Personalized System for Human Gym Activity Recognition using an RGB Camera

Preetham Ganesh*
Department of CSE
University of Texas at Arlington
United States
preetham.ganesh@mavs.uta.edu

Reza Etemadi Idgahi
Department of CSE
University of Texas at Arlington
United States
reza.etemadiidgahi@mavs.uta.edu

Chinmaya Basavanahally Venkatesh
Department of CSE
University of Texas at Arlington
United States
chinmaya.basavanahallyve@mavs.uta.edu

Ashwin Ramesh Babu
Department of CSE
University of Texas at Arlington
United States
ashwin.rameshbabu@mavs.uta.edu

Maria Kyrarini
Department of CSE
University of Texas at Arlington
United States
maria.kyrarini@uta.edu

ABSTRACT

Human Activity Recognition is one of the most researched topics in the field of computer vision. It is a powerful tool mainly used to aid medical systems, smart homes, surveillance, and many more areas. In this paper, an RGB camera was used to record gym activities such as push-up, squat, plank, forward lunge, and sit-up. Features were extracted from the recorded videos and were fed into classification algorithms such as Support Vector Machines, Decision Tree classifier, K-Nearest Neighbor classifier, and Random Forest classifier. The developed models were evaluated using metrics such as accuracy, balanced accuracy, precision score, recall score, and F1 score. The Random Forest Classifier outperformed all the other attempted methods with an accuracy of 98.98%. A repetition counter was developed, which splits workouts based on local minima analysis, and correctness of the workout was calculated for each skeletal point using dynamic time warping. An interactive android application was built for the user to gain insights on the performed workouts.

CCS CONCEPTS

• **Human-Centered Computing** → Gestural input; Information Visualization; • **Computer Vision** → Scene understanding; Activity recognition and understanding; Object Recognition; Tracking; • **Machine Learning** → Supervised learning by regression; Transfer Learning; Neural Networks.

KEYWORDS

Human Activity Recognition, Gym Activity Recognition, Open Pose, Classification

*Corresponding author.

PETRA '20, June 30-July 3, 2020, Corfu, Greece

ACM ISBN 978-1-4503-7773-7/20/06…$15.00
https://doi.org/10.1145/3389189.3392611

ACM Reference Format:
Preetham Ganesh, Reza Etemadi Idgahi, Chinmaya Basavanahally Venkatesh, Ashwin Ramesh Babu, and Maria Kyrarini. 2020. Personalized System for Human Gym Activity Recognition using an RGB Camera. In *The 13th PErvasive Technologies Related to Assistive Environments Conference (PETRA '20), June 30-July 3, 2020, Corfu, Greece.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3389189.3392611

1 INTRODUCTION

Machine Learning has its impact in almost every field. Improvement in hardware has paved the way to perform complex computation even on small personal devices such as tablets, mobile phones, and wearable devices. Extensive work has been done to apply machine learning in building mobile applications to perform various tasks such as detection of malicious android application [34, 37], driver behavior prediction [24], identification of alcohol use in young adults [30], gesture recognition for cognitive assessment [3, 9] and many more. Most of these applications were designed to assist users, thereby reduce their efforts in those tasks.

Data that was collected in the year 2015 stated that nearly forty percent of Americans suffered from obesity, and thirty-two percent were overweight [6]. The authors in [1] stated that a sedentary lifestyle leads to higher risks of cardiovascular disease, diabetes, depression, and many more. According to work done in [11], the authors concluded that high muscular and cardio-respiratory fitness improves the overall quality of life. A healthy lifestyle can be achieved through proper food habits, sound sleep, and frequent workouts.

There are many mobile phone applications and wearable technology services for tracking food, sleep, and fitness. Tracking food is trivial as it requires the user to enter the food consumed, and the application provides insights on the data entered. Similarly, tracking sleep is achieved through various wearable sensors, which give insights on sleeping patterns. When it comes to fitness, one of the significant concerns with the existing mobile applications is that it cannot validate the correctness of the workouts. This problem can be solved by hiring a trainer/expert, which might be expensive. Our proposed system uses computer vision and machine learning to monitor the correctness of the workout performed and provides personalized feedback.

The composition of the proposed work is split into five sections and is listed as follows. Section 2 discusses the list of previous works related to the various approaches used by researchers to solve problems in Human Activity Recognition (HAR). Section 3 talks about the system architecture used for the proposed system. Section 4 discusses the experimental results obtained using the proposed method. Section 5 concludes the paper based on the achieved results.

2 RELATED WORKS

This section talks about the previous works done by researchers in the field of HAR using both Inertial-Sensor based and Vision-based approaches. It also discusses the works related to Human Gym Activity Recognition and draws a comparative study with the various approaches used in the referred papers, along with their limitations.

