An Analysis of Human Activities Recognition using Smartwatches Dataset

Saadia Karim[1], SM Aqil Burney[2], Nadeem Mahmood[3]

College of Computer Science and Information System, IoBM, Karachi, Pakistan[1, 2]

Umaer Basha Institute of Information Technology, University of Karachi, Karachi, Pakistan[3]

Abstract—**Today, the era of smart devices evolving the human behavior interaction to a changing environment where the learning of activities is monitored to predict the next step of human behavior. The smart devices have these sensors built-in (accelerometer and gyroscope), which are continuously generating a large amount of data. The data used to identify the novel patterns of human behavior, together with machine learning and data mining techniques. Classification of human motions with motion sensor data is among the current topics of study. The classification is an important part of data mining techniques and used in this work to find the accuracy of instances in the given dataset. Thus, it is possible to follow the activities of a user carrying only a smartwatch. The smartwatches consisting of four different models from two manufacturers are used. Furthermore, the experiment contains nine users and seven activities performed by them. After the classification was determined, the data set to which the principal component analysis has been applied was classified by decision stump, j48, Bayes net, naive Bayes, naive Bayes multinomial text, random forest, and logit boost methods, and their performances were compared. The most successful result was obtained from the random forest method. The accuracy of the Random Forest classification algorithm on nominal datasets is 99.99% on both accelerometer and gyroscope sensors.**

Keywords—Human activity recognition; smartwatches; big data; machine learning; random forest

I. INTRODUCTION

In this technology era, various advanced applications are developing to support human daily activities. Among them, smartphones, and smartwatches are more popular to use. Smartwatches are wearable computer in the form of a wristwatch. Smartwatches are providing a touchscreen interface for daily use and become an integrated part of everyday life [1] [2]. They frequently use smartwatches producing huge amounts of data on daily basis. The growing era of technology, make smartwatches more power full device because it comprises the efficient computational power, internet facilities, and hardware sensors such as accelerometer [3][4][5], gyroscope [3][6], Face lock mechanism, touch with finger lock, and GPS receivers in it [7]. All these topographies inspire to study human activity recognition system and make the smartwatches to become a rich environment for many systems like:

- Blood pressure and temperature measuring in COVID19 situation [8].

- Heterogeneities Activity Recognition [9].

- Activity Recognition Using Active Learning [1].

- Classification of Various Daily Activities [10].

- Sport-Related Human Activity Detection and Recognition [11].

Smartwatches sensors become monitoring tools to recognize human physical activities such as Biking, Walking, Stair Up, Stairs down, Sitting, Standing, and doing nothing (null). Analyzing and learning of these activities will help to monitor human health and provide health services, and security services.

Serkan with other authors [3] describes the motion detection of human daily life, is an important task such as for healthcare, fitness, children's motion, and older person care. The smartwatches are easy to carry on hand. So, the activities of human daily life are observed in the form of data set through smartwatches. The eight different activities are collected for analysis to be performed. The analysis used the machine learning approach. The features are determined by using random forest, support vector machine, C4.5, and k-nearest neighbor methods. The most successful result was obtained from the random forest.

Hamid and Ali in [7] used smartwatches, which were fixed on the right human ankle-foot. The accelerometer and gyroscope sensor are used to classify the walk, jog, and run base on fast, normal, and slow motion. A threshold-based analysis using 20 different human activities are classified. The accuracy of 97.5% is achieved from the raw dataset of smartwatches used in the experiment. The authors in [9] used two different brands of smartwatches, which were fixed on the hand. The smartwatches contain different types of embedded sensors for example accelerometer, Global positioning system (GPS), gyroscope, Wi-Fi, and NFC. The output from these sensors has been used for the calculation of accurate sensing and impact on HAR by Sensor Biases (SB), Sensor Rate Heterogeneity (SRH), and Sampling Rate Instability (SRI). The clustering approach (low-pass filtering) has been used for finding the justification deficiencies in HAR systems. The 36 devices have been used for getting datasets and examination identifies the type heterogeneity has present in the datasets. The experiments show that the sensor biases have an 8% deviation in it, whereas the heterogeneity in data causes a huge challenge for monitoring the tasks. There is the time domain, feature domain and ECDF types of feature extraction has used with four types of learning approaches: C4.5, SVM, K-NN, and random forests. The author adopts the F1-score,

which is the harmonic mean of precision and recall, as the primary evaluation metric.

