IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, April 2022
Impact Factor: 6.252

Deep Neural Network for Complex Human Activity Recognition

M. Kusuma Sri and K. Sai Krishna
Assistant Professor
Anurag University, Hyderabad, TS, India
kusumasriece@cvsr.ac.in and saikrishnaece@cvsr.ac.in

Abstract: *The problem statement for our project is to recognize the automatic emotions from the analysis of body movement. It has tremendous potential to bring in some changes into domains like robotics, biometric identity recognition etc. A computer can identify human emotions from the facial expressions; this may bring about changes in the way we interact with the computer. One of the challenges is to identify emotion specific features from a vast number of descriptors of human body. In this project we used feature selection frame work to accurately recognize basic emotions namely happiness, sadness, anger, depression and neutral. This project consists of two layers, the first layer, a unique combination of Analysis of variance (ANOVA) and multivariate Analysis of variance (MANOVA), to eliminate irrelevant features. The proposed model can detect complex actions like sitting, walking and few others. The proposed model achieved recognition accuracy of 90.0% during walking, 96.0% during sitting, demonstrating high accuracy and robustness of the developed method.*

Keywords: Movement Recognition, CNN, SVM

I. INTRODUCTION

This project involves recognition of complex human activities using Convolutional Neural Networks algorithm. This project is divided into two parts, one is facial feature extraction and the other is pattern recognition (SVM). The problem statement of our project is given below. There are many problems associated with the mental health of humans especially who are working in online-digital platform. As humans are working for hours and hours in front of systems, laptops or any other smart devices, they might feel stress, anxiety, angry and happy etc., so they are not assessing themselves about their state of mind mentally while they are working continuously.

Figure 1.1: Person stressed due to work from home

Many psychological studies have found evidence that the human perception can discern various affective states expressed only through body movements.

Most of the recent research on emotion recognition is focusing on developing a system that can recognize emotions based on nonverbal cues expressed through body movements. Based on the above discussion, an increasing number of applications, that use body movement information for emotion recognition, has emerged. One of the recent works used a robot as a social mediator to increase the quality of human robot interaction. Emotion recognition from body movement encompass a large number of applications including biometric security, healthcare, gaming, and behavior modeling. Use of emotion recognition in the medical domain includes identification of the signature behavior of patients having specific psychological conditions.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, April 2022
Impact Factor: 6.252

Researchers have mostly attempted to recognize emotions from various modalities, such as the face, head, and hand. Very few studies have focused on whole-body expressions for emotion analysis. However, as stated in a computer model is not only suitable but may even exceed a human observer's ability to recognize emotion, as it can detect subtle movement changes not readily apparent to the naked eye. Moreover, body movement information can be obtained noninvasively from a distance which may be beneficial for many practical applications. Previous research focused only on a limited number of movement features from a vast number of computable features.