Human Activity Recognition (HAR) mainly focuses on recognizing actions performed by one or more agents, from a series of observations on agents' actions and the environmental conditions. HAR has various applications in real-life such as assistive living in a smart home [19, 29], health care monitoring system [12, 22, 23], and many more [28]. Different sensors such as RGB camera, motion sensors, depth camera, and many more can be used to obtain data of the activities performed. In general, there are two approaches to solve HAR problems, namely, Inertial-Sensor based approach and the Vision-based approach [10].

In the Inertial Sensor-Based approach, an inertial sensor is attached to the body of the user performing the activities. The data of the activity performed is collected using the inertial sensor and is fed into the classification algorithms [33]. It has been used by Hsu et al. in [13] to develop an inertial sensor network capable of recognizing 10 domestic activities and 11 sports activities. The developed inertial sensor network consisted of an accelerometer and gyroscope. Also, Lee et al. in [31] used data from an accelerometer to classify daily human activities, where a one-dimensional Convolutional Neural Network (CNN) was used for the classification.

The Vision-based approach uses an RGB or RGB-D camera to record the participants' activities. The recorded videos are fed into CNN-based feature extraction algorithms, and the extracted features are fed into classification algorithms [25]. It has been used by Babiker et al. in [2] to develop a computer vision-based surveillance system capable of recognizing malicious human movement. For this purpose, the authors used various digital image processing techniques and a multi-layer feed-forward perceptron neural network to classify the activities.

Liu et al. in [20] used a distributed RGB-D camera network to capture action sequences from different angles and combine using Information-Weighted Consensus Filter (ICWF). Based on the analysis, the authors concluded that the model improved the accuracy of action recognition. Bakalos et al. in [4] used a deep-NARMA filter with multi-modal fusion to recognize unusual behavior. The authors concluded that the model was able to adapt to dynamic events such as unusual behavior. Kosmopoulos et al. in [17] used a Bayesian filter based classification supported by Hidden Markov Model (HMM) to recognize behavior in workflows. The developed model produced promising results in real-time when tested with challenging real sequences.

Devanne et al. in [8] used a Dynamic Naive Bayes classifier to model segmented human motions from an RGB-D setup for capturing the different dynamics of human motion. The authors concluded that the developed model outperformed the other state-of-the-art methods for the online detection of human behavior. Wu et al. in [35] developed a 3D motion scale-invariant feature transform for describing the information in RGB-D videos. The output was then fed into an HMM to recognize human behavior. The authors concluded that the model provided better recognition performance in real-time.

The problem statement chosen in this paper is a subset of HAR called Gym Activity Recognition (GAR), where the chosen activities are a set of exercises and the personalized system was created using an android application, where each user can use the application to track the details of his/her workouts. Many authors have tried to solve GAR using different approaches by using different sensors. Qi et al. [26] used an electrocardiogram for recognizing the gym activities where a neural network was used for classifying sedentary and aerobic activities, and an HMM was used classifying free weight activities. The electrocardiogram was attached to the subject's chest for this purpose.

The authors in [5] used a passive capacitive based approach to detect gym activities and count the repetitions. A microwatt level power consumption based sensor was attached to one of the subject's leg. Qi et al. in [27] used a wristband and a belt based accelerometer to detect and classify physical activities where the obtained data was then fed into an Artificial Neural Network for classification. Ten subjects wore the wrist band and the belt for obtaining the data. In [38], the authors used an accelerometer for recording the subject's data, which was then used for classifying the gym activities with similar movements and calculating the amount of energy exerted by the subject in performing the same.

A personalized system using HAR was developed, keeping in mind that every person differs in height, weight, and current needs. Authors have developed many machine learning-based personalized systems such as recognition of driver state [36], a textile-based sensor system for health care [21], machine learning-based heating system [16], and many more. Generally, when a machine learning-based personalized system is developed, the machine learning algorithms are trained on an existing dataset or a dataset developed by the authors to solve a specific problem. The trained model is then used for predicting the result on the data generated by the user and provide insights specific to the user's data. Many authors have used a similar setup for developing applications such as facial recognition based course attendance [32], hand gesture recognition [18], computer vision-based assistance for visually impaired [15], and many more.

The model proposed in this paper comprises of a personalized system used for giving feedback to the user using an interactive android application, where the feedback is for a gym activity performed in front of a web camera. The feedback consists of the following: (1) name of the activity performed; (2) number of repetitions; (3) the level of correctness at which the activity was performed.

