_____

Facial Emotions and Behaviour Monitoring System using DNN

Raghavendra Reddy[1], Bindu Bhargava Reddy Chintam[2], Maram Venkata Nagasai Teja[3], M Sumanth[4], Lokireddy Sai Siddhardha Reddy[5]

[1,2,3,4,5,] School of Computer Science and Engineering, REVA University.
Email: raghavendra.reddy@reva.edu.in[1], bhargavchintam@gmail.com[2], maramsaiteja22@gmail.com[3], sumanthkumar6677@gmail.com[4], siddhardh2605@gmail.com[5]

**Corresponding author's E-mail:* raghavendra.reddy@reva.edu.in*

Article History	Abstract
                           	In this paper, Machine Learning Algorithms are used to implement the proposed approach 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' faces, which are available 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 facemask, age-gender, emotion, and drowsiness altogether using frontal-face detection and eye-detection via haarcascade dataset loaded into Convolutional Neural Network (CNN) to train and test the models on colour mapped images. In the proposed model, we are using machine-learning techniques such as linear discriminant Analysis (LDA), Independent Component Analysis (ICA), Principal Component Analysis (PCA) for age-gender detection and emotion detection, K-Nearest Neighbours (KNN) for Social Distancing, and Support Vector Machine (SVM) for facemask detection and drowsiness detection. The accuracy of proposed system depends on frame (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 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.the systemic health of the patient and avoiding possible drug interactions. **Keywords:** *Social Distance, Facemasks, Drowsiness Detection, Age-Gender Detection, And Emotion Detection*

1. 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 until 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 [15,16]. 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 many 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 [6].

Face and Body are major parts of the human body to identify a person and their attributes as well. Perceptions, emotions, behavioural reactions, and a degree of pleasure or frustration are all part of the emotional state, which is linked to the nervous system [10, 12, and 13]. In video surveillance, tracking children, injured, and aged persons, emotion recognition has a lot of potential [7, 8]. Human facial expressions are a common and straightforward way to communicate feelings and thoughts,

particularly in nonverbal communication. To solve these issues, we do need a system that helps to prevent losses and gain features from resources [9,11,14]

In the proposed work, we are performing various tasks Face Mask Detection, Social Distance Detection, Emotion Detection, Drowsiness Detection, and Age-Gender Detection. Face Mask Detection is done by using Region-based Convolutional Neural Networks (R-CNN) image processing algorithm. Social Distance Detection is implemented by using KNN. In Gender Detection, as there are 2 genders to identify as they belong to different class. Hence, we use SVM. For Age prediction, we use multi-label classification and multi-class classification techniques SVM. For drowsiness detection, we consider eyes that are completely open by excluding all closed and partially closed eyes. It is implemented by using logistic regression. This paper is structured in the following way: Section II outlines the associated studies in the field of activity detection research. Section III explains the new method's ultimate perspective. Section IV describes the implementation specifics, followed by Section V's conclusion and future enhancement work.

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. Some of the papers discussed here will help us to identify the research gaps. Christopher Streiffer et. al. [1] proposed that Dark Net is a multi-layered feature accretion & study framework planned to recognize and classify unfocussed driving conduct. Driver distraction has been identified as a significant contributor to motor vehicle crashes and injuries. A sporadic convolutional neural network that can deconstruct a driving picture is known as Dark Net. Learn how to correctly identify up to six types of driving events 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. Author 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 a Cu-DNN model for body detection and accuracy results are stated by using the learned model. They implemented using the DNN model along with the SVM technique. This helped to identify 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 CNN. They categorized emotions into 7 types and developed a dataset as well. They used CNN to grasp emotions from multiple persons in a single frame.

