Multimedia Tools and Applications (2024) 83:73607–73625
https://doi.org/10.1007/s11042-024-19285-7

1232: HUMAN-CENTRIC MULTIMEDIA ANALYSIS

PIF dataset: a comprehensive dataset of physiological and inertial features for recognition of human activities

Manpreet Kaur Dhaliwal[1] · **Rohini Sharma**[1] · **Rajbinder Kaur**[1]

Received: 13 June 2023 / Revised: 23 January 2024 / Accepted: 22 April 2024 /
Published online: 2 May 2024

Abstract
Activities and falls monitoring systems using wearable technology have a promising future. The publicly available datasets are based on a few inertial features only acquired with an accelerometer, gyroscope, smartphone or smart Watches. The activities and falls performed are also less. In this study, a dataset is created by collecting physiological features along with inertial features which will help in developing and validating systems studying the effect of physiological features on the detection and prediction of falls and activities. The dataset consists of 7 activities and 8 falls for inertial data; 2 activities for ECG data; 6 activities for EMG data and 6 activities for GSR data. Basic body parameters like height, weight, etc. along with beats per minute, SpO2 and blood pressure are also recorded for 12 subjects. The collected data is analyzed statistically using a boxplot, pair plot, correlation heatmap and p value. The activities are classified using SVM, KNN, RF and DT. For GSR, more than 90% accuracy is achieved and for EMG, the accuracy is less than 80%. For IMU data, more than 95% accuracy is achieved. The results encourage combining inertial, physiological and basic body parameters to detect and predict falls and activities.

Keywords ECG · EMG · GSR · IMU · Machine learning · Statistics · Falls · Activities

1 Introduction

Global data volume presently generated by the healthcare sector is about 30%. The annualized average growth rate of healthcare data will be 36% by 2025. This is 6% swifter than the industrial sector, 10% swifter than the financial sector, and 11% swifter than the media & entertainment sector [1]. In 2010, per subject gadget interaction is 298 per day, by 2025 it will reach 5000 gadget interactions per day and the major contributor of this data is the

✉ Rohini Sharma
rohini@pu.ac.in

Manpreet Kaur Dhaliwal
preet2016@pu.ac.in

Rajbinder Kaur
raj1995@pu.ac.in

[1] Department of Computer Science and Applications, Panjab University, Chandigarh, India

healthcare sector [2]. The accuracy of machine learning techniques is dependent on the availability of large amounts of data. With the advancement of technology and availability of data, the world is moving towards automation leading to smart applications in healthcare. Smart healthcare applications rely on real-time datasets for validation and wearable technology has made it possible to collect real-time health-related data. Seshadri et al. [3] provide in-depth knowledge about the various wearables available in the market. Wearable technology, features and the activities are useful to monitor various diseases. Wu et al. [4] projected a 7-day early prediction of Acute Exacerbations of Chronic Obstructive Pulmonary Disease (AECOPD) on 67 subjects using wearable and sensor devices. According to Wong et al. [5] use of wearables assists us in the early prediction of covid illness. In [14], EMG and gyro sensors are used for the detection of freezing of gait in Parkinson's disease. Trembling and shuffling muscle activities help in the assessment of falls for the affected person. In [15], EMG data collected through Myo Armband is used to classify hand-posed activities including stretching, wrist, waving, relaxed, and waving out to control wheelchair movement. The information related to prediction of falls and various types of activities for elderly people living alone can be shared with the caretakers that can ensure their well being [16–18, 50]. Ichwana et al. [19] monitored Transient Ischemic Attack (TTA) by classifying ten various activities pertaining to sitting, standing, walking and falling using an MPU6050 sensor affixed to their waists. For stress and anxiety disorder and emotion prediction, Castro-Garcia et al. [20] used EEG, ECG, breathing rate (BR), electrodermal activity and Skin temperature, Deger et al.'s study [21] used GSR signals, Luz Santamaria-Granados et al. [22] used ECG and GSR features and Chueh et al. [25] used skin temperature variation, ECG and GSR to classify emotions using various machine learning techniques.

Synthetic datasets have been used for the prediction of diseases [6–10]. [11–13]. used real datasets but these datasets have limited features. Moreover, for detecting falls and activities of daily living, the datasets available are based on inertial features. This study is a move toward the development of a dataset that is based on physiological and inertial features. We used a prototype using various sensors that can collect the physiological and motion features of a person. As the data are collected in a real environment it can assist researchers in the development and validation of an efficient Machine learning system that can help to detect and predict the condition.

This study is divided into 5 sections. Section 2 discusses the existing datasets. Section 3 covers the details of PIF dataset. Section 4 deals with the data analysis that further consists of two parts i.e., statistical and machine learning-based analysis of the dataset. Lastly, the study is concluded in Section 5.

2 Existing datasets

Kausik Sen et al. [23] conducted a study using GSR and EMG sensors for pain assessment on BioVid heat pain database. After statistical features extraction various machine learning algorithms are implemented to estimate the level of pain. From the EMG signal, features like Root Mean square are extracted. This feature has also been considered in this study. Fu et al. [24] used EMG with ECG features to detect fatigue in drivers and Kolmogorov–Smirnov Z test is implemented for statistical analysis. Two states such as fatigue and normal are classified with an accuracy of 86.67%. Chueh et al. [25] analyzed the data related to ECG, GSR and temperature of a person using MANOVA and six machine learning methods to predict emotions. The author concluded that MANOVA

and Logistic regression significantly improves the performance of the system. Kim et al. [26] considered physiological features i.e., ECG, EMG Skin conductance and respiration. After the extraction of features using pseudoinverse LDA (pLDA) is used to classify four musical emotions. Soleymani et al. [27] used the ANOVA test to analyze the emotions during movie scenes using physiological signals captured from ECG, EMG, temperature, respiration and GSR. The author concluded that MANOVA and Logistic regression significantly improves the performance of the system. Nadeem et al. [28] acquired the data for three activities and one fall using a Shimmer inertial sensor on the waist of 114 participants but in the dataset, the falls are performed by only four participants. ECG data are collected in parallel using a Shimmer ECG Sensor attached to the chest of a person to record 10 to 15 min ECG signals. A total of 39 records are collected of which 33 records contain twelve attributes and six contain 8 attributes. The PTB-XL dataset [29] is recorded from 18,885 participants. The Schiller AG device is used to acquire ECG signals from participants. Nebojša Malešević et al. [30] collected EMG signals from the right-hand using Force sensors from 20 subjects performing 65 hand movements. These movements are collected in an isometric way. Ozdemir et al. [31] used a BIOPAC MP36 device to acquire EMG signals of the forearm from forty participants performing ten hand gestures. In the CASE dataset [32], data related to physiological features such as EMG, GSR, ECG, Temperature, BVP and respiration are collected in parallel from thirty young subjects watching videos. Ojetola et al. [33] collected a dataset using seven activities and six falls using a Shimmer sensor placed on the chest and thigh of 42 subjects. This dataset has also data from only young healthy persons. DEAP dataset [34] is collected from thirty-two subjects by playing 40 music videos and rating the videos in four categories to acquire physiological signals such as GSR, EEG, Respiration Amplitude, Blood Volume, EMG, Skin Temperature, and EOG.