The approach used by the authors in [4, 8, 20, 35] for recording the data (i.e., using RGB-D camera setup) performed by the subject

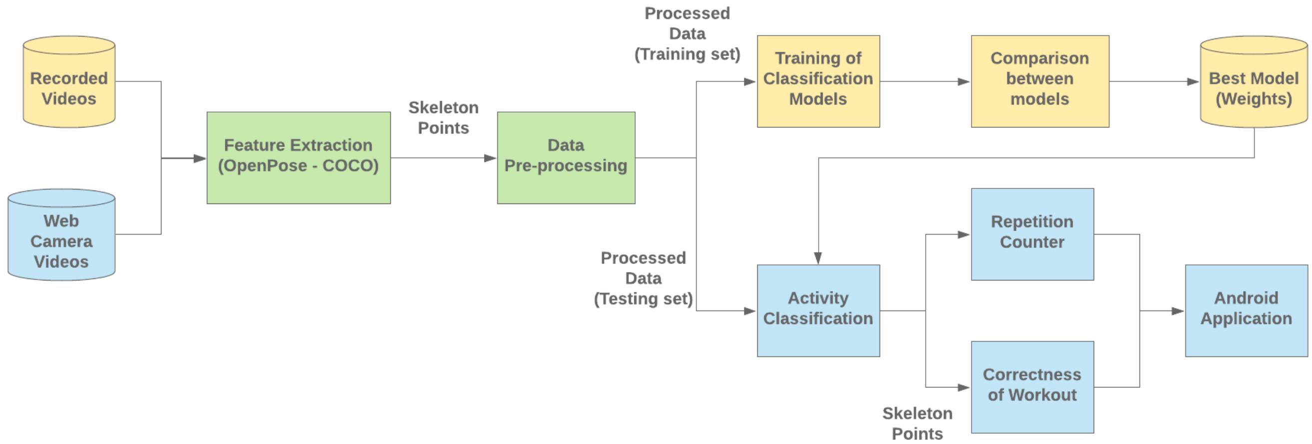


Figure 1: System Architecture

was not used in this paper because the paper was aimed at providing the end-user a simple product with cost-effective components. Similarly, the approach in [17] was aimed at detecting unusual behaviors of multiple users in a video frame, whereas this paper was focused on developing a model for a single user in a video frame. Also, the various works mentioned above under the area of GAR use wearable sensors for implementation, which are obtrusive, whereas our proposed system is non-obtrusive.

3 PROPOSED SYSTEM

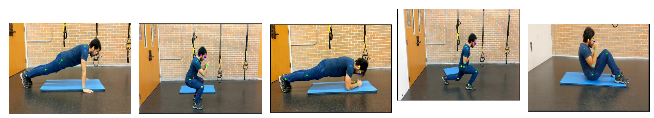
Fig. 1 shows the system architecture used for the proposed system and the modules in the figure are colored based on the following basis: (1) modules that were used only during the training stages are colored yellow, (2) modules that were used only during the testing stages are colored blue, and 3) modules that were used only during both the training and testing stages are colored green.

3.1 Dataset Creation

The videos used to train the models were recorded using an RGB camera in the Google Pixel 3a mobile phone, and the videos used to test the models were recorded using a Logitech Web Camera. The videos were recorded with a frame size of 1920 x 1080 at a frame rate of 30 frames per second. The recorded videos contain five gym activities, namely: push-up, squat, plank, forward lunge, and sit-up, and four subjects were asked to perform the workouts individually. All the subjects were of gender male, with a mean age of 21.75 (standard deviation = 1.758), and with a mean height of 168.96 cm (standard deviation = 3.793).

Each subject had to perform one repetition of plank for 10 seconds and 15 repetitions of the remaining workouts. Only one repetition of the plank was recorded because the subject's body movement does not change throughout the activity. The examples of frame representation for each of the chosen activities are shown in Fig. 2.

It can be observed from all the images in Fig. 2 that only the right-hand side of the subject's body is visible to the camera. This configuration was chosen because the end product will be using a single web camera, and the objective was to capture the maximum



(a) Push-up **(b) Squat** **(c) Plank** **(d) Forward Lunge** **(e) Sit-up**

Figure 2: Examples of frame representation for each of the activities

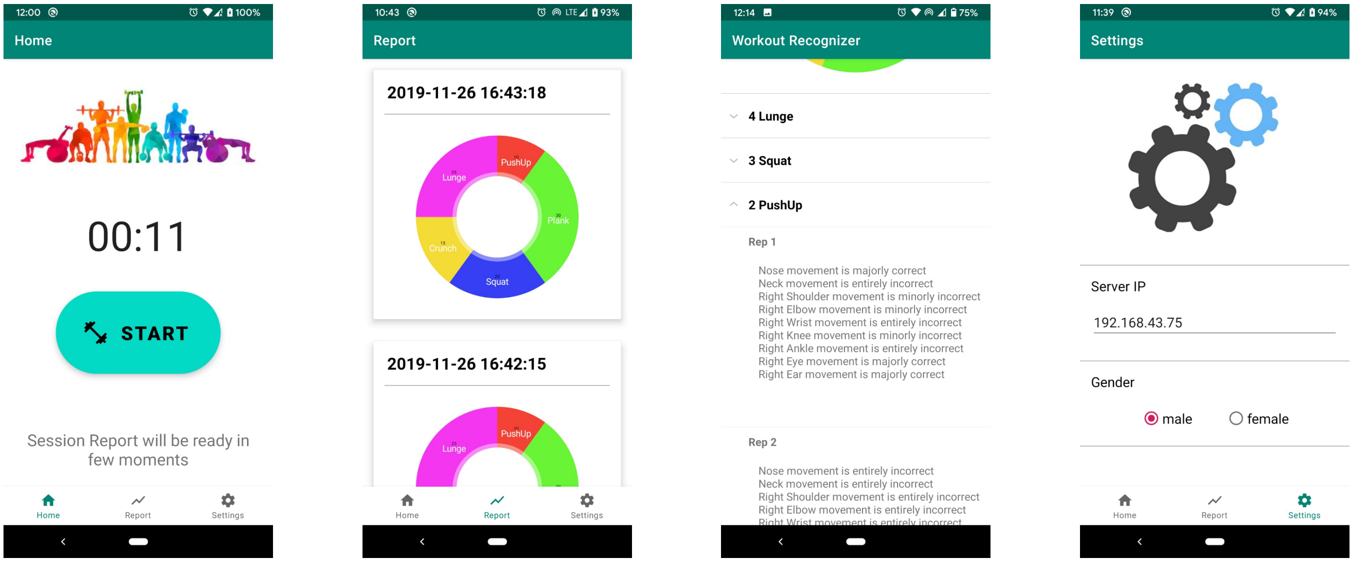
number of skeleton points in a frame. It was also assumed that the movement of the left-hand side of the subject's body would be more or less similar to the right-hand side movement for the chosen list of activities.