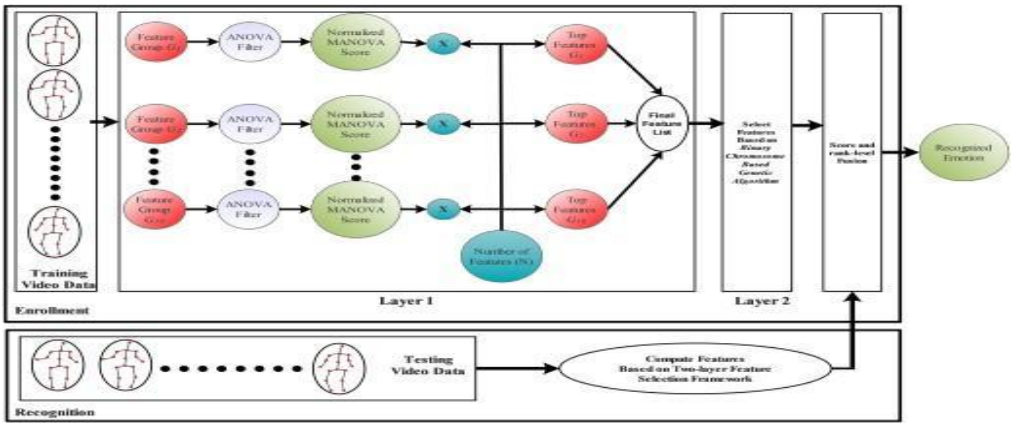

Figure 1.2: Network with Anova and Manova

In fig. 1.2, we leveraged the knowledge acquired from other disciplines such as computer animation and graphics for computing the body movement featuresThe number of features considered from each group was derived using normalized Multivariate Analysis of Variance (MANOVA) score computed for each group separately. Several popular feature ranking algorithms were investigated, including Mutual Information, Chi-squared Score, Relief and Ensemble of Decision Tree. The method outperformed all of the state-of-the-art approaches tested on our proprietary dataset. Information fusion techniques such as score and rank-level fusion further improved the emotion recognition accuracy of the proposed system. The proposed system also achieved 81.25% accuracy on a public dataset. outperforming existing state-of-the-art methods reported on this dataset. The overall contributions of the presented research are summarized as follows:

- Proposal of a unique structuring of motion features into ten groups, each describing a different aspect of a human body movement.
- Development of a two-layer feature selection architecture that combines the power of a traditional filter-based approach with a genetic algorithm.
- Identification of the most relevant motion features for emotion recognition from a comprehensive list of motion features. The relevance factor was computed for a univariate case where the features were considered independently, and a multivariate case, where features were considered as part of a group.
- Computation of feature relevance during two action scenarios, which provides an additional insight on importance of features during emotion recognition.
- Proposing a unique combination of score and rank level fusion with two-layer feature selection algorithm to maximize the emotion recognition accuracy.

II. Literature Review

Emotion can be expressed through eye gaze direction, iris extension, postural features, and movement of the human body. Pollick et al. showed that arm movements are significantly correlated with the pleasantness dimension of the emotion model. Bianchi-Berthouze et al. introduced an incremental learning model through gestural cues and a contextual feedback system to self-organize postural features into discrete emotion categories[1]. However, those works were limited to only parts of the body. Several researchers attempted to recognize emotion from dance movement. Camurri et al. in extracted the quantity of motion and contraction index from 2D video images depicting dance movements of the subjects to recognize discrete emotion categories[2]. Very recently, Durupinar et al. in conducted a perceptual study to establish a relationship between the LMA (Laban Movement Analysis) features and the five personality traits of a human. Senecal et al. in analyzed body

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, April 2022
Impact Factor: 6.252

motion expression in theater performance based on LMA features. Researchers have also focused on recognizing emotion in arbitrary recording scenarios using deep learning architectures. However, those attempts were limited to specific dance movements[3].

One of the biggest challenges of emotion recognition is the high dimensionality representation of the motion features. Also, the literature provides very little guidance as to what type of motion features are suitable for emotion classification. Most of the existing research have considered a very limited number of features[4]. Feature relevance was also not considered for emotion recognition. Therefore, most of the existing research is biased towards a particular set of motion features. For instance, Glowinski et al. in extracted energy, spatial extent, symmetry, and smoothness related features and then used Principal Component Analysis (PCA) to create a minimal representation of affective gestures. Saha et al. in picked nine features related to velocity, acceleration, and angular features to identify six emotions. This work successfully addresses the above deficiencies through the proposed comprehensive framework for emotion recognition, described in details in the next section[5].

III. Proposed Method

The first step of the proposed system involved the extraction of various geometric and kinematic features. Researchers have yet to establish a consensus on the right combination of various motion features. Therefore, in the proposed emotion recognition system, a comprehensive list of motion features was extracted maximizing the available body movement information. The motion features were computed either on a single frame or over a sequence of frames spanned over a short period. As a result, computed motion features characterize various aspects of human motion, such as trajectories or geometric properties of the postures.