Table 1: shows the summary of various existing work approaches, advantages, limitations, and accuracy

Approach	Advantages	Limitations	Accuracy
CNN	Intermitted CNN.	Fails capture in low light conditions.	93.2%
Open-CV and SVM	Works in bright backgrounds.	Fails while wearing glasses.	89.2
3D-CNN with HMM	Detects both images and videos.	Fails to detect multiple faces.	91.5%
R-CNN.	Works even in-crowd.	Fails in dark background.	95.2%
CNN and LDA.	Detects multiple faces.	Scope to improve Accuracy.	85.3%

2. Materials And Methods

Human facial recognition methods were used to develop image-processing tasks. The number of emotions identified and the complexity of the algorithms are linked to the success of different facial emotion recognition techniques. In all human societies, there are six fundamental feelings that are uniformly felt. A smile on a person's face denotes joy, and the mouth and eyes have a curled appearance. With rising tipped eyebrows and frowns, the face expresses sorrow with bagginess. Cuddled lashes, lean and strained eyelids highlight the human face's indignation. Drooping ears and a wrinkled nose convey disgust, while surprise is expressed by widening eyes and a broad jaw. Human

facial recognition methods have been introduced as image-processing techniques have progressed. We need to incorporate LDA, KNN, and Naive Bayes so we need to distinguish between different groups.

Meanwhile, the Yolo and COCO identity datasets are used to identify gender. Computers process human face photographs to remove characteristics, which are then classified to determine gender. Some of the gender features like hairs, beard/moustaches, lips, and skin. As there are two 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 of 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. One of the leading causes of motor vehicle collisions is driver drowsiness. For these reasons, a danger warning system for drivers that uses a drowsiness monitor is strongly recommended. The warning system will either wake up the drowsy driver or transfer control to the self-driving car. The drowsiness of drivers has been measured using a variety of methods. 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 of LDA, SVM, and KNN [17-22].

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. The complete flow of the system is given in figure 1.

A. 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 haas-cascade 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.

B. 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. Deep neural networks, in particular, need a huge volume of training data. Furthermore, the images used in the training model make up a significant portion of the final model's output. It necessitates the collection of objective data of good quality. There are many datasets available for emotion recognition studies, ranging from a few thousand high-resolution photographs to hundreds of thousands of smaller images. We train the network with GPU for 20 epochs to ensure that the precision converges to the optimal level. The network will be trained on a wider variety of topics than the one described previously to improve the model even further. Training would be performed with 20,000 photos from the dataset, rather than a smaller amount, as seen in fig 2.
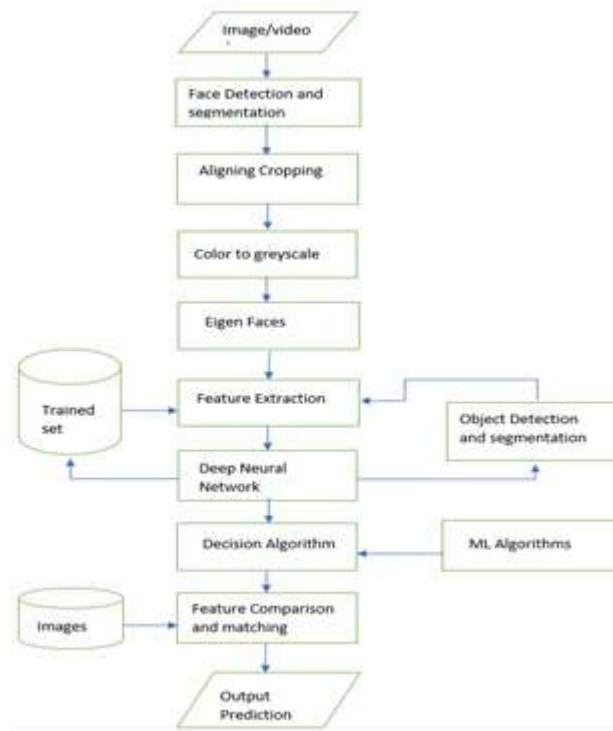
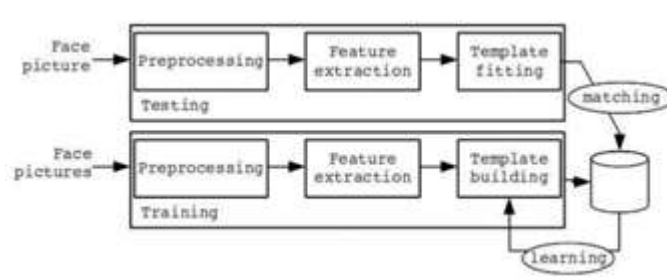**Figure 1:** Simplified flow-chart for the entire system