SisFall Dataset [35] consists of 19 activities and 15 falls collected from 38 participants using two accelerometers and one gyroscope attached to the waist of the participants. During data collection, few activities and falls are not performed by the elderly. Only one elderly who has expertise in judo performed all falls and ADLs. KU-HAR [36] dataset is only related to ADLs data collected from 90 subjects using Smartphones. This dataset was collected using only young participants aged between 18 and 34. This dataset has also data from only young healthy persons. The Real-World Falls dataset [37] is collected while performing falls by 100 subjects using a tri-axial accelerometer. This dataset faced the problem of class imbalance. AMIGOS dataset [38] is collected using Shimmer Sensor from forty subjects watching 16 short videos and 4 long videos. The author selected GSR, EEG and ECG features including Audio, Visual and Depth. Phinyomark et al. [39] used EMG signals to classify the gestures of six hands movements. Thirty-seven features are extracted in this study. The authors concluded that frequency domain features are not good for EMG signals. Falih et al. [15] used an EMG signal to control the wheelchair movement using the forearm motion to detect the muscle signals using the Myo Armband device. In this study, features are extracted based on the time domain and classified the movement using the Naïve Bayes technique in Weka. A total of five poses are correctly classified with 93.5%.

Researchers have focused on features based on single sensor or in certain circumstances, dual sensors like GSR and EMG. In PIF dataset, physiological and inertial features are collected for different types of activities and fall. The basic parameters of the human body are also available in the dataset. The motivation is that in the real world, holistic information about patients helps in detecting a condition with more accuracy.

3 PIF dataset

The dataset collected in this study is a comprehensive dataset that is based on **P**hysiological as well as **I**nertial **F**eatures (PIF). Table 1 gives the details of the dataset including source location, subject area etc.

PIF dataset is useful for the following:

- To develop and validate algorithms for the classification of various events.
- To compare existing datasets with existing algorithms.
- To support further studies in bioinformatics, bioengineering, health informatics etc.
- To develop systems for detection, prediction and health monitoring for the elderly, differently-abled persons or persons with major health problems like heart attack, emotional instability, and mental issues.
- To detect anomalies in different events that include heart rate, blood pressure spike, abnormal values of sdnn and rmssd, anomalies in accelerometer and gyroscope values etc.

3.1 Design, materials, and methods

In this section, several aspects of the data-acquiring process are discussed, that includes sensors used for data collection, types of data collected, storage, subjects, size of data samples, etc.

3.1.1 Data collection

Data acquisition process is shown in Fig. 1 and extracted features from various sensors are shown in Table 2. Figure 1 indicates the placement of sensors on the body during the data collection process, the flow of data, storage of data in .csv file with names and the flow of methodology used for this study.

Table 1 Specification table of PIF dataset

Specification table	
Subject area	Computer science in healthcare
More specific subject area	Physiological features of electromyography, electrocardiography, skin response and inertial measurements, GOQII smartwatch
How data was acquired	Using wearable Sensors MPU6050 Sensor, AD8232 Sensor, AD8226 Sensor, GSR Sensor, GOQII smartwatch
Data format	Raw and analyzed
Experimental factors	Measure of the dynamic behavior of subjects during different activities
Experimental features	12 healthy subjects while wearing wearable sensors.
Data source location	https://data.mendeley.com/datasets/phb9y6cp5c/1
Data accessibility	Dataset will be made available on different open platforms.
Related research article	Improving detection of falls and activities using machine learning model (Accepted, not published yet)

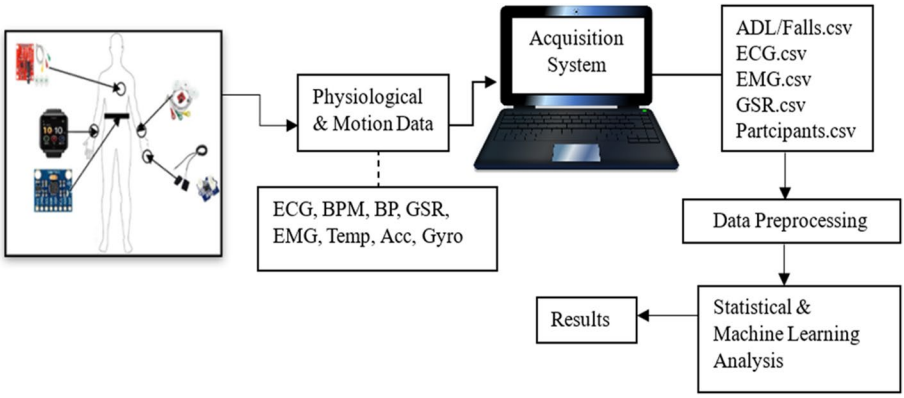

Fig. 1 The schematic diagram of the system for data acquisition and analysis

Table 2 List of sensors with extracted features

Sensors	Extracted features
ECG	PID, BPM, ECG, SpO_2, blood pressure, label
EMG	EMG, RMS, MAV, Label
GSR	Raw value, resistance, GSR value, SCL, label
IMU	TEMP, AccX, AccY, AccZ, GyRoX, GyRoY, GyRoZ, AccANGX, AccANGY, GyRoANGX, GyRoANGY, GyRoANGZ, GetANGX, GetANGY, GetANGZ, label

In Table 2 PID represents the unique identification number for the person, BPM means beats per minute, ECG represents the raw ecg value from the sensor, blood pressure represents the systolic and diastolic values and Label divides the data into various classes. EMG represents raw muscle value. RMS stands for Root Mean Square; MAV stands for Mean Absolute Value. SCL stands for Skin Conductance Level. Features in IMU sensor represent GyRoX, GyRoY, GyRoZ represent 3-axis gyroscope values, AccX, AccY, AccZ represent 3-axis accelerometer values, AccANG X, AccANG Y represent 3-axis accelerometer angle values, and GetANG X, GetANG Y, GetANG Z represent 3-axis combined angles of accelerometer and gyroscope.