3.2 Feature Extraction

Each of these videos was processed using the trained CNN called OpenPose used in [7]. The videos were loaded and saved as a list of frames. Each frame was resized from 1920x1080 resolution to 640x480 resolution to reduce the time taken to provide the feedback to the user. The OpenPose model used the resized frames to generate the skeleton points, where the model was compiled on an Apple MacBook Pro (13 inches 2019) with a CPU configuration of Intel 8^{th} Generation Core i5 processor and GPU configuration of Intel Iris Plus Graphics 645. The average execution time was 4 seconds per frame, and a total of 18 skeleton points were generated for each frame, which includes nose, neck, eyes, ears, shoulders, elbows, wrists, hips, knees, and ankles [14]. The generated skeleton points were saved as a text file for further processing.

3.3 Data Pre-Processing

The text file generated for every video was converted into a Comma Separated Values (CSV) file, which has 37 columns. The first 18 columns represent the X values of skeleton points in the frame; the next 18 columns represent the Y values of skeleton points in the frame, and the last column represents the frame number. The generated CSV file for every video had a lot of missing value columns



(a) Home Page **(b) Report Page** **(c) Detailed Workout Page** **(d) Settings Page**

Figure 3: Screenshot of pages on android application

because only the right-hand side of the body was visible to the RGB camera during the recording session. Hence, columns belonging to the left-hand side of the body were dropped.

The balance missing values in the data frame were imputed using the frame numbers, i.e., if a missing value was found in a subject's data, then the frame number was chosen and was then used for obtaining the values from the corresponding columns of the subject's other data for the same activity. If the values were found, the mean was taken and imputed in the current file, and if no value was found, the current file's corresponding column's mean value was imputed. Except for the frame number column, all the columns in the CSV file were normalized using the Min-Max Normalization, as it helped in calculating the correctness of workout performed by the user. The formula used for normalization is shown in (1). The normalization was performed for each video separately.

$$X'_{\mathrm{i}} = \frac{X_{\mathrm{i}} - X_{\mathrm{min}}}{X_{\mathrm{max}} - X_{\mathrm{min}}} \tag{1}$$

Where X_i is the i^{th} element in the feature, X_{min} is the minimum value of the feature, X_{max} is the maximum value of the feature and X'_i is the normalized value of the i^{th} element in the feature. The normalized columns were added into a dictionary along with the frame number column and converted into a CSV file. The CSV files from each video were combined into a single file for the training set. However, for the testing set, there is only one video; hence, there is no combining of CSV files.

3.4 Android Application

The workout patterns of the user were monitored using an android application, which consisted of four pages, namely home page, setting page, report page, and workout detail page. Django server was

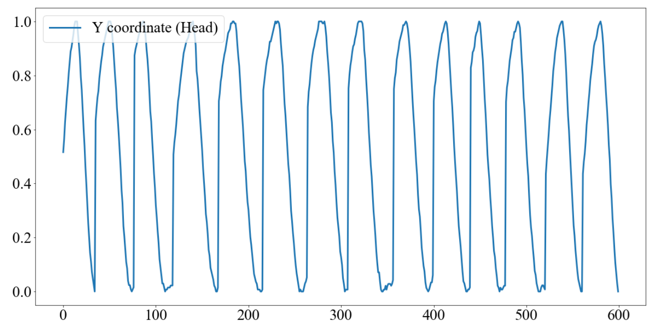


Figure 4: A sample of the y coordinate movement of the head skeleton point in a pushup workout

used to integrate and run the web camera and received the processed data and results. The android application pings the Django server for every 10 seconds to check whether the required results were available to populate the android application. Fig. 3a depicts the home page where the user can start and stop the time of his/her exercise. Fig. 3b represents a pie chart with the number of repetitions for each workout. Fig. 3c represents the detailed workout page, along with the accuracy of each posture. Fig. 3d represents the settings page where the user sets the connection to the Django server.