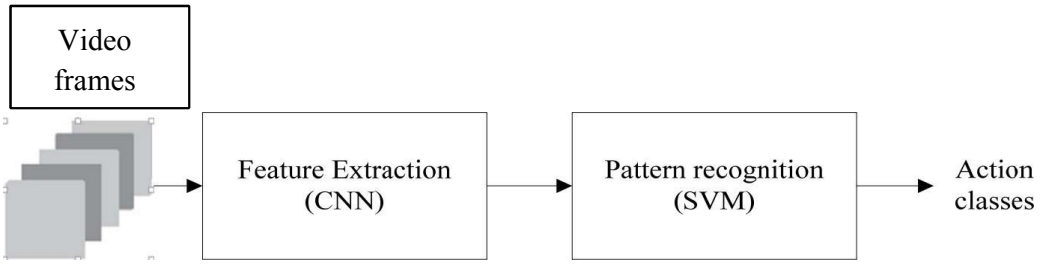

Figure 3.1: Implemented block diagram consists of CNN network and SVM Algorithm

A temporal profile was computed for each of the features. The temporal profile consists of twelve, time series functions, as described in section. A temporal profile computed in this way performs better than a histogram with fixed number of bins. The number of bins of a histogram determines the level of discretization of the calculated features. A limitation of using histogram is that the number of bins must be set empirically for the dataset. The values of a histogram are also sparse and most of the bins remain empty after the histogram computation. The main component of the proposed framework involves a two-layer feature selection process, as shown in fig. 3.1.

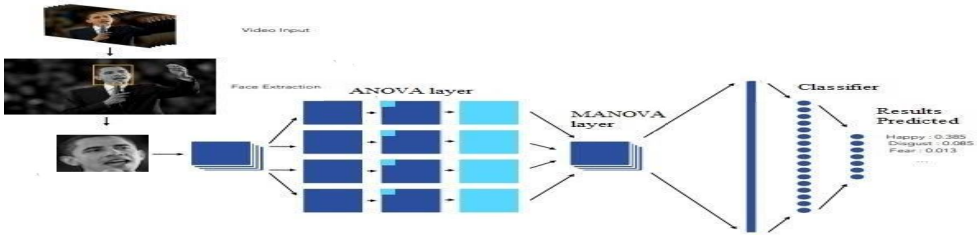

Figure 3.2: CNN model used in the model

In the fig. 3.2., the first layer, irrelevant features are eliminated using a combination of ANOVA and MANOVA. ANOVA is used to sort the features according to their relevance at recognizing emotions. ANOVA provides two measures: f-score and p-score to compute relevance of a feature. The f-score is a measure of total variation that exists among the arithmetic means of the target emotions. The p-score is a measure that determines the probability associated with rejecting the f-score. The features that failed to pass a significance test are discarded immediately. After removal of these features, remaining features are considered as statistically relevant for further analysis. MANOVA was used to compute group significance and

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, April 2022
Impact Factor: 6.252

to distribute features among various feature groups. The number of features considered from each group was derived using normalized MANOVA score computed for each group separately. The first layer may not be enough to attain optimum model performance based on the computed relevant features. The reason for this may be attributed to the performance improvement of specific feature combination for certain expert models. Thereby, several top features from each feature group were used as an input to the second layer of the framework. The objective of the second layer is to find the best subset of features that maximizes the emotion recognition rate of the expert models. Statistically relevant features were ordered based on the computed f-score. To reduce the number of possible combinations for computing feature subset that maximizes the emotion recognition rate, a predefined number of features were selected from the top ANOVA features. Typically, this number is set empirically. According to, the number of features can be selected as a function of the sample size, N, and the maximum feature size is N. In the proposed system, the total number of computed features was set based on the sample size of N. Since each group of features describes a different aspect of human body movement, the total number of features were distributed among the feature groups. Top ANOVA features were selected from each feature group based on the total number of features and the normalized MANOVA score computed for each group. The group significance scores were normalized so that each score ranges from 0 to 1 and their sum equals to 1. Then, the computed MANOVA score was used to distribute the total number of features from each motion feature group.

This way only some of the top features from each motion feature group remained for the subsequent steps. The features can be ranked based on their relevant factors, and irrelevant features can be removed based on an empirical threshold value. In our experiments, the features were ranked based on the computed f-score. From the top features based on the computed f-score, the features that produced a p-score, which was higher than a predefined threshold, was chosen for the genetic algorithm.