The focus of this article is to classify the given dataset on the performance evaluation of selected classifiers. The method used in this article are quite straightforward, but to generate results that are competitive with available articles. This article is organized as follows. Section 2 describes the related work already done for smartwatches. Section 3 then describes the process for collecting the smartwatches sensor data and transforming it into a form suitable for data mining. Section 4 then describes the performance evaluation based on accuracy and time using classifiers on smartwatches datasets with the results from applying those classifiers. The article concludes with Section 5, which summarizes the main conclusions, identifies areas for future work.

II. RELATED WORK

Akram and other authors in [12] illustrate that recognizing human physical activities are a significantly challenging research region for the researcher to work and many new applications are functioning as healthcare, smart cities, and home security. The new capable method is to track human activities by adding sensors to smartphones. The model is proficient in identifying multiple activities of humans in a real-time environment achieved dataset daily by triaxial accelerometer construct in smartphones. The experiment does not work on some constraints like fast walking, dancing. The design of the system works on the new digital low-pass filter with some classification methodology on high and low speed of data to make the system more accurate and faster. The selected classifiers are available in the Weka toolkit: Multilayer Perception, Random Forest, SVM, Simple Logistic, Logit Boost, and LMT. To improve the accuracy, proficiency, and toughness of classifiers, the fusion method has been used to combine different classifiers and found the average probability accurate rate of 91.15% [12].

Whereas, in [13], the authors express the Weka tool for data mining purposes. Data mining involves the following steps as data preprocessing, cleaning, transformation, reduction, feature selection, classification, and evaluation of data. The Weka tool is used to process all steps involves in data mining. The Weka tool accepts the ARFF format and CSV format. The dataset loads in Weka to be preprocessed and apply the classifications methods, which help predictions and apply algorithm techniques. Classification, clustering, and association rules are discussed to get interaction and give an overview of using the Weka tool [13].

According to [14], the authors describe the advanced method of monitoring human activities by an accelerometer, Global positioning system (GPS), gyroscope, radio frequency identification, and NFC available in smartphones using discrete data variables. The enabling of automatic sensors extend the time for ubiquitous computing. The data collected from devices are analyzed by some classification methods then perform monitoring by some algorithms like Bayesian decision (BDM), support vector machines (SVM), K-nearest neighborhood (KNN), etc. The usage of different devices and these devices has different consumption of batteries. This energy consumption is a major problem by smartphone IMU sensors to execute all tasks needed for activity classifications. The dynamic Ameva algorithm has been applied to monitor 32Hz sitting, standing, and lying to 50Hz of walking. Ameva classification system has a 98% average accuracy for recognizing different activities. To identify the accuracy in heterogeneous activities the Dynamic Sample Rate and Duty Cycle section approaches are used. The system analysis of smart activities by items is additional information to provide support to the system. The comparison of work has been performed with other classification approaches and concluded that all approaches give related results but the computational cost is applied [14]. As smartwatches are used to recognize the activities of humans. In [1], the author elaborates on the usage of smartwatches based on hardware and software sensors and active learning mechanisms. The hardware and software-based sensors using accelerometer, rotation vector, and linear acceleration sensors. These sensors produce data for 5 daily activities perform by users such are standing, sitting, laying down, walking, and none. These activities are classified by Random Forest, Extra Tree, Naive Bayes, Logistic Regression, and Support Vector Machine (SVM). Among them, the Extra Tree model is used as a baseline for the active learning model. Through the active learning model, 95% of accuracy is proved using 46% fewer samples [1]. Furthermore, in [10], the author proof 97.19% accuracy using a deep learning model through a Convolutional Neural Network (CNN). The classification uses 11 different activities (high or less use in daily life). This study shows that if accuracy measures are predicted accurately then the smart-home concept saves energy and this will help us to go the concept of IoT.