Subjects For the study, we have taken twelve participants, all are of Indian descent and free from physical or mental disorders. Each participant is given detailed information about the experiment and the entire method, and the goal of data collection and utilization is clearly defined. The Declaration of Helsinki was followed in all of the trials. Informed consent is provided by all the subjects and anonymous data is collected.

Basic health information Basic health information is often used to assess overall health and wellness, and it can help doctors and other healthcare professionals provide you with the care and treatment you need. Basic information of participants is recorded manually that includes the information give in Table 3.

Table 3 List of basic health information of subjects

Pid	Weight in(kg)	Height	Age	Gender	Medical History	Country	Humidity	Smoking	Drinking	Diabetes	Heart Disease	Obese	Thyroid

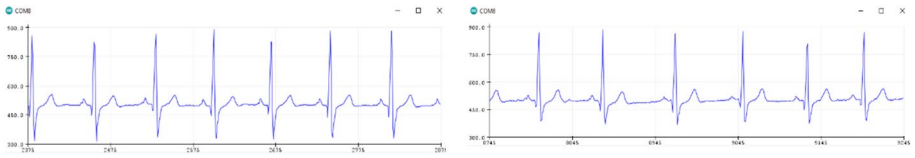

Fig. 2 ECG plot of a subject in sitting and deep breath while sleeping

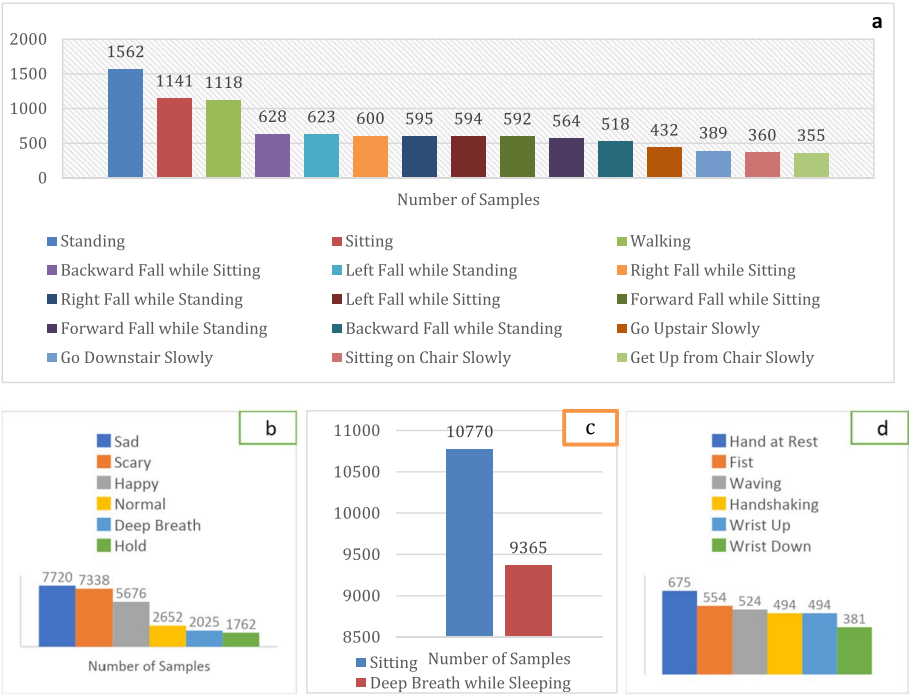

Fig. 3 **a-d** Number of samples in each class

In this article, the physiological features of subjects are collected using AD8232, AD8226, GSR, MPU6050, sensors and GOQII smartwatch. Details of the sensors used are given below:

- **ECG Sensor**: An electrocardiogram (ECG) sensor known as the ECG8232 is used to measure and record the electrical activity of the heart. This sensor is often used to find irregular heartbeats or other heart-related disorders using wearable technology and medical monitoring systems. A microcontroller Arduino 2560 receives the recorded ECG data from the ECG8232 sensor, where it is analyzed and interpreted. Figure 2 shows the raw ECG value of a subject. ECG data is collected using the AD8232 sensor and data was saved in (.csv) file format. On each subject electrodes are placed on Left Arm, Right Arm and Right Leg. We have recorded the values of ECG in sitting and sleeping mode for 5 min and the size of samples in each category is shown in Fig. 3c.