During the experiment, the p-score was chosen as 0.005. This ensures that there exists a minimal chance that the computed f-score was produced from a different distribution. In this way, the first layer used the relevance of the features to prepare for the second layer of the proposed feature selection framework. The second layer uses the genetic algorithm that evaluates the distinctive ability of the features to maximize emotion recognition accuracy. In the second layer of the two-layer framework, a binary chromosome-based genetic algorithm was used to identify the optimal feature subset that maximizes the emotion recognition rate. The genetic algorithm used in the proposed system achieved a plateau within 800 generations. The mutation rate was set to 0.03, as described in section IV. A detailed explanation of the genetic algorithm is presented in section IV. Finally, the expert models were fused using score and rank-level fusion as described in section IV-A. Figure shows how features for a feature group were selected using the first layer of a two-layer framework.

IV. Motion Feature Groups

Based on a thorough analysis of the existing literature, a comprehensive list of 3D motion features was extracted. Group of Features 1 This group of features consists of low-level feature descriptors that measure the speed of the motion, such as velocity, acceleration, and jerk. If X defines a motion that is described as n consecutive poses, where X = x(t1), x(t2), x(t3), ..., x(tn). Then, the velocity is defined in equation 1 and the magnitude of the velocity is determined using the equation 2 accordingly. In equations 1 and 2, v k (ti) is the velocity of the k th joint at time ti , v k x (ti) is the x component of the velocity of the kth joint at time ti , and δt refers to a small fraction of time required for transitioning between consecutive frames. Usually, δt is set to a very small value. During the experiment, the value was set to 1 30 seconds as Kinect v2 has a frame rate of 30 fps.

The indices of the ones' correspond to the indices of the selected features. The population size of a genetic algorithm affects the ability of exploration of the feature space. If the value is set too low, the genetic algorithm may not produce enough variability among the chromosomes. For this reason, the population size of the chromosomes was set empirically to 30. The fitness function determines the ability of a chromosome to survive a generation of reproduction. The main goal of using the genetic algorithm is to find a subset of features that maximizes the emotion recognition rate. Therefore, the emotion recognition rate was an obvious choice for the fitness function. Each chromosome in the population was evaluated using the fitness function. The list of chromosomes was sorted based on the result of the fitness function. Half of the population chromosomes were automatically scheduled to survive when the next generation of chromosomes were reproduced. The remaining half of the population was chosen based on a crossover operator performing recombination among the top half of the chromosomes. The crossover operation used in the proposed system is shown using equation. Equations shows how

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, April 2022
Impact Factor: 6.252

crossover operator reproduces chromosomes at a crossover point x from two chromosomes C1 and C2 in a m-dimensional feature space.

The crossover operator chooses a crossover point with a uniform probability distribution for each pair of consecutive chromosomes survived for the next generation. Thus, the crossover point was chosen randomly for each pair of chromosomes for reproduction. The mutation operator introduces randomness so that the crossover operation can avoid repeated reproduction of the same chromosomes. Mutation rate is a hyper parameter to balance exploitation and exploration ability of a genetic algorithm. If this value is set too high, genetic algorithm may not converge to a plateau during which the maximum recognition rate is not changed. If this value is set too low, genetic algorithm may get stuck in a local maxima. Typically, the mutation rate is set to a small value of 2–5%. In our experiments, mutation rate of 0.03 was chosen. The mutation rate indicates that there is a 3% probability of reversing a single bit value randomly in a chromosome

V. Results and Discussion

The model developed using CNN & SVM for facial features extraction, we have trained the model using the data sets and have got an accuracy of 98%. We have modified an existing model to detect the human activity and have increased the accuracy of prediction from 50% to almost 60%.