3.5 Repetition Counter

This module is used during the testing phase, where the user clicks the start button in the android application before performing the workout and clicks the stop button after completing the workout. Once the stop button is clicked, the feature extraction module in Section 3.2 starts extracting skeleton points from each frame.

Once all the frames are processed, the data is then fed into the pre-processing module in Section 3.3, and the processed data is then fed into the best model in Table 2 in a tuple by tuple manner, where each tuple consists of normalized skeleton points.

Once the prediction is performed, the data is split into multiple CSV files based on the movement of the nose skeleton point in the user's workout. The sample output is given in Fig. 4. More specifically, when a local minimum is obtained in the movement. Since the model is not 100% accurate, the majority count of the classified activity in the separated CSV file is taken as the activity for the file. The total count is also updated based on this final activity per file (repetition). As a whole, when using a CPU for computation, the system takes about 4 minutes to process and provide output for a workout of 30 seconds in length.

3.6 Correctness of the Workout

Subject 2 happened to be a personal trainer. Hence his workout data was chosen as the reference file for each of the activities. So each of the new user's workout data is compared with the corresponding reference file for each selected skeleton point. Since the speed at which the new user performs the workout would be different, a distance measure called as dynamic time warping was used for calculating the correctness. For each skeleton point, the calculation is done separately and is displayed in the android application. The correctness value for each skeleton point was split into six categories, which are shown in Table 1.

Table 1: Categories of Correctness of Workout Splitting

Range	Definition
>100	Movement is entirely incorrect
<100 and >80	Movement is majorly incorrect
<80 and >60	Movement is minorly incorrect
<60 and >40	Movement is minorly correct
<40 and >20	Movement is majorly correct
<20 and >0	Movement is entirely correct

4 RESULTS AND DISCUSSIONS

This section discusses in detail the results obtained on applying the Pearson Correlation Coefficient on the training CSV file generated in Section 3.3. It also discusses the results obtained on feeding the CSV file into the classification algorithms.

4.1 Correlation between the Skeleton Points and the Workout

The correlation was found between X values, Y values of the skeleton points, and the activity using the Pearson Correlation Coefficient. The formula for which is given in the equation in (2).

$$\rho_{\mathrm{A,\,B}} = \frac{\sum(A_{\mathrm{i}} - A_{\mathrm{m}})(B_{\mathrm{i}} - B_{\mathrm{m}})}{\sqrt{\sum(A_{\mathrm{i}} - A_{\mathrm{m}})^2 \sum(B_{\mathrm{i}} - B_{\mathrm{m}})^2}} \quad (2)$$

Where A and B are the continuous variables, A_i and B_i represent the i^{th} element in the vectors and A_m, and B_m are the mean values of the corresponding vectors. A heat map was used for visualizing the generated correlation matrix, where the image in Fig. 5 shows the correlation between the X values of the selected skeleton points and the activity.

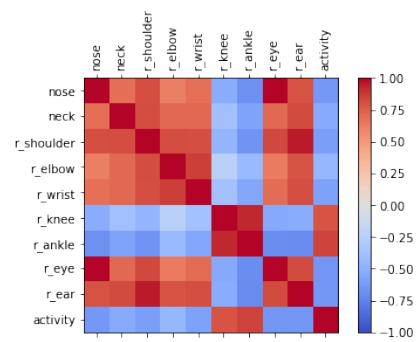


Figure 5: Correlation between the X values of the selected skeleton points and the activity

It can be observed from Fig. 5 that the skeleton points at the nose, neck, right shoulder, right elbow, right wrist, right eye, and right ear were negatively correlated with the activity, and the other attributes were positively correlated with the activity. It is because of the location of the skeleton points.

4.2 Workout Classification

The combined CSV file was fed into the following classification algorithms: Decision Tree classifier, Support Vector Machine, K-Nearest Neighbor classifier, and Random Forest classifier. The attempted models were validated using repeated k-fold cross-validation, where the number of folds is ten, and the number of repeats is ten. The total number of iterations is 100, and in each iteration, nine folds of data would be used for training the model, and the balance one fold would be used for validating the trained model. The classification models were evaluated using metrics such as accuracy, balanced accuracy, macro precision score, macro recall score, and macro F1 score. The performance of the algorithms on the CSV file is shown in Table 2.

It can be observed that all the models have high performance in all the metrics. It should be noted that the results shown in Table 2 used the default values of the parameters in the algorithm. It can also be observed that the Random Forest classifier model outperforms all the other attempted models and hence was chosen as the best model for the personalized system. The confusion matrix for the Random Forest classifier is given in Table 3.