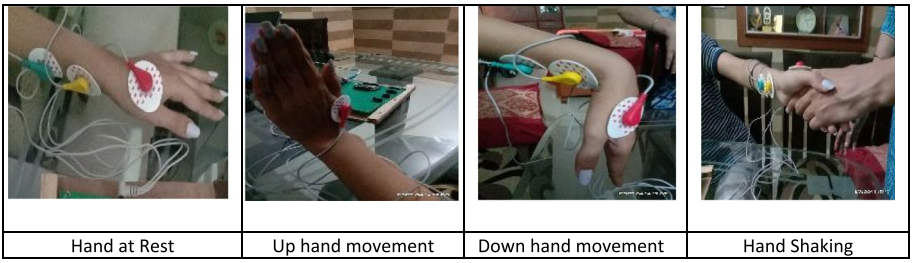

Fig. 4 Subject performing hand movements using EMG sensor for data collection

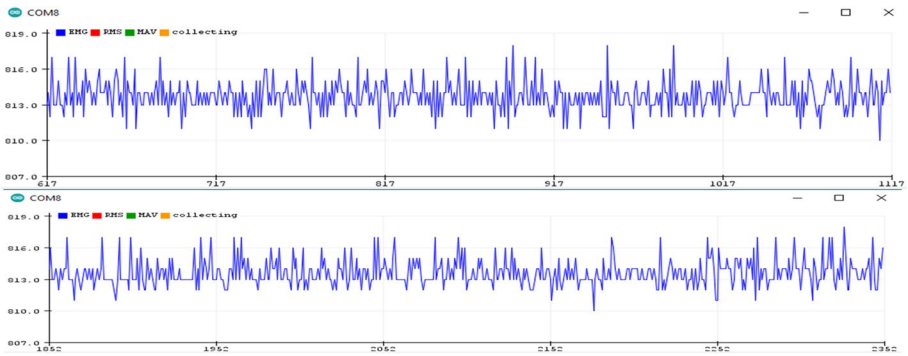

Fig. 5 EMG plot of subject performing hand at rest and fist

- **EMG Sensor**: An electromyography (EMG) sensor known as the EMG8226 is used to assess the health of muscles and nerve cells. These cells provide electrical signals that trigger the contraction or relaxation of muscles. These signals are read using the EMG sensor, which is a small form factor, multichannel, low-cost, and low-power amplifier. The 3-lead differential EMG sensor utilized in this study is shown in Fig. 4. Figure 5 shows the Raw EMG values of the subject during the data collection process. EMG electrodes can be attached to the hand and wrist using the onboard 3.5 mm cable connection and 30s data is recorded in a .csv file. All the movements are explained to subjects before performing. In this study, 6 movements have been considered, hand at rest, fist, up, down, handshaking and waving so that multi-classification can be possible. The number of samples in each class is shown in Fig. 3d.
- **GSR Sensor**: The Galvanic Skin Response (GSR) sensor is a biosensor that measures the electrical conductance of the skin. It is also known as a skin conductance or skin resistance sensor. It is generally used in psychological research studies to monitor changes in a person's autonomic nervous system that includes emotional state, stress levels, and physiological arousal. When a person experiences emotional or physical changes, the skin's electrical conductance varies, and the GSR sensor can pick up these changes. In this study, a sensor is placed in the first two fingers of a subject as shown in Fig. 6. Figure 7 gives the plotted values of GSR sensor in scary and normal moods. GSR sensor data is collected in 6 different moods; the number of samples in each class is shown in Fig. 3b. To capture values for scary, sad

Fig. 6 Recording emotion data of subject using GSR sensor

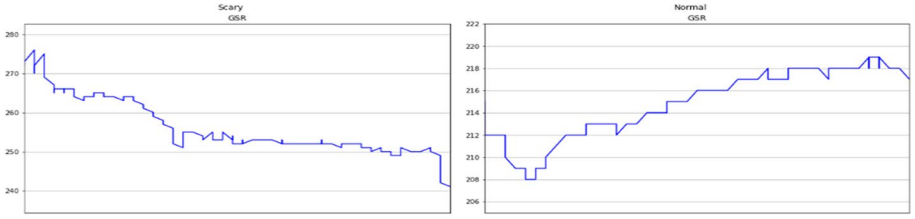

Fig. 7 GSR plot of subject in scary and normal mood

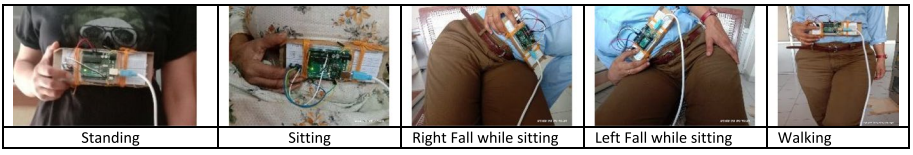

Fig. 8 Subject performing activities for data collection

and happy moods, a horror clip, a sad clip and a happy clip of 1 min each are shown to the subjects. In deep breathing the subject is breathing deeply and in hold breath, the subject is holding the breath.

- **IMU Sensor**: An MPU6050 Sensor is an Inertial Measurement Unit consisting of 9° of freedom device that includes a 3-axis Accelerometer, 3- axis Gyroscope and Temperature sensors. It operates on 3.3 V to 5 V DC. The different ranges of the Gyroscope are ± 250, 500, 1000, and 2000 °/s and Accelerometer is $\pm 2 \pm 4 \pm 8 \pm 16$ g. It is used to measure the acceleration, angular velocity, orientation, rotation and temperature of the body. MPU6050-based prototype is mounted on the waist of the subjects as shown in Fig. 8. It describes the features extracted from sensors using the acquisition system and stores in various .csv files for each subject. A total of 12 subjects with different age, weight and height profiles are selected for performing 7 activities and 8 falls and two trials are performed per subject. The features are extracted using tockn arduino library. The number of samples collected in each motion is shown in Fig. 3a. We have collected 1-minute data for activities and 20s data for falls. The plotted graph of 2 activities and 2 falls is shown in Fig. 9.

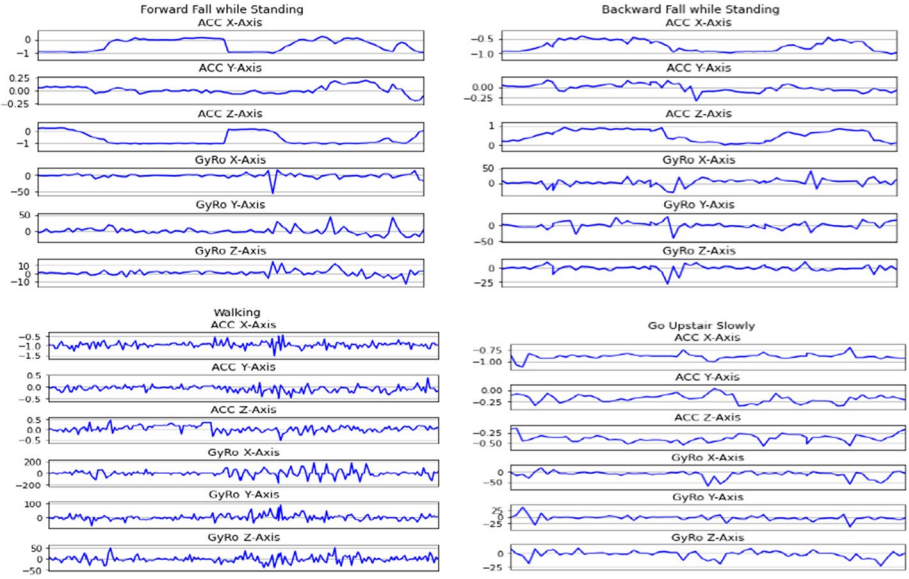

Fig. 9 Plotted graphs of two activities and two falls using IMU sensor

- **GOQII Smartwatch**: It is used to collect beats per minute, SpO_2 and blood pressure of the subject and data is recorded after every 10 s and stored in a .csv file.

4 Dataset analysis

4.1 Statistical analysis

Statistical analysis of the dataset is essential because it helps us to understand the data better. Descriptive statistics and graphs provide a general overview of the dataset that makes it simple to grasp. Graphs also summarize data to make it more intelligible by highlighting trends, patterns, regression, and correlations, among other measures. We have used Python as a statistical analysis language. In statistics, box plot is used to show the dispersion of the dataset, with the center line indicating the median value of the dataset. Box plot in Fig. 10 shows the data pertaining to GSR, EMG, ECG and AccX values. For the EMG plot, data is mostly negatively skewed dispersed as compared to GSR where data is less dispersed. For scary and sad classes, data is positively skewed and for hold and happy classes, data is normally distributed. For Falls, data is more dispersed and for activities it is very less dispersed. For ECG values data is symmetric and equally dispersed.