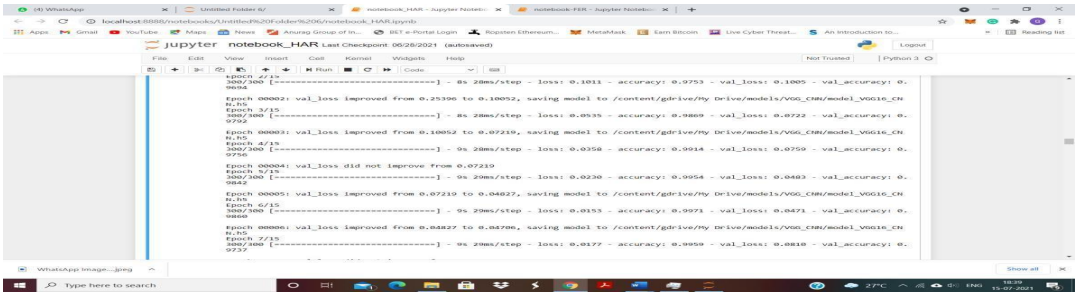

Figure 5.1: Results after testing the model for facial emotions

It can be observed from the fig. 5.1 that the model developed has an accuracy in between 97% and 98% for the facial emotion recognition. We can conclude that the model has high accuracy to predict the facial emotions.

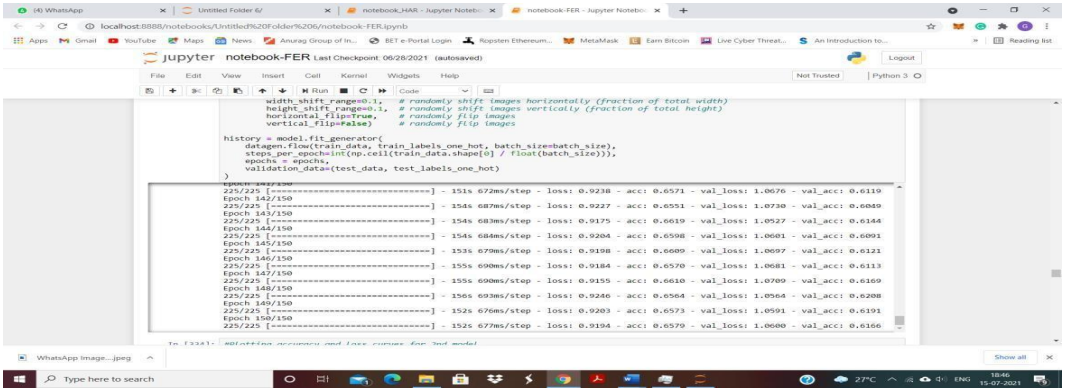

Figure 5.2: Results after testing the model for Human Activity.

It can be observed from fig. 5.2 that the model developed for human activity recognition has an accuracy in between 65% and 69%, the reason for it being the camera resolution and angle of capture. Both the models designed for Facial Emotions and Human activity are clubbed into a same HTML page using JAVA& JSON script, taking help of inbuilt libraries, which makes it user friendly to use and to asses the results obtained from the graphs. The code for web pages is written in HTML 5 which is easier to use.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, April 2022
Impact Factor: 6.252

Figure 5.3: Initial page

When the program is made to run using the command prompt window, the web page in fig. 5.3 appears which consists of facial detection and human activity detection. The facial emotions of a human are captured live from the internal webcam of the laptop or system the person is using. As soon as the user clicks on the 'cam' button the page displayed in fig. 5.4 appears, and prompts the user to start using webcam for the capture of live video. The video is captured from the webcam and the program used to detect various human emotions is made to run in the background. The code divides the video captured into frames which makes it easy to detect the emotions.

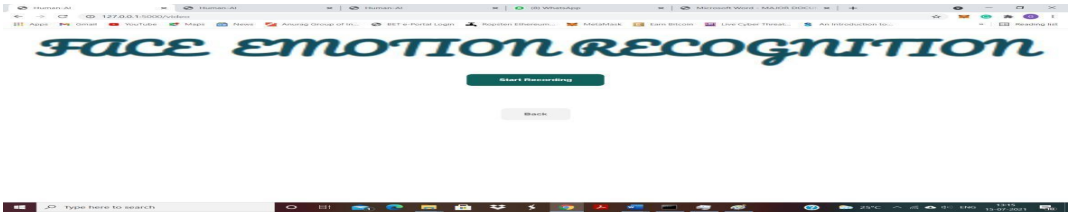

Figure 5.4: Page prompted for webcam to start.