Figure 2: Training and testing process for the proposed system

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, as shown in fig 3.
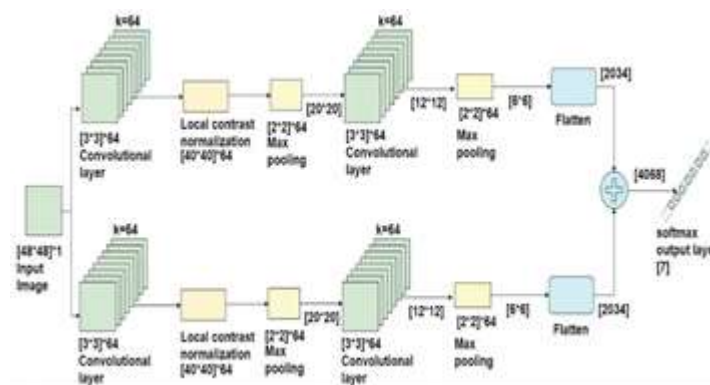

Figure 3: CNN Architecture used in the proposed system

CNN has various layers, Initially Input Layer - Input frame size is about [48*48] and with one color frequency is taken here. Preprocessed images are given as input to this layer. CNN Layer - Computes the productivity of all neurons which are connected with the input layer, outputs the dot product of their weights. RELU Layer - where each element works on function *max (0, x)* zero and where batch normalization is executed. Max Pooling Layer - It down samples the data as it moves through the dimensions and for each layer pooling is performed separately. Next layer Convoluted with 64 modifiers that recover the scale to its original state. Firstly, Convolution with 128 filters and again restores the original size. Secondly, Convoluted with 128 layers, 64 output sizes, and 128-layer levels. Later Entirely linked with neurons. Weights are determined by backpropagation. Fully Connected

Layer - computes the class nicks, resulting in capacity size. Max Layer - through neurons to expect the yield.

Algorithms:

Algorithm for 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

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 as shown in equation 1. 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.

$$R(z) = \max(0, z) \text{ --------------------} \quad (1)$$

R(z) shows the output RELU function.

Sigmoid Activation Function:

For the output units, a sigmoid activation mechanism was used as shown in equation 2. 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.

$$\sigma(z) = 1 / (1 + e^{-z}) \text{ ----------------------} \quad (2)$$

$\sigma(z)$ shows the sigmoid output function.

3. Results and Discussion

The proposed work is implemented by using python and its modules such as OpenCV, Keras, Matplotlib, SciPy, TensorFlow, NumPy. Webcam/GoProCam has been used as a part of the hardware component. Table 2 shows the observations from the proposed system for different 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.

Table 2: Observations From The Proposed System For Different Cases

Test Parameters	No. of Tests	No. of Successes	Percentage of Success
Face Detect Bad Light Condition	50	39	78.0%
Moving Body Detection	50	42	84.0%
Face Mask Detect	100	93	93.0%
Eye Blink Detection	100	92	92.0%
Hand Detection	50	45	90.0%
Face Detect Normal	100	95	95.0%

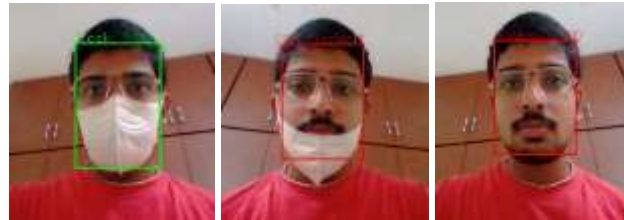|			
Eye Detect Normal	50	47	94.0%
Fully Body Detection	50	44	88.0%
Eye Detect Wearing	50	45	90.0%
Face Shape Detect	100	46	92.0%
Lip movement Detect	100	89	89.0%
Hair Color Detect	100	90	90.0%
Hair Style Detect	50	47	94.0%
Face Appearance Detect	50	43	86.0%
Cheek Movement Detect	50	42	84.0%