In addition to the field of smartwatches and human activity recognition, Weiss with other authors [15] added biometric efficiency to the devices. For evaluating the 18 different activities in the WISDM lab dataset. The biometric authentication in smartwatches and smartphones gives zero-effort feasible results and help to identify the subject and apply access to them. The IoT and energy saving using smartwatches same as smartwatches used for monitoring player health. The activities performed by the player have many variations and recognize them is a complex task.

The author in [11], explains the complication of monitoring player activities on the sliding window method as candidate motion and action changes rapidly. The activities related to sports are monitored based on the duration of each task to be done within an interval of time and these intervals of time are classified by the CNN technique. The CNN technique is based on the periodic and non-periodic duration of activities. This helps to identify the weak and strong candidates in the team as compared to the previous sliding window method. Same as in [8], Rozita Jamili with other authors explains the importance of IoT during the pandemic era of Corona Virus Disease (COVID-19). There are many sensors used in smart-home among them smartwatches are a reliable source of monitoring human activities. The smartwatches on the human hand enable us to monitor the health and take the required action as soon as possible. The method includes the MQTT, Web Socket, and HTTP programming to access and secure the devices. The data collected for the emergency of the patient gives 95% accuracy,

which helps to provide good health services from the health department [8].

III. METHODOLOGY

This section focused on step by step procedures perform on different activities using accelerometer and gyroscope data collected from smartwatches. First, we need to collect data then. Secondly, we describe the tool with which the analysis will perform. Then which type of classifiers are used to extract information from the data. Finally, we use the analysis to justify the accuracy of the classifier used in the experiment and draw the comparative analysis.

A. Data Collection

For data collection. The dataset is downloaded from the UCI machine learning repository [9] [16]. UCI machine learning repository is a gathering of databases that are used by the machine learning community for the experimental examination of machine learning procedures. The data from the accelerometer and gyroscope includes 10 attributes among them the motion of smartwatches is shown by x, y, and z axes. The dataset has nine different users as shown in Table I. The four smartwatches from two different brands (Samsung and LG as shown in Fig. 1 are used for recording the seven different activities of humans.

B. Weka Tool

Weka is a collection of machine learning algorithms and preprocessing virtual tools for different purposes of data mining, written in Java, and developed at the University of Waikato, New Zealand. Weka has multiple learning algorithms with various transformations on datasets. The datasets can be preprocessed, captured into a learning schema to analyze result-using classifiers without writing any code. Weka provides filters for preprocessing, classifiers, clusters, association, and visualization in 2D and a 3D interface that help in analyzing data more deeply and very helpful in prediction. For the classification of data, Weka 3.8 is used in the experiment [13], [19]–[21]. The data set used in this experiment is large. So, when uploading in Weka with default java memory setting than found the Java Virtual Heap size error which is solved by increasing the heap size, using the command prompt and change directory to Weka-3-8 and use the command: Weka-3-8-3> Java –Xmx5120m –jar weka.jar that increased the heap size up to 5GB, which make the loading of dataset easier.

C. Classifier

In this article, we used seven different built-in classifiers in Weka 3.8 to classify the seven activities (biking, walking, sitting, and standing, stair up, stair down, and null). The seven classifiers used are Decision Stump, J48, Bayes Net, Naive Bayes, Naive Bayes Multinomial Text, Random Forest, and Logit boost. The input data size is not fixed because Weka handles the huge amount of data to be processed and convert into. arff format but the machine Java virtual heap memory size can affect the loading of data in Weka. In this experiment, a huge amount of data, in which each file size is greater than 1GB of size having 6,746,393 instances. By increasing the Java Heap size, it becomes easier to execute the huge size of data in Weka.

The data is loaded in the preprocess tab available in Weka. The Preprocess screen describes the data related to the watch accelerometer and watches gyroscope files. Both files have different instances like the watch accelerometer has 3,540,962 instances and the watch gyroscope has 3,205,431 instances. After loading the. arff format in Weka. The screen shows the index attribute is selected by default and the dataset has been checked for having no missing values, no unique values. In Preprocess, a histogram shows how often an individual 10 selected values of classes like ground truth of nominal data set gt (nom) arises for the individual value of users are shown in Fig. 2.