It can be observed from Table 3 that the model recognized most of the workouts with a very high distinctive accuracy. It can also be observed that push-up, plank, and sit-up workouts had minimal misclassification rates, and the squat had the lowest recognition rate of 97.6%.

4.3 Correctness of Workout

The distance between each repetition for each activity of subjects 1, 3, and 4 are compared with the selected file of subject two's corresponding workout data. The average value of distance for each skeleton point for all the activities is given in Table 4. From

Table 2: Performance of the classification algorithms on the CSV file

Model	Accuracy	Balanced accuracy	Precision score	Recall score	F1 score
Support vector machine	0.8833	0.8833	0.9099	0.7703	0.8008
K-nearest neighbor	0.9779	0.9779	0.9841	0.965	0.9737
Decision tree classifier	0.965	0.965	0.9533	0.9493	0.951
Random forest classifier	**0.9898**	**0.9898**	**0.9911**	**0.9835**	**0.9872**

Table 3: Confusion matrix of the random forest classifier model for the validation data

		Predicted Label				
		Push-up	**Squat**	**Plank**	**Forward Lunge**	**Sit-up**
	Push-up	359	3	0	0	0
	Squat	0	414	0	10	0
True Label	**Plank**	0	0	37	0	1
	Forward Lunge	0	6	0	528	0
	Sit-up	0	0	0	1	536

Table 4: Result on using Dynamic Time Warping

Activity	Subject	Nose	Neck	R Shoulder	R Elbow	R Wrist	R Knee	R Ankle	R Eye	R Ear
	1	8.988	16	9.117	15.578	26.735	10.896	33.508	8.915	8.262
1	3	6.389	13.31	4.602	9.697	18.738	23.332	35.197	6.115	5.216
	4	7.387	12.412	5.67	16.835	22.82	14.272	20.184	7.512	5.569
	1	13.076	14.349	13.28	12.647	16.923	18.505	24.803	12.529	13.254
2	3	14.118	9.851	13.356	9.445	19.219	12.558	18.711	13.217	11.378
	4	13.194	13.569	13.31	15.106	16.42	18.413	22.377	13.146	13.355
	1	30.125	43.169	35.014	35.534	27.134	26.729	48.025	32.742	34.933
3	3	28.697	51.996	49.356	59.986	71.681	53.776	69.676	25.08	47.075
	4	51.038	56.639	33.197	48.94	56.73	35.949	39.652	56.344	51.521
	1	3.943	2.949	3.145	4.013	4.575	12.404	25.956	3.648	3.133
4	3	6.418	7.562	7.495	10.043	10.081	13.84	28.593	6.599	6.965
	4	5.079	4.938	5.037	6.723	6.921	24.281	31.082	5.027	4.592
	1	8.709	3.167	9.722	52.555	11.636	26.031	30.22	10.034	7.144
5	3	89.983	13.924	38.3	54.501	33.464	28.508	36.997	87.562	37.662
	4	19.563	6.042	11.776	50.771	26.044	22.298	43.723	17.234	11.854

Table 4, it can be observed that each subject's distance from the reference subject is different, and all of them are non-zero values indicating that the model had a variety of data to train on. The difference could also be due to the difference in height and length of limbs.

5 CONCLUSION

In this paper, gym activities such as the push-up, squat, plank, forward lunge, and sit-up were used as the list of activities for recognition. OpenPose was used for extracting the skeleton points from the videos, and the extracted features were pre-processed and fed into classification algorithms, namely Decision Tree classifier, Support Vector Machine, K-Nearest Neighbor classifier, and Random Forest classifier. Metrics such as accuracy, balanced accuracy, macro precision score, macro recall score, and macro F1 score, were used for evaluating the attempted classification models, and it was concluded that the Random Forest classifier outperformed the other attempted classification models. A repetition counter was developed used for calculating the number of repetitions of each activity performed by the user. It was developed with the help of local minima analysis of the movement of nose skeleton point in the user. Also, the correctness of the workout performed by the user was calculated with the help of dynamic time warping. The detailed report was presented to the user in an android application with an interactive Graphical User Interface.

REFERENCES

[1] Pedro Acosta-Manzano, Víctor Segura-Jiménez, Irene Coll-Risco, Milkana Borges-Cosic, José Castro-Piñero, Manuel Delgado-Fernández, and Virginia A. Aparicio. 2019. Association of sedentary time and physical fitness with ideal cardiovascular health in perimenopausal women: The FLAMENCO project. *Maturitas* 120 (2019), 53 – 60.

[2] M. Babiker, O. O. Khalifa, K. K. Htike, A. Hassan, and M. Zaharadeen. 2017. Automated daily human activity recognition for video surveillance using neural network. In *2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*. IEEE, Putrajaya, Malaysia, 1–5.

[3] Ashwin Ramesh Babu, Mohammad Zakizadeh, James Robert Brady, Diane Calderon, and Fillia Makedon. 2019. An Intelligent Action Recognition System to assess Cognitive Behavior for Executive Function Disorder. In *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*.