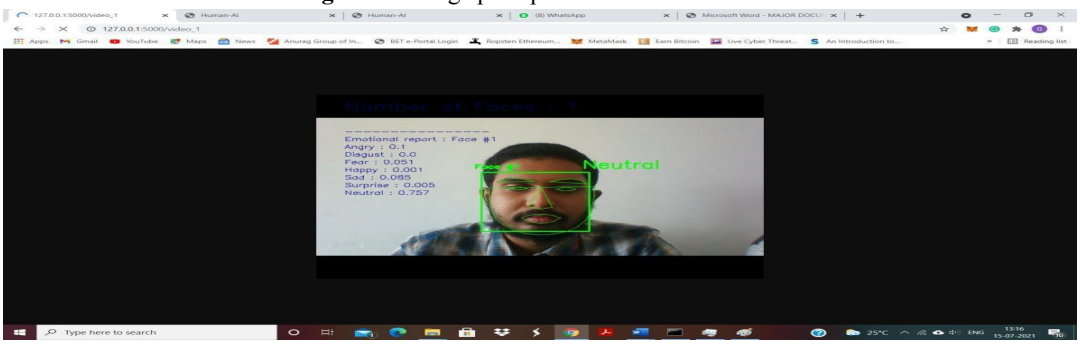

Figure 5.5: Detection of neutral emotions.

The code that is running in the background detects various types of emotions from the set of images that it is trained for and displays the probabilities beside the face of the person as shown in fig. 5.5. and fig. 5.6. The average of all the values that are calculated is converted into a graph, which are easy to perceive than the numerical values.

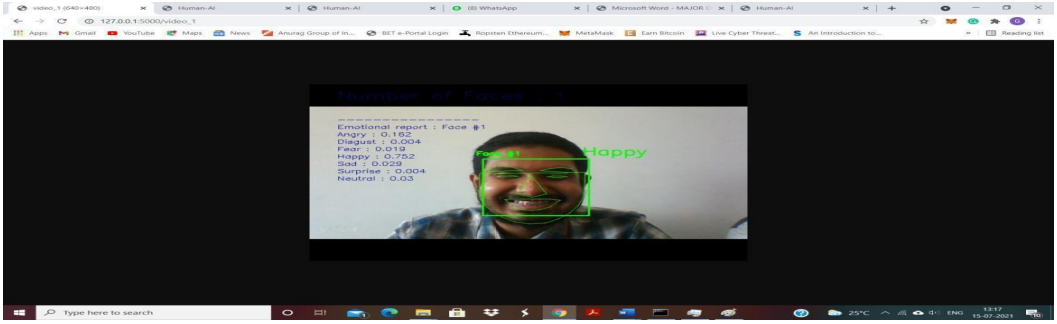

Figure 5.6: Detection of happy emotion.

The fig. 5.5 captures neutral emotion and it can be seen that the probability shown by the trained model is 0.75 which means that the model is able to detect the happy emotion with a percentage of 75%. The fig. 5.6 captures happy emotion

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, April 2022
Impact Factor: 6.252

with a probability percentage of 75%.

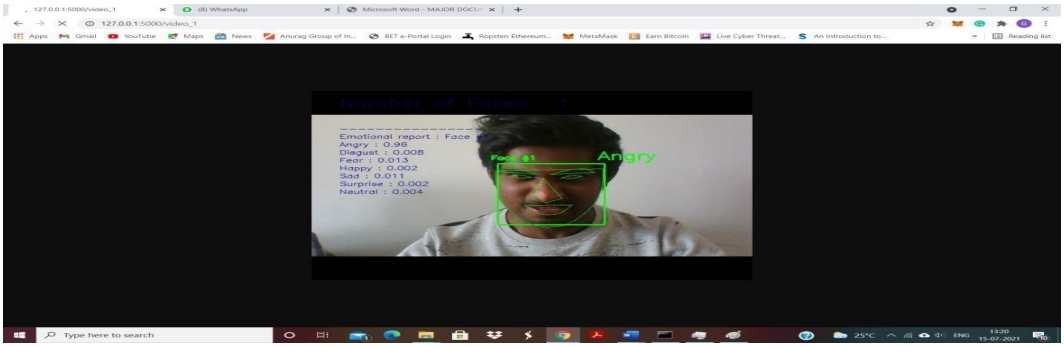

Figure 5.7: Detection of Angry emotion.