Fig 4. Face Mask Detection

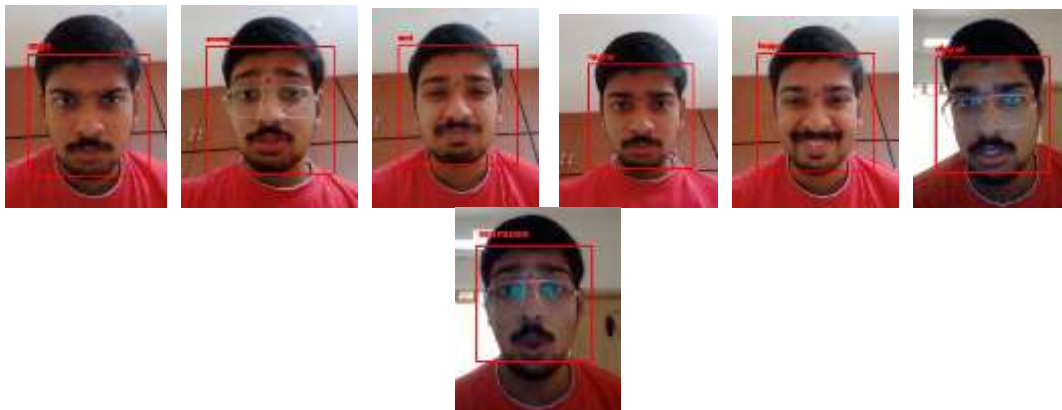

Fig 5. Social Distancing Detection

Fig 6. Emotion Detection

Fig 7. Drowsiness Detection

Fig 8. Age & Gender Detection

Figures 4&8: shows the snapshot of different operations performed by using the proposed system.

are taking images, video, or live capturing of the data (by using camera) as an input to our system for performing Face Mask Detection, Social Distancing Detection, Emotion Detection, and Drowsiness Detection. Model accuracies are calculated based on the probability of correctness and perfectness in the prediction of tasks assigned as shown in table 3. 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. Table 4 and fig 9. show the Accuracy comparison of the proposed system with the existing system.

Table 3: Module Accuracy Of The Proposed System For Different Cases

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%

Table 4: Accuracy Comparison Between The Proposed System And The Existing System

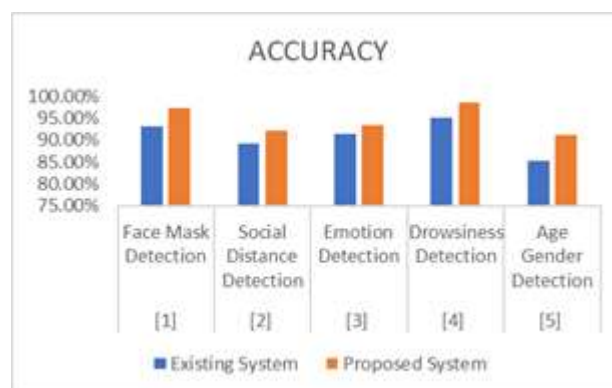| Different task	Existing System	Proposed System
Face Mask Detection	93.2%	97.3%
Social Distance Detection	89.2%	92.1%
Emotion Detection	91.5%	93.3%
Drowsiness Detection	95.2%	98.6%
Age Gender Detection	85.3%	91.3%

Fig 9: Accuracy comparison of the proposed system with the existing system

4. Conclusion

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 Behaviour 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. 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, R. Raghavendra, Theophilus A. Benson, M. Srivatsa, "DarNet: A Deep Learning Solution for Distracted Driving Detection", 18th ACM/IFIP/USENIX Middleware Conference, pp 22–28, ACM, 2019.