TABLE I. INFORMATION ABOUT DATASET

Sensors	2	Accelerometer, Gyroscope
Smartwatch's	4	2 Samsung Galaxy Gears 2 LG watches
Users	9	a, b, c, d, e, f, g, h, i
Activities	7	Biking, Sitting, Standing, Walking, Stair up, Stair down, null
Attributes	10	Index, Arrival Time, Creation Time, X, Y, Z, User, Model, Device, Ground Truth

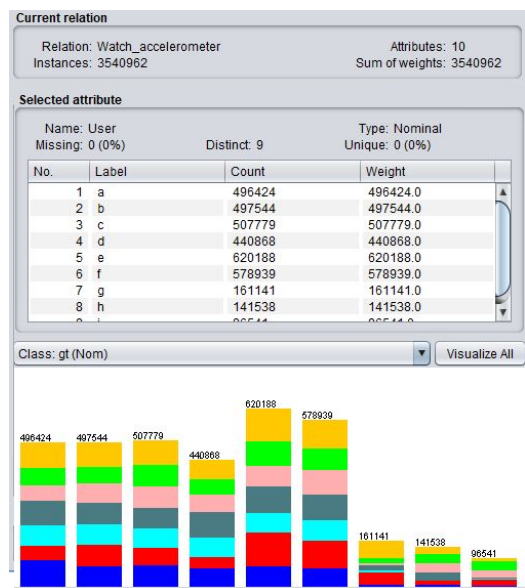Fig. 1. Samsung Galaxy Gear & LG Smartwatch [17] [18].

Fig. 2. Weka –Preprocess – Colors In Histogram Define the Users as "blue – Standing, Red – Null, Teal – Sitting, Steel – Walking, Light Pink – Stair Up, Green – Stair Down, Yellow – Bike".

Now onwards, after checking the dataset for no noisy data or missing values in it. The second step is to perform the classification of data. To apply classification on the dataset, the seven available classifiers in Weka such as Decision Stump, J48, Bayes Net, Naive Bayes, Naive Bayes Multinomial Text, Random Forest, and Logit boost are used. A decision stump is a decision tree, which uses only a single attribute for splitting.

Decision stump illustrates the learners with a high bias and low variance that's why often used as a weak learner. Decision stumps perform notably well on some commonly used benchmark datasets from the UCI repository [22][23].

The **J48 algorithm** is used to classify different applications and perform accurate results of the classification. J48 algorithm is one of the best machine learning algorithms to examine the data categorically and continuously [24].

A **Bayesian network** is a form of a directed graphical model for representing multivariate probability distributions [25][22].

Naïve Bayes is a simple learning algorithm that utilizes the Bayes rule together with a strong assumption that the attributes are conditionally independent, given the class [22].

Multinomial naive Bayes for text data. Operates directly (and only) on string attributes other. types of input attributes are accepted but ignored during training and classification [26].

Random Forests is an ensemble learning technique. It is a hybrid of the Bagging algorithm and the random subspace method and uses decision trees as the base classifier. Each tree is constructed from a bootstrap sample from the original dataset. The *"n"* is suggested to be $log_2 (N + 1)$, where *N* is the size of the whole feature set. Random forest classifier handles the missing values and maintains the missing data. Random forest handles the large dataset with higher dimensionality. In this research, the handling of the large dataset is the best fit with Random forest and the below results prove that also [3][2][22][23][27][28].

In machine learning and computational learning theory, **LogitBoost** is a boosting algorithm. The original paper casts the AdaBoost algorithm into a statistical framework. Specifically, if one considers AdaBoost as a generalized additive model and then applies the cost function of logistic regression, one can derive the LogitBoost algorithm [29].

The analysis is performed on both accelerometer and gyroscope datasets with the same seven classifiers. The Classify tab in Weka shows the Test Option, Result List, and Classifier output screen. In the Test Option, the percentage split option is chosen based on 66% training data and 34% testing data. Table II shows the splitting percentage of data.