IEEE, IEEE, Vancouver, BC, Canada, 164–169.
[4] N. Bakalos, A. Voulodimos, A. Doulamis, and N. Doulamis. 2019. A Deep-narma Filter for Unusual Behavior Detection from Visual, Thermal and Wireless Signals. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Brighton, United Kingdom, 4005–4009.
[5] S. Bian, V. F. Rey, P. Hevesi, and P. Lukowicz. 2019. Passive Capacitive based Approach for Full Body Gym Workout Recognition and Counting. In *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*. IEEE, Kyoto, Japan, 1–10.
[6] David Blumenthal and Shanoor Seervai. 2018. Rising Obesity in the United States Is a Public Health Crisis.
[7] Z. Cao, T. Simon, S. Wei, and Y. Sheikh. 2017. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Honolulu, HI, USA, 1302–1310.
[8] Maxime Devanne, Stefano Berretti, Pietro Pala, Hazem Wannous, Mohamed Daoudi, and Alberto Del Bimbo. 2017. Motion segment decomposition of RGB-D sequences for human behavior understanding. *Pattern Recognition* 61 (2017), 222 – 233. https://doi.org/10.1016/j.patcog.2016.07.041
[9] Alex Dillhoff, Konstantinos Tsiakas, Ashwin Ramesh Babu, Mohammad Zakizadehghariehali, Benjamin Buchanan, Morris Bell, Vassilis Athitsos, and Fillia Makedon. 2019. An Automated Assessment System for Embodied Cognition in Children: From Motion Data to Executive Functioning. In *Proceedings of the 6th International Workshop on Sensor-Based Activity Recognition and Interaction* (Rostock, Germany) *(iWOAR '19)*. Association for Computing Machinery, New York, NY, USA, Article 9, 6 pages. https://doi.org/10.1145/3361684.3361693
[10] M. Ehatisham-Ul-Haq, A. Javed, M. A. Azam, H. M. A. Malik, A. Irtaza, I. H. Lee, and M. T. Mahmood. 2019. Robust Human Activity Recognition Using Multimodal Feature-Level Fusion. *IEEE Access* 7 (2019), 60736–60751.
[11] Sofia Evaristo, Carla Moreira, Luís Lopes, André Oliveira, Sandra Abreu, César Agostinis-Sobrinho, José Oliveira-Santos, Susana Póvoas, Rute Santos, and Jorge Mota. 2019. Muscular fitness and cardiorespiratory fitness are associated with health-related quality of life: Results from labmed physical activity study. *Journal of Exercise Science & Fitness* 17, 2 (2019), 55 – 61.
[12] Yu-Jin Hong, Ig-Jae Kim, Sang Chul Ahn, and Hyoung-Gon Kim. 2010. Mobile health monitoring system based on activity recognition using accelerometer. *Simulation Modelling Practice and Theory* 18, 4 (2010), 446 – 455. Modeling and Simulation Techniques for Future Generation Communication Networks.
[13] Y. Hsu, S. Yang, H. Chang, and H. Lai. 2018. Human Daily and Sport Activity Recognition Using a Wearable Inertial Sensor Network. *IEEE Access* 6 (2018), 31715–31728.
[14] Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. In *Proceedings of the 22nd ACM International Conference on Multimedia* (Orlando, Florida, USA) *(MM '14)*. Association for Computing Machinery, New York, NY, USA, 675–678. https://doi.org/10.1145/2647868.2654889
[15] H. Jiang, T. Gonnot, W. Yi, and J. Saniie. 2017. Computer vision and text recognition for assisting visually impaired people using Android smartphone. In *2017 IEEE International Conference on Electro Information Technology (EIT)*. IEEE, Lincoln, NE, USA, 350–353.
[16] Katarina Katić, Rongling Li, Jacob Verhaart, and Wim Zeiler. 2018. Neural network based predictive control of personalized heating systems. *Energy and Buildings* 174 (2018), 199 – 213.
[17] Dimitrios I. Kosmopoulos, Nikolaos D. Doulamis, and Athanasios S. Voulodimos. 2012. Bayesian filter based behavior recognition in workflows allowing for user feedback. *Computer Vision and Image Understanding* 116, 3 (2012), 422 – 434. Special issue on Semantic Understanding of Human Behaviors in Image Sequences.
[18] Houssem Lahiani and Mahmoud Neji. 2018. Hand gesture recognition method based on HOG-LBP features for mobile devices. *Procedia Computer Science* 126 (2018), 254 – 263. Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 22nd International Conference, KES-2018, Belgrade, Serbia.
[19] Daniele Liciotti, Michele Bernardini, Luca Romeo, and Emanuele Frontoni. 2019. A sequential deep learning application for recognising human activities in smart homes. *Neurocomputing* (2019).
[20] G. Liu, G. Tian, J. Li, X. Zhu, and Z. Wang. 2018. Human Action Recognition Using a Distributed RGB-Depth Camera Network. *IEEE Sensors Journal* 18, 18 (Sep. 2018), 7570–7576.