The fig. 5.7 detects the Angry emotion of the person with the accuracy of 96% which is better when compared to many other available models.

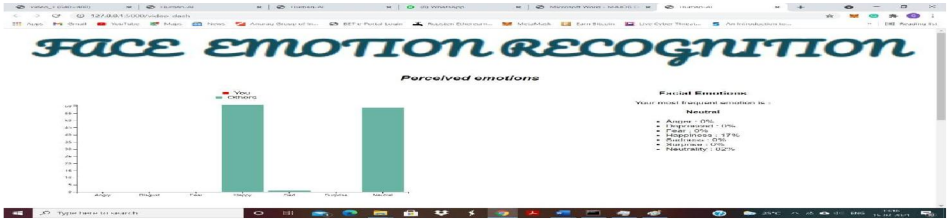

Figure 5.8: Graphical representation of recognized emotions.

Figure 5.9: Overall average of emotions detected.

Graphs are always a better way to understand the results, the graph in fig. 5.8 shows the results of the video captured live. The fig. 5.9 shows the average of all the predictions done till date which can be used to analyze.

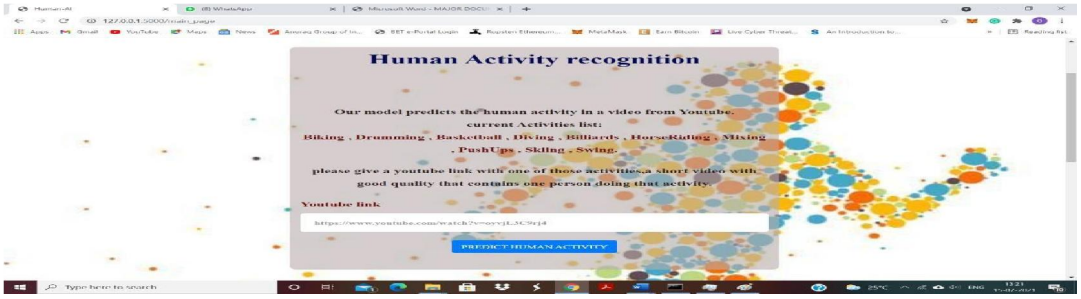

Figure 5.10: Human activity page for complex activity recognition.

The page displayed in fig. 5.10 is the page where the user has to give the link of YouTube, the trained model runs in the background and predicts the activity in the video.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, April 2022
Impact Factor: 6.252

Figure 5.11: YouTube video given as input.

The video shown in the fig. 5.11 is given as the input video, the link is copied from the chrome browser and is given the box below as shown in fig. 5.12

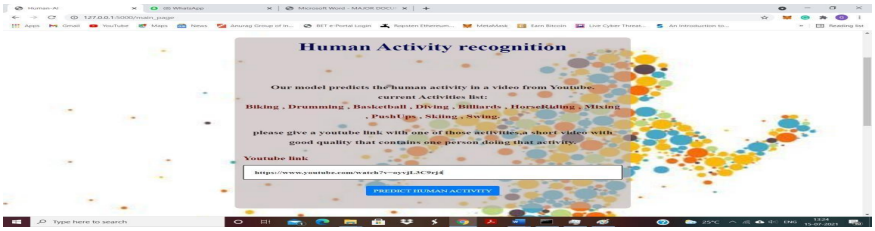

Figure 5.12: Activity prediction page.

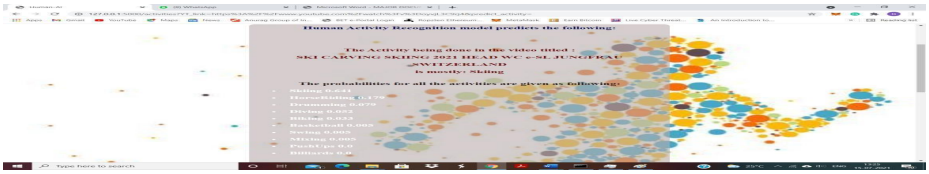

Figure 5.13: Results from the predicted model trained.

After giving the link in the activity page, the video is taken by the code and processes the video and produces the results as shown in fig. 5.13. The input video given is of Skiing which is a 30 seconds video and the results produced can be seen that the video is of skiing which shows a probability of 64%.