TABLE II. DATA SPLITTING PERCENTAGE

Training Data (66%)		Testing Data (34%)	
Accelerometer	Gyroscope	Accelerometer	Gyroscope
2337035	2115584	1203927	1089847

IV. PERFORMANCE EVALUATION

The experiment is based on the performance of the following classifiers available in Weka software: Decision Stump, J48, Bayes Net, Naive Bayes, Naive Bayes Multinomial Text, Random Forest, and Logit boost. The classification is trained and tested using nine users and 10 attributes. The experiment is based on two parts: one is the time performance of the classifier and the other is the accuracy of the classifier.

A. Time Performance

The time performance of the classifier is shown in Table III on both the accelerometer and gyroscope dataset. In the time performance, the time is taken by the model to train itself, and onwards time is taken to test the model based on splitting the data as shown in Table II. Lesser the time is taken by the classifier justify that it doesn't mean that classification is done accurately. As in Table III, the Decision Stump shows that the time taken to test the model is 1.66s on the accelerometer and 1.13s on the gyroscope sensor. On the second, the J48 shows the time taken to test the model is 2.54s on the accelerometer and 1.72s on the gyroscope sensor. According to this result, the decision stump takes a fewer second to test the result but these results are accurate or not will be proof in Accuracy Rate shown in Table IV.

TABLE III. TIME PERFORMANCE

Time is taken to	build the model		test the model	
	Accelerometer	Gyroscope	Accelerometer	Gyroscope
Decision Stump	13.39s	12.94s	1.66s	1.13s
J48	247.41s	203.11s	2.54s	1.72s
Bayes Net	78.02s	88.25s	7.4s	10.36s
Naive Bayes	4.74s	4.63s	15.7s	13.4s
Naive Bayes Multinomial Text	0.23s	0.24s	5.08s	4.91s
Random Forest	1912.08s	1713.5s	39.52s	40.27s
Logit Boost	552.25s	498.24s	3.07s	2.24s

TABLE IV. ACCURACY SUMMARY OF THE CLASSIFIERS

Classifiers	Accelerometer	Gyroscope
Decision Stump	26.27%	27.13%
J48	99.98%	99.98%
Bayes Net	99.99%	99.98%
Naive Bayes	53.21%	48.58%
Naive Bayes Multinomial Text	17.90%	16.28%
Random Forest	99.99%	99.99%
Logit Boost	72.89%	74.39 %

B. Accuracy Rate

The classification accuracy is calculated as shown in equation (1). This value is the ratio of the correctly classified sample number to the total sample number.

$$\text{Accuracy (CA)} = \frac{TP + TN}{TP + FP + FN + TN} * 100\% \qquad (1)$$

Where TP is the number of true positives, TN defines the number of true negatives, FP is the number of false positives and FN is the number of false negatives [3].

The accuracy summary of all classifiers shown in Table IV. Among all classifiers the Random Forest, Bayes Net, and J48 perform were well. As in Table III, the time performance of Decision Stump and J48 is fast to execute but accuracy to classify the data is not efficient to be used for a huge amount of data. However, our experiment shows that the Random Forest took much time to build the model and test the model but the result shown are accurately classified. The Random Forest [14] [7] [4] verified the high accuracy rate of 99.99% on both datasets of accelerometer and gyroscope. Below Fig. 3 shows the graphical representation of the accuracy rate of the used classifiers.

In Tables V and VI, the confusion matrix is based on the Random Forest classifier. The matrix shows the accurately classified instance for accelerometer and gyroscope sensors. It gives an overview of all activities like standing, null, sitting, walking, stair up, stair down, and on a bike.

Fig. 3. Shows the accurate of the Classifiers.

TABLE V. CONFUSION MATRIX FOR ACCELEROMETER SENSOR

Accelerometer							
a	b	c	d	e	f	g	
153126	1	0	0	0	0	0	a
2	177357	2	3	2	0	6	b
0	5	144415	0	0	0	0	c
0	1	0	186628	0	0	0	d
0	5	0	0	160844	30	0	e
0	4	0	0	36	165873	0	f
0	2	0	0	0	0	215585	g

TABLE VI. CONFUSION MATRIX FOR GYROSCOPE SENSOR

Gyroscope							
a	b	c	d	e	f	g	
146222	4	0	0	0	0	0	a
2	159839	4	3	1	1	6	b
1	3	142790	0	0	0	0	c
0	6	0	166121	0	0	0	d
0	2	0	0	151856	25	0	e
0	0	0	0	44	145395	0	f
0	2	0	0	0	0	177520	g

*a = stand, b = null, c = sit, d = walk, e = stairs up, f = stairs down, g = on bike

V. CONCLUSION

In this study, the features obtained after applying classification to the data set generated by the accelerometer and gyroscope sensor. The smartwatches are classified using machine learning methods. In this study, where seven different daily human activities are classified, the most successful result is obtained from the random forest method. The Random Forest is easy to understand and handle the large dataset. The result shows that Random Forest classified 99.99% correct classification on both datasets using the Weka toolkit.