Figure 11 represents the pair plot to identify patterns and relationships between multiple variables. We have plotted the pair graph for Accelerometer and gyroscope values according to ADLs and Falls. The relationship and difference in variable values are displayed by the scatter plot in the paired graphs. It can be observed from the pair plot that features AccZ shows a symmetrical relation with other features. In contrast, other features are dispersed so it makes it easy to classify into the classes.

Figure 12 shows that there is a positive relationship between MAV and EMG but a negative relationship between for the RMS feature. GSR is positively correlated with

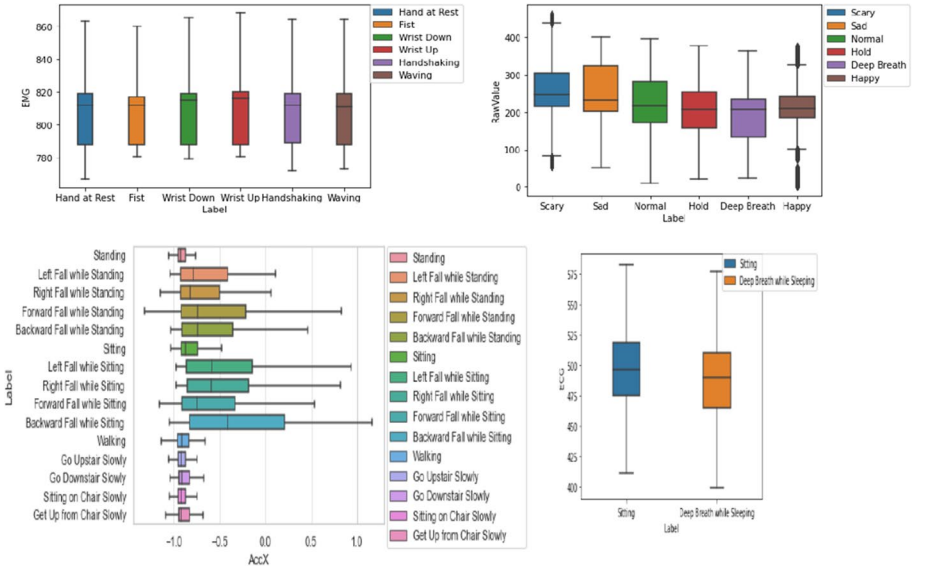

Fig. 10 Boxplots for ECG, EMG, IMU and GSR

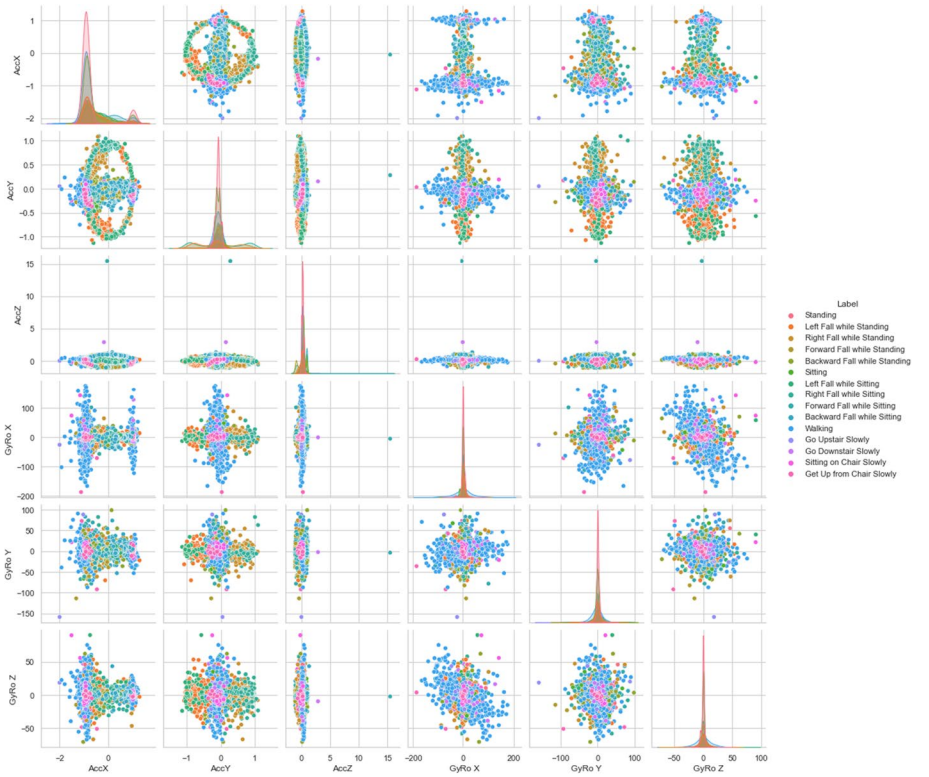

Fig. 11 Pair Plots for features captured using IMU for ADLs and fall

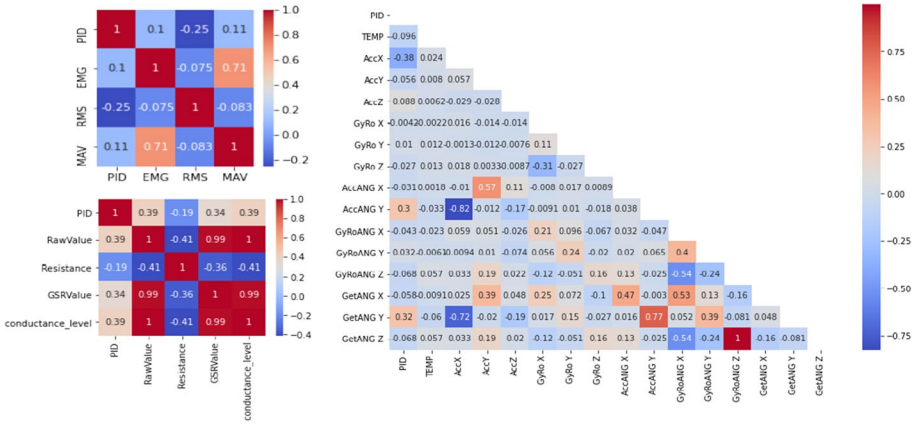

Fig. 12 Correlation heatmap of features

conductance level. AccAngY is positively correlated with GetANGY and negatively with AccX.