[21] Keyu Meng, Shenlong Zhao, Yihao Zhou, Yufen Wu, Songlin Zhang, Qiang He, Xue Wang, Zhihao Zhou, Wenjing Fan, Xulong Tan, Jin Yang, and Jun Chen. 2020. A Wireless Textile-Based Sensor System for Self-Powered Personalized Health Care. *Matter* (2020).
[22] Henry Friday Nweke, Ying Wah Teh, Ghulam Mujtaba, and Mohammed Ali Al-garadi. 2019. Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions. *Information Fusion* 46 (2019), 147 – 170.
[23] Venet Osmani, Sasitharan Balasubramaniam, and Dmitri Botvich. 2008. Human activity recognition in pervasive health-care: Supporting efficient remote collaboration. *Journal of Network and Computer Applications* 31, 4 (2008), 628 – 655.
[24] Eleonora Papadimitriou, Anastasia Argyropoulou, Dimitrios I. Tselentis, and George Yannis. 2019. Analysis of driver behaviour through smartphone data: The case of mobile phone use while driving. *Safety Science* 119 (2019), 91 – 97.
[25] Okyza M Prabowo, Kusprasapta Mutijarsa, and Suhono Harso Supangkat. 2016. Missing data handling using machine learning for human activity recognition on mobile device. In *2016 International Conference on ICT For Smart Society (ICISS)*. IEEE, IEEE, Surabaya, Indonesia, 59–62.
[26] J. Qi, P. Yang, M. Hanneghan, S. Tang, and B. Zhou. 2019. A Hybrid Hierarchical Framework for Gym Physical Activity Recognition and Measurement Using Wearable Sensors. *IEEE Internet of Things Journal* 6, 2 (April 2019), 1384–1393.
[27] J. Qi, Y. Yang, X. Peng, L. Newcombe, A. Simpson, and P. Yang. 2019. Experimental Analysis of Artificial Neural Networks Performance for Physical Activity Recognition Using Belt and Wristband Devices. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, Berlin, Germany, 2492–2495.
[28] Suneth Ranasinghe, Fadi Al Machot, and Heinrich C Mayr. 2016. A review on applications of activity recognition systems with regard to performance and evaluation. *International Journal of Distributed Sensor Networks* 12, 8 (2016), 1550147716665520.
[29] Majdi Rawashdeh, Mohammed GH. Al Zamil, Samer Samarah, M. Shamim Hossain, and Ghulam Muhammad. 2017. A knowledge-driven approach for activity recognition in smart homes based on activity profiling. *Future Generation Computer Systems* (2017).
[30] Denzil Ferreira Anind K. Dey Brian Suffoletto Sangwon Bae, Tammy Chung. 2018. Mobile phone sensors and supervised machine learning to identify alcohol use events in young adults: Implications for just-in-time adaptive interventions. *Addictive Behaviors* 83 (2018), 42 – 47. Ambulatory Assessment of Addictive Disorders.
[31] Song-Mi Lee, Sang Min Yoon, and Heeryon Cho. 2017. Human activity recognition from accelerometer data using Convolutional Neural Network. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*. IEEE, Jeju, South Korea, 131–134.
[32] Dwi Sunaryono, Joko Siswantoro, and Radityo Anggoro. 2019. An android based course attendance system using face recognition. *Journal of King Saud University - Computer and Information Sciences* (2019).
[33] Muhammad Ehatisham ul Haq, Muhammad Awais Azam, Usman Naeem, Yasar Amin, and Jonathan Loo. 2018. Continuous authentication of smartphone users based on activity pattern recognition using passive mobile sensing. *Journal of Network and Computer Applications* 109 (2018), 24 – 35.
[34] L. Wei, W. Luo, J. Weng, Y. Zhong, X. Zhang, and Z. Yan. 2017. Machine Learning-Based Malicious Application Detection of Android. *IEEE Access* 5 (2017), 25591–25601.
[35] Yuexin Wu, Zhe Jia, Yue Ming, Juanjuan Sun, and Liujuan Cao. 2016. Human behavior recognition based on 3D features and hidden markov models. *Signal, Image and Video Processing* 10, 3 (2016), 495–502.
[36] Dewei Yi, Jinya Su, Cunjia Liu, Mohammed Quddus, and Wen-Hua Chen. 2019. A machine learning based personalized system for driving state recognition. *Transportation Research Part C: Emerging Technologies* 105 (2019), 241 – 261.
[37] Z. Yuan, Y. Lu, and Y. Xue. 2016. Droiddetector: android malware characterization and detection using deep learning. *Tsinghua Science and Technology* 21, 1 (Feb 2016), 114–123.
[38] Z. Zhang, N. Wang, and L. Cui. 2018. Fine-Fit: A Fine-grained Gym Exercises Recognition System. In *2018 24th Asia-Pacific Conference on Communications (APCC)*. IEEE, Ningbo, China, 492–497.