The limitation of this work is the hardware used and Weka toolkit package support on 6GB RAM and 5GB Java. The used hardware effects time performance of many classifiers also restrict some move to use because of processing took longer time. Furthermore, if the resources are updated then the time performance evaluation can become better and some classifiers can be used.

In the future, the updated resources and predictions from this analysis work in the human recognition system. The result of this study can help in the recognition of activities to save energy consumption. Applications of detection of human activities can support people in healthy life subject. It can help detect and prevent dangerous actions such as falling and disappearing of older people and young children, or actions which are not good for the health of a person. Furthermore, the dataset can be changed to unsupervised data and use clustering and association algorithms to perform prediction and making the human recognition support system better.
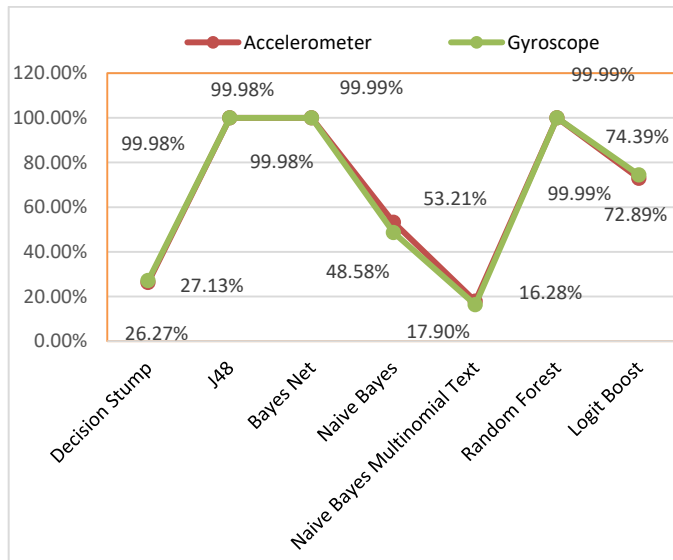
REFERENCES

[1] F. Shahmohammadi, A. Hosseini, C. E. King, and M. Sarrafzadeh, "Smartwatch Based Activity Recognition Using Active Learning," in Proceedings - 2017 IEEE 2nd International Conference on Connected Health: Applications, Systems, and Engineering Technologies, CHASE 2017, 2017, pp. 321–329.

[2] P. Fulfillment et al., "Developing Information and Communication Technology ' S (ICT) Acceptance Model in the Context of Social Environment and Measuring the Performance using Fuzzy Models," 2017.

[3] S. Balli, E. A. Sağbaş, and M. Peker, "Human activity recognition from smartwatch sensor data using a hybrid of principal component analysis and random forest algorithm," Meas. Control (United Kingdom), vol. 52, no. 1–2, pp. 37–45, 2019.

[4] D. D. Adrian Fernandez, "Accelerometer Sensor," 2013. [Online]. Available: https://www.sciencedirect.com/topics/engineering/accelerometer-sensor.

[5] 2019 Goleman et al., Chapter 4, Using the Accelerometer, vol. 53, no. 9. O'Reilly Media, Inc., 2019.

[6] M. Barela, Make: Getting Started with Circuit Playground Express. 2018.

[7] H. M. Ali and A. M. Muslim, "Human Activity Recognition Using Smartphone and Smartwatch," Int. J. Comput. Eng. Res. Trends, vol. 3, no. 10, pp. 2349–7084, 2016.