T-test and Anova test have also been performed to see how varied and similar the labels on the various datasets are. T test is a parametric test used for hypothesis testing to check whether two groups are different or not. It compares the mean of two different groups. ANOVA test stands for Analysis of Variance. It is used to examine the variance among two groups. The test on GSR values is illustrated in Fig. 13, where we have taken the raw values of the GSR and compared them to various scenarios, such as scared and happy. We have established the null hypothesis that there is no difference between scary and hold breath and performed t-test and anova test. P value for various GSR classes are shown in Table 4. As P values are smaller than 0.05, we reject the null hypothesis that there is no difference between Scary and hold breath etc.

We have also performed t-test on ECG values as shown in Fig. 14. It receives the Pvalue 0.0012 which is less than the significance level 0.05. So, we can conclude that sitting and deep breath while sleeping are significantly different.

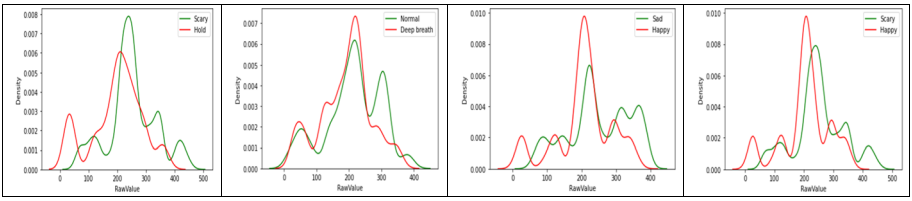

Fig. 13 Represents the distplot of GSR feature according to labels

Table 4 *P* values for various GSR classes

GSR classes	*P* value
Scary and hold breath	1.02E-111
Normal and deep breath	5.83E-19
Sad and happy	3.78E-159
Scary and happy	2.53E-184

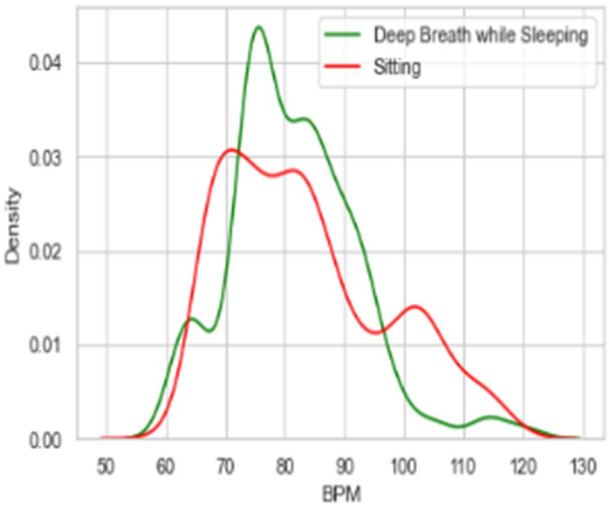

Fig. 14 Distplot for ECG values

Figures 15 and 16 shows the Distplot for IMU classes and EMG classes respectively. Tables 5 and 6 show the P values for IMU and EMG classes. It is concluded from both figures and tables that there are significant differences among class values for IMU and EMG.

4.2 Analysis of PIF dataset using machine learning

Statistical analysis was performed to check the similarities and differences in the class values of the dataset. Based on Distplot and significant values it is concluded that classes are different. In this section, we are classifying the same classes using Machine learning techniques: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF) and Decision Tree (DT). Apart from KNN, RF, SVM and DT, Logistic Regression, Linear Discriminant Analysis, and Gaussian Naïve Bayes were also implemented and based on the

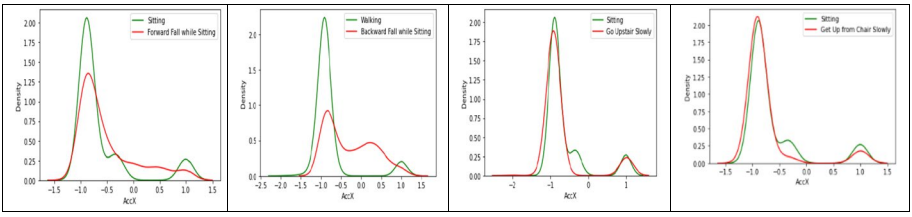

Fig. 15 DistPlot for IMU classes

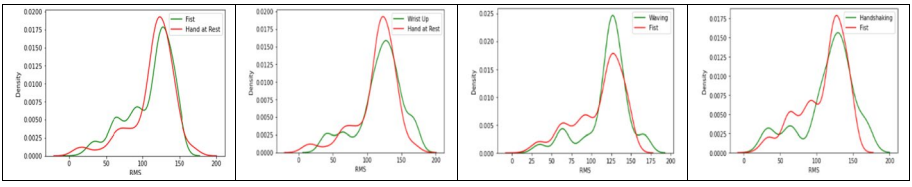

Fig. 16 DistPlot for EMG values

Table 5 *P* value for IMU classes

IMU	*P* value
Sitting and forward fall while sitting	1.12E-05
Walking and backward fall while sitting	1.08E-65
Sitting and get up from chair slowly	0.0004
Sitting and go upstair slowly	0.0188

Table 6 *P* value for EMG classes

EMG	*P* value
Hand at rest and fist	0.03983778
Wrist up and hand at rest	0.001819062
Waving and fist	2.70E-07
Handshaking and fist	0.003150364

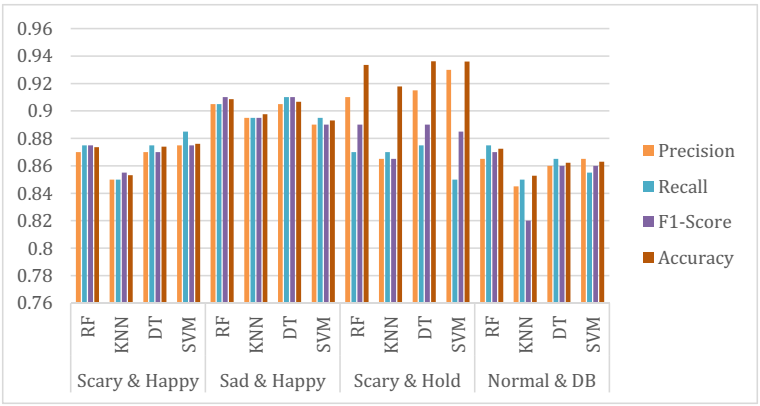

Fig. 17 Classification results of various classifiers according to GSR classes

results top four have been included in this study. Results are shown in Figs. 17, 18 and 19. This is based on the raw values of GSR, EMG, ECG and AccX values.