It can be seen that the trained model for Facial Emotions has a prediction percentage of Anger with 98% which the other model has not achieved. The main reason for achieving such high probability is because of the use of facial geometry points, as the geometry of face is different for different persons.

The model trained for the prediction of Human Activity has a probability of 65% which is less due to the alignment of camera from which the video is captured, the detection percentage can be increased by clearly capturing the video using a good quality camera.

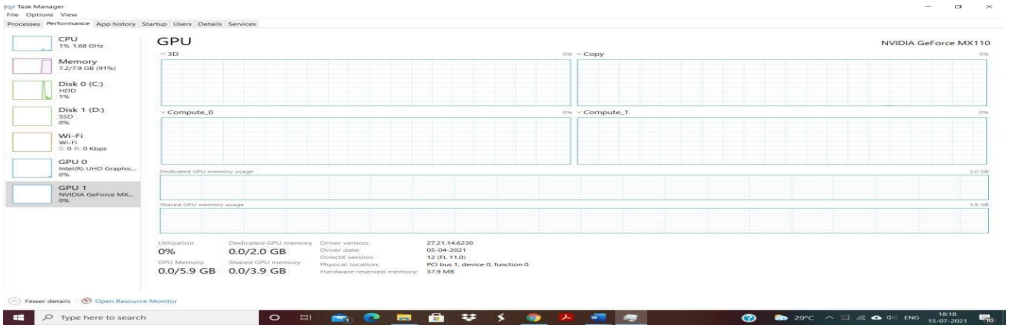

Figure 5.14: System utilization.

The fig. 5.14 shows the resources used by the system to predict activities of Human and facial recognition. It is not that the code has not used GPU for the computation and the developed code can be run on machines with no GPU. The only disadvantage is with the occupancy of the RAM used by the code in order to compute the results.

The process of the computation of results can be improved if the user uses an advanced process such as Intel i9 or any other, for this project we have used Intel i5 8th generation processor. The results will vary if a different processor is used, in terms of results display and computation power and speed.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 2, April 2022
Impact Factor: 6.252

VI. Conclusion And Future Scope

In this paper, emotion recognition using body movement is implemented by using facial features extraction and pattern recognition. This project has addressed the problem of creation of a complete system that can accurately recognize five basic emotions: happiness, sadness, anger, fear, and neutral based on body movement features. Significant benefits of this project can be achieved for biometric security, patient behavior monitoring, gaming, and robotics with the creation of a movement-based emotion aware computer system. Body movement information can provide valuable cues related to the emotional state of a person. Despite showing great potential to be an essential indicator of perceived emotions, body movement information is one of the least explored modalities for emotion recognition.

The experimental results showed that it is possible to build a computer system capable of recognizing human emotion only based on body movement information. Experiment results provided critical information regarding the perceived emotion in walking and sitting action scenarios. During walking action, the quantity of movement in the arm and the upper body region were essential indicators. On the other hand, body space utilization, elbow angle, and spatial extension were essential cues to recognize emotion during the sitting action. During action-independent cases, motion features important during all action scenarios need to be considered to maximize the emotion recognition rate.

References

[1]. N. Fragopanagos and J. G. Taylor, "Emotion recognition in human-computer interaction," Neural Networks, vol. 18, n0. 4, pp. 389-405, 2017.

[2]. P. Tarnowski, M. Kolodziej, A. Majkowski, and R. J. Rak, "Emotion recognition using facial expressions," Procedia Computer science, vol. 108, pp. 1175-1184, 2017.

[3]. F. Noroozi, C. A. Corneanu, D. Kaminska, T. Sapinski, S. Escalera, and G. Anbarjafari, "Survey on emotional body gesture recognition," arXiv preprint arXiv:1801.07481, 2018.

[4]. H. G. Wallbott, "Bodily expression of emotion," European Journal of Social Psychology, vol. 28, no.6, pp. 879-896, 2018.

[5]. XIAOTONG ZHANG, "InnoHAR: A Deep Neural Network for Complex Human Activity Recognition", 10.1109/ACCESS.2018.2890675, 2019.