[8] R. J. Oskouei, Z. MousaviLou, Z. Bakhtiari, and K. B. Jalbani, "IoT-Based Healthcare Support System for Alzheimer's Patients," Wirel. Commun. Mob. Comput., vol. 2020, pp. 1–15, Oct. 2020.

[9] A. Stisen et al., "Smart devices are different: Assessing and mitigating mobile sensing heterogeneities for activity recognition," SenSys 2015 - Proc. 13th ACM Conf. Embed. Networked Sens. Syst., pp. 127–140, 2015.

[10] M. C. Kwon, H. You, J. Kim, and S. Choi, "Classification of Various Daily Activities using Convolution Neural Network and Smartwatch," in Proceedings - 2018 IEEE International Conference on Big Data, Big Data 2018, 2019, pp. 4948–4951.

[11] Z. Zhuang and Y. Xue, "Sport-related human activity detection and recognition using a smartwatch," Sensors (Switzerland), vol. 19, no. 22, pp. 1–21, 2019.

[12] A. Bayat, M. Pomplun, and D. A. Tran, "A study on human activity recognition using accelerometer data from smartphones," Procedia Comput. Sci., vol. 34, no. December, pp. 450–457, 2014.

[13] S. Srivastava, "Weka: A Tool for Data preprocessing, Classification, Ensemble, Clustering, and Association Rule Mining," Int. J. Comput. Appl., vol. 88, no. 10, pp. 26–29, 2014.

[14] L. M. S. Morillo, L. Gonzalez-Abril, J. A. O. Ramirez, and M. A. A. De La Concepcion, "Low energy physical activity recognition system on smartphones," Sensors (Switzerland), vol. 15, no. 3, pp. 5163–5196, 2015.

[15] G. M. Weiss, K. Yoneda, and T. Hayajneh, "Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living," IEEE Access, vol. 7, pp. 133190–133202, 2019.

[16] G. M. Weiss, "WISDM Smartphone and Smartwatch Activity and Biometrics Dataset," UCI Mach. Learn. Repos. WISDM Smartphone Smartwatch Act. Biometrics Dataset Data Set, vol. 7, pp. 133190–133202, 2019.

[17] Samsung, "Samsung Galaxy Gear," Samsung. [Online]. Available: https://www.samsung.com/in/support/mobile-devices/is-samsung-galaxy-gear-waterproof/. [Accessed: 18-Dec-2020].

[18] LG, "LG Smartwatch." [Online]. Available: https://www.lg.com/us/smart-watches/lg-W100-lg-watch.

[19] E. Frank, M. A. Hall, and I. H. Witten, "The WEKA workbench," Data Min., pp. 553–571, 2017.

[20] "Weka 3: Data Mining Software in Java." [Online]. Available: https://www.cs.waikato.ac.nz/~ml/weka/.

[21] C. J. P. Ian H. Witten, Eibe Frank, Mark A. Hall, Data Mining Practical Machine Learning Tools and Techniques, 4th ed.

[22] K. M. Ting, Precision and Recall BT - Encyclopedia of Machine Learning. 2010.

[23] P. Morgan, Introduction to Data Science: Essential Concepts, vol. 53, no. 9. AI Sciences LLC, 2013.

[24] N. S. anaN and V. G. thri, "Performance and Classification Evaluation of J48 Algorithm and Kendall's Based J48 Algorithm (KNJ48)," Int. J. Comput. Trends Technol., vol. 59, no. 2, pp. 73–80, 2018.

[25] S. M. Aqil Burney and J. Naseem, "Decision Making in Uncertainty : A Bayesian Network for Plant Disease Diagnoses," Int. J. Comput. Sci. Inf. Secur., no. April 2018.

[26] Weka, "Class NaiveBayesMultinomialText." [Online]. Available: https://weka.sourceforge.io/doc.dev/weka/classifiers/bayes/NaiveBayesMultinomialText.html. [Accessed: 28-Nov-2020].

[27] L. Breiman, "Random Forests," 2001.

[28] Scott Hartshorn, Machine Learning With Random Forests And Decision Trees A Visual Guide For Beginners. 1393.

[29] J. Friedman, T. Hastie, and R. Tibshirani, "Additive Logistic Regression," The Annals of Statistics, vol. 28, no. 2. pp. 337–374, 2000.