SVM: Support Vector Machine is a type of supervised machine learning algorithm that is used both for classification and regression. This algorithm is used to find hyperplane that differentiate the classes with maximum margin. Main advantage of using SVM is resulting classifier uses less numbers of training points (support vector) to classify new points.

DT: A decision tree is a supervised machine learning algorithm which is a non-parametric algorithm widely used for classification and prediction. In this algorithm the decisions are represented in the form of tree and flowchart like structure using divide and conquer method. The output is in the form of leaf nodes.

Random forest: It is also known as Ensemble algorithm which is a supervised learning algorithm. It solves the classification problem with a bagging mechanism and also solves regression problems. It reduces the problem of overfitting which makes

Fig. 18 Classification results of various classifiers according to EMG classes

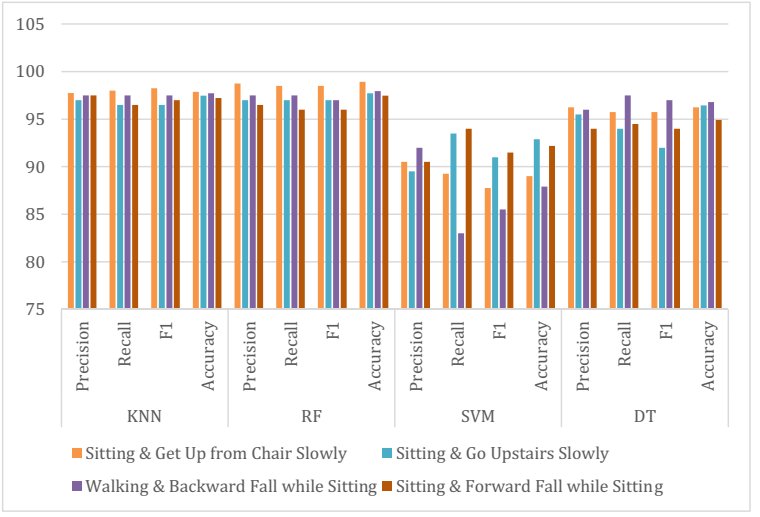

Fig. 19 Classification results of various classifiers according to IMU classes

it better from the decision tree. The best solution (on the basis of voting technique performed for every predicted result) is selected from every decision tree created on dataset samples. It performs on large datasets very fast and provides accurate result.

KNN: K-Nearest Neighbors algorithm is a non-parametric method, used to solve both classification and regression problems. The algorithm provides a solution to a problem that depends on identifying similar objects. Euclidean distance, Mahalanobis distance or Hamming distance are applied to find the new data points. The main advantages of this algorithm are simple, easily implemented, versatile, no need to build models and if numbers of parameters increase then this algorithm works slowly.

Dataset is preprocessed before applying machine learning techniques. Missing values are dropped and NAN values are replaced by median values or replicating the previous values. Machine learning techniques: SVM, KNN, RF, and DT are applied to the processed data. The data are split 75:25 between training and testing and 10-fold cross-validation is used. The experimental results are shown in Table 7 in terms of accuracy, precision, recall and F1 score for ECG.

From Fig. 17, it is concluded that in GSR data, the performance of all the classes is above 82%. All the machine learning methods perform well on our dataset but KNN performance is low compared to other methods. In Fig. 18, on EMG data, SVM outperforms other methods by achieving results above 70% in terms of accuracy, precision, recall and F1-Score. In Fig. 19, on IMU data, the results are dropped for SVM but RF and KNN perform well. KNN and RF achieved an accuracy of 97%.

From Statistical and Machine learning-based analysis, it is drawn that data is imbalanced at some points but overall, the distribution is normal. Both methods are classifying the classes correctly and it is shown using graphical plots also. Except for EMG's data values, all classifiers are performing very well.

5 Conclusion

This paper describes the dataset acquired from 12 subjects performing various activities which include the electromyography data while simulating various hand movements that include hand at rest, fist, handshaking, waving and up/down movement of hand, Electrocardiography data collected while a subject is sitting and sleeping, emotions data using the GSR sensor when subjects are watching scary, sad, and happy clips and when the subject is holding her breath, breathing deeply and breathing normal, IMU data collected for 7 activities and 8 falls and smartwatch data which includes blood pressure, SpO2 and heart rate values. Basic health information is recorded manually that includes height, weight, smoking, drinking, medical history, etc. This dataset is very useful in the development of machine learning algorithms for detecting and predicting various activities and falls based on physiological and inertial features of subjects. Further, it is useful for developing and validating health monitoring systems for prediction and detection of events. The statistical and machine learning techniques-based analysis has given greater insights of the dataset which will be useful for researchers.

5.1 Limitations of the dataset

- Dataset is imbalanced. The performance of Machine learning model is significantly impacted by imbalanced and noisy data [40] but there are balancing techniques [41–

Table 7 Classification results of various classifiers for ECG classes

Sitting & deep breath while sleeping	SVM	KNN	DT	RF
Accuracy	98.1	81.74	98.09	99.45
Precision	98	82	98	99
Recall	98	81	97.5	99
F1-Score	98	81.5	98	99

49] that improve the performance of models by over-sampling or under-sampling the dataset.
- The dataset is created by collecting individual physiological and inertial features. Simultaneous collection of all types of features can give more promising results.

Acknowledgements The authors wish to acknowledge the assistance of Mr. Rakesh Sharma, Lab Technician, Computer Science and Engineering Department, at Thapar Institute of Engineering and Technology, Patiala, India for prototype designing.

Funding This research did not receive any specific grant from funding agencies in the public, commercial, not-for-profit sectors.

Data availability The dataset generated and analyzed during the current study is available at: https://data.mendeley.com/datasets/phb9y6cp5c/1.

Declarations

Competing interests The author declare that they have no competing of interest.