[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", Association for the Advancement of Artificial Intelligence (AAAI),2020

[3] S. Alizadeh, A. Fazel, "Convolutional Neural Networks for Facial Expression Recognition", CoRR, 2019.

[4] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A face detection benchmark", Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, pp. 5525-5533, 2016.

[5] B Subarna, Daleesha M Viswanathan, "Real-Time Facial Expression Recognition Based on Deep Convolutional Spatial Neural Networks", International Conference on Emerging Trends and Innovations In Engineering And Technological Research (ICETIETR), IEEE, pp. 1-5, 2018.

[6] Tanveer, M., Khan, M., Qureshi, M.J., Naseer, N., & Hong, K., "Enhanced Drowsiness Detection Using Deep Learning: An fNIRS Study", IEEE Access, vol 7, pp. 137920-137929, 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) IEEE, pp. 335-339, 2020.

[8] Alavudeen Basha A, Parthasarathy P, "Detection of Suspicious Human Activity based on CNN-DBNN Algorithm for Video Surveillance Applications", Innovations in Power and Advanced Computing Technology (i-PACT), IEEE, pp. 1-7, 2019.

[9] Akriti Jaiswal, A. Krishnama Raju, Suman Deb, "Facial Emotion Detection Using Deep Learning", International Conference for Emerging Technology (INCET), IEEE, pp. 1-5, 2020.

[10] Insha Rafique, Awais Hamid, Sheraz Naseer, "Age and Gender Prediction using Deep Convolutional Neural Networks", International Conference on Innovative Computing (ICIC), IEEE, pp. 1-6,2019.

[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," Recent Findings in Intelligent Computing Techniques, Springer, pp. 193–204, 2019.

[12] G. Levi and T. Hassncer, "Age and gender classification using convolutional neural networks", Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, pp. 34–42, 2015.

[13] G. Ozbulak, Y. Aytar, and H. K. Ekenel, "How transferable are CNN-based features for age and gender classification", International Conference of the Biometrics Special Interest Group (BIOSIG), IEEE, pp. 1-6, 2018.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", IEEE CVPR, pp. 770–778, 2019.

[15] S. Singh, P. Raj, R. Kumar and R. Chaujar, "Prediction and forecast for COVID-19 Outbreak in India based on Enhanced Epidemiological Models", Second International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 93-97, 2020.

[16] World Health Organisation Novel CoronaVirus 2019 (COVID-19): https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[17] Sudhan Murugan Bhagavathi, Anitha Thavasimuthu, Aruna Murugesan, Charlyn Pushpa Latha George Rajendran, A Vijay, Raja Laxmi, Rajendran Thavasimuthu, Weather forecasting and prediction using hybrid C5.0 machine learning algorithm International Journal of Communication Systems, Vol. 34, Issue. 10, Pp. e4805, 2021.

[18] PM Surendra, S Manimurugan, A New Modified Recurrent Extreme Learning with PSO Machine Based on Feature Fusion with CNN Deep Features for Breast Cancer Detection, Journal of Computational Science and Intelligent Technologies, Vol. 1, Issue. 3, Pp. 15-21, 2020.

[19] PK Sadineni, Comparative Study on Query Processing and Indexing Techniques in Big Data, 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 933-939, 2020.

[20] AH Omar Baabood, Prajoona Valsalan, Tariq Ahmed Barham Baomar, IoT Based Health Monitoring System, Journal of Critical Reviews , Vol. 7, Issue. 4, pp. 739-743, 2020.

[21] Sajay KR, Suvanam Sasidhar Babu, Vijayalakshmi Yellepeddi, Enhancing The Security Of Cloud Data Using Hybrid Encryption Algorithm, Journal of Ambient Intelligence and Humanized Computing, 2019. https://doi.org/10.1007/s12652-019-01403-1

[22] Bindhia K Francis, Suvanam Sasidhar Babu, Predicting academic performance of students using a hybrid data mining approach, Journal of Medical Systems, 43:162, 2019. https://doi.org/10.1007/s10916-019-1295-4