Accuracy: 96.3%

Module Accuracies:

Model accuracies are calculated based on the probability of correctness and perfectness in the prediction of tasks assigned.

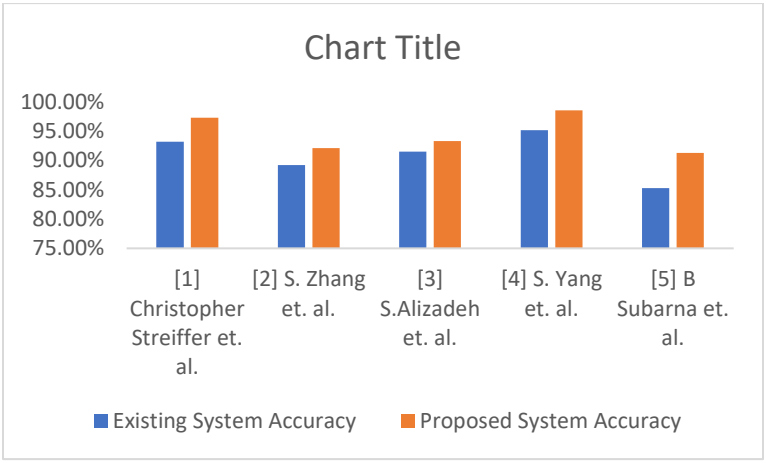
As there are various modules with various approaches and techniques, hence accuracies vary with each other. Here, we measure accuracy based on the probability ratio of real-time prediction vs actual prediction. Therefore, the below table shows the max threshold of accuracies of various modules that presents the project.

Model Name	Accuracy
Face Mask Detection	97.3%
Social Distance Detection	92.1%
Emotion Detection	93.3%
Drowsiness Detection	98.6%
Age Gender Detection	91.3%



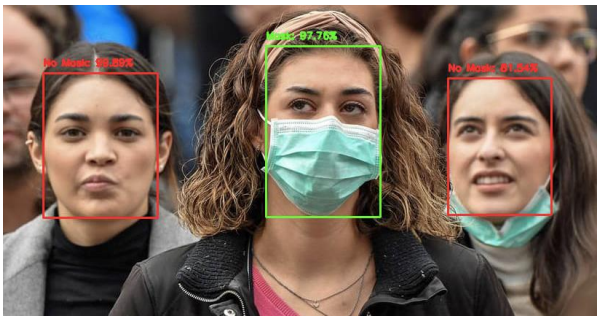
Comparison proposed system with existing system:

Authors	Existing System Accuracy	Proposed System Accuracy
[1] Christopher Streiffer et. al.	~93.2%	97.3%
[2] S. Zhang et. al.	~89.2%	92.1%
[3] S.Alizadeh et. al.	~91.5%	93.3%
[4] S. Yang et. al.	~95.2%	98.6%
[5] B Subarna et. al.	-85.3%	91.3%



Resultant Images:

Face Mask Detection O/P:



Social Distancing Detection O/P:

Emotion Detection O/P:

Drowsiness Detection O/P:

CONCLUSION & FUTURE ENHANCEMENT

Some of the major applications of the models, Emotion detection has great potential in video surveillance, monitoring infants, disabled and elderly people. Used to retain the facial feedback based on their expressions after and before visiting a Movie, Market, Park, Complex, and Malls, etc. Age and gender recognition play a major role in the Police investigation and Intelligence department as it helps find the actual suspect based on his age and gender. They could get a filtered-out result of that person who has performed a criminal act or any other activity. The alert system can awaken the drowsy driver or hand over the control to an

autonomous vehicle. Suspicious activity detection identified using video surveillance to transform into a too bad legitimate verification to recognize offenders after the occasion of bad behavior. Social Distancing and Face Mask Monitoring is only the solution to prevent this pandemic situation in and around you and wide-spreading.

Finally, we conclude that our model is more effective at multitasking and has a higher rate of accuracy. This model will assist 70% of the current industry in performing tasks such as reporting and guiding operations. As a result, this will boost and grep the current age of technology to find consistency and take necessary action. We are confident that our model can provide information on Social Distancing, Emotion Recognition, Gender/Age Classification, Driver Drowsiness, and Illegal Behavior with a 96.3 percent accuracy. This model has the potential to inspire large-scale enterprises to incorporate and operate them in a way that maintains stability and robustness.

Validation accuracy, computational complexity, detection rate, learning rate, validation failure, and computational time per stage are all used to assess the proposed facial emotion detection model's efficiency. We compared the performance of our proposed model to that of an original version using qualified and test sample images. The results of the experiment show that the proposed model outperforms previous models published in the literature in terms of detection results. On both datasets, the suggested framework produces state-of-the-art results, according to the observations.

We would to like improve and sort our algorithms to make them more accurate in the upcoming days. We would like to add few addons that could deliberately improve our model on both sides, Trust & Security, improve hardware config's: GPU Enhancement, Algorithm Tuning: Improving Stability and Feature Engineering: More Refined and Accurate.

REFERENCES

1. Christopher Streiffer DarNet: A Deep Learning Solution for Distracted Driving Detection, 2019 ACM.
2. S. Zhang, R. Zhu, X. Wang, H. Shi, T. Fu, S. Wang, T. Mei, and S. Z. Li, "Improved selective refinement network for face detection", 2020.
3. S. Alizadeh, A. Fazel, "Convolutional Neural Networks for Facial Expression Recognition", CoRR, abs/1704.06756,2019.
4. S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark," in IEEE CVPR.
5. "Real-Time Facial Expression Recognition Based on Deep Convolutional Spatial Neural Networks", B Subarna; Daleesha M Viswanathan, " 2019 International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR) "
6. M. ASJID TANVEER1, M. JAWAD KHAN1, M. JAHANGIR QURESHI2 "Enhanced Drowsiness Detection Using Deep Learning: An NIRS Study", 2019
7. Amrutha C.V, C. Jyotsna, Amudha J." Deep Learning Approach for Suspicious Activity Detection from Surveillance Video" Proceedings of the Second International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2020) IEEE Xplore.
8. Alavudeen Basha A, Parthasarathy P" Detection of Suspicious Human Activity based on CNN-DBNN Algorithm for Video Surveillance Applications" 2019 Innovations in Power and Advanced Computing Technology (i-PACT).
9. Akriti Jaiswal, A. Krishnama Raju, Suman Deb" Facial Emotion Detection Using Deep Learning", 2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020.
10. Insha Rafique, Awais Hamid, Sheraz Naseer" Age and Gender Prediction using Deep Convolutional Neural Networks", 2019 International Conference on Innovative Computing (ICIC).
11. R. Sharma, T. S. Ashwin, and R. M. R. Guddeti, "A Novel Real-Time Face Detection System Using Modified Affine Transformation and Haar Cascades," in Recent Findings in Intelligent Computing Techniques, Springer, 2019, pp. 193–204.
12. G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2019.
13. G. Ozbulak, Y. Aytar, and H. K. Ekenel, "How transferable are CNN-based features for age and gender classification" in 2016 International Conference of the Biometrics Special Interest Group (BIOSIG), 2018.
14. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in IEEE CVPR, pp. 770–778, 2019.