References

1. Coughlin S, Roberts D, O'Neill K, Brooks P (2018) Looking to tomorrow's healthcare today: a participatory health perspective. Intern Med J 48:92–96. https://doi.org/10.1111/imj.13661
2. RBCCM (2020) RBC Capital Markets | The healthcare data explosion. In: Rbccm. https://www.rbccm.com/en/gib/healthcare/episode/the_healthcare_data_explosion . Accessed 10 Jun 2023
3. Seshadri DR, Davies EV, Harlow ER et al (2020) Wearable sensors for COVID-19: a call to action to Harness our digital infrastructure for remote patient monitoring and virtual assessments. Front Digit Health 2:1–11. https://doi.org/10.3389/fdgth.2020.00008
4. Wu CT, Li GH, Huang CT et al (2021) Acute exacerbation of a chronic obstructive pulmonary disease prediction system using wearable device data, machine learning, and deep learning: development and cohort study. JMIR mHealth and uHealth 9. https://doi.org/10.2196/22591
5. Wong CK, Ho DTY, Tam AR et al (2020) Artificial intelligence mobile health platform for early detection of COVID-19 in quarantine subjects using a wearable biosensor: protocol for a randomised controlled trial. BMJ Open 10:1–5. https://doi.org/10.1136/bmjopen-2020-038555
6. Sareen S, Sood SK, Gupta SK (2018) IoT-based cloud framework to control Ebola virus outbreak. J Ambient Intell Humaniz Comput 9:459–476. https://doi.org/10.1007/s12652-016-0427-7
7. Sareen S, Gupta SK, Sood SK (2017) An intelligent and secure system for predicting and preventing Zika virus outbreak using fog computing. Enterp Inf Syst 11:1436–1456. https://doi.org/10.1080/17517575.2016.1277558
8. Thakur P, Kaur S (2018) An intelligent system for predicting and preventing Chikungunya virus. In: 2017 International conference on energy, communication, data analytics and soft computing, ICECDS 2017. Institute of Electrical and Electronics Engineers Inc., pp 3483–3492. https://doi.org/10.1109/ICECDS.2017.8390109
9. Sandhu R, Gill HK, Sood SK (2016) Smart monitoring and controlling of Pandemic Influenza A (H1N1) using Social Network Analysis and cloud computing. J Comput Sci 12:11–22. https://doi.org/10.1016/j.jocs.2015.11.001
10. Ahanger TA, Tariq U, Nusir M et al (2022) A novel IoT–fog–cloud-based healthcare system for monitoring and predicting COVID-19 outspread. J Supercomput 78:1783–1806. https://doi.org/10.1007/S11227-021-03935-W
11. Quer G, Radin JM, Gadaleta M et al (2021) Wearable sensor data and self-reported symptoms for COVID-19 detection. Nat Med 27:73–77. https://doi.org/10.1038/s41591-020-1123-x
12. Mishra T, Wang M, Metwally AA et al (2020) Pre-symptomatic detection of COVID-19 from smartwatch data. Nat Biomed Eng 4:1208–1220. https://doi.org/10.1038/s41551-020-00640-6

13. Dhaliwal MK, Sharma R, Bindra N (2023) Analyzing wearable data for diagnosing COVID-19 using machine learning model. In: Lecture notes in electrical engineering, Springer, Singapore, pp 285–299. https://doi.org/10.1007/978-981-19-5868-7_22
14. Mazzetta I, Zampogna A, Suppa A et al (2019) Wearable sensors system for an improved analysis of freezing of gait in Parkinson's disease using electromyography and inertial signals. Sens (Switzerland) 19. https://doi.org/10.3390/s19040948
15. Falih ADI, Adhi Dharma W, Sumpeno S (2017) Classification of EMG signals from forearm muscles as automatic control using Naive Bayes. 2017 Int Semin Intell Technol Its Appl Strength Link Between Univ Res Ind to Support ASEAN Energy Sect ISITIA 2017 - Proceeding 346–351. https://doi.org/10.1109/ISITIA.2017.8124107
16. Ajerla D, Mahfuz S, Zulkernine F (2019) A real-time patient monitoring framework for fall detection. Wirel Commun Mob Comput 2019:1–13. https://doi.org/10.1155/2019/9507938
17. Li Q, Stankovic JA, Hanson MA et al (2009) Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information. Proc – 2009 6th Int Work Wearable Implant Body Sens Networks, BSN 2009 138–143. https://doi.org/10.1109/BSN.2009.46
18. Majumder S, Mondal T, Deen MJ (2017) Wearable sensors for remote health monitoring. Sens (Switzerland) 17. https://doi.org/10.3390/s17010130
19. Ichwana D, Arief M, Puteri N, Ekariani S (2018) Movements monitoring and falling detection systems for transient ischemic attack patients using accelerometer based on Internet of Things. In: 2018 International Conference on Information Technology Systems and Innovation (ICITSI). IEEE, pp 491–496
20. Castro-García JA, Molina-Cantero AJ, Gómez-González IM et al (2022) Towards human stress and activity recognition: a review and a First Approach based on low-cost wearables. Electron 11:1–30. https://doi.org/10.3390/electronics11010155
21. Ayata D, Yaslan Y, Kamasak M (2017) Emotion recognition via galvanic skin response: comparison of machine learning algorithms and feature extraction methods. Istanbul Univ - J Electr Electron Eng 17:3129–3136
22. Santamaria-Granados L, Munoz-Organero M, Ramirez-Gonzalez G et al (2019) Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS). IEEE Access 7:57–67. https://doi.org/10.1109/ACCESS.2018.2883213
23. Sen K, Pal S (2021) Alternative method for pain assessment using EMG and GSR. J Mech Med Biol 21:1–24. https://doi.org/10.1142/S0219519421500391
24. Fu R, Wang H (2014) Detection of driving fatigue by using noncontact emg and ecg signals measurement system. Int J Neural Syst 24. https://doi.org/10.1142/S0129065714500063
25. Chueh TH, Chen TB, Lu HHS et al (2012) Statistical prediction of emotional states by physiological signals with manova and machine learning. Int J Pattern Recognit Artif Intell 26:1–18. https://doi.org/10.1142/S0218001412500085
26. Kim J, Andr´ E (2009) Detection of affective patterns in physiological signals towards improving automatic emotion recognition. Handb Pattern Recognit Comput Vis Fourth Ed 415–434. https://doi.org/10.1142/9789814273398_018
27. Soleymani M, Chanel G, Kierkels JJM, Pun T (2009) Affective characterization of movie scenes based on content analysis and physiological changes. Int J Semant Comput 3:235–254. https://doi.org/10.1142/S1793351X09000744
28. Nadeem A, Mehmood A, Rizwan K (2019) A dataset build using wearable inertial measurement and ECG sensors for activity recognition, fall detection and basic heart anomaly detection system. Data Br 27:104717. https://doi.org/10.1016/j.dib.2019.104717
29. Wagner P, Strodthoff N, Bousseljot RD et al (2020) PTB-XL, a large publicly available electrocardiography dataset. Sci Data 7:1–15. https://doi.org/10.1038/s41597-020-0495-6
30. Malešević N, Olsson A, Sager P et al (2021) A database of high-density surface electromyogram signals comprising 65 isometric hand gestures. Sci Data 8:1–10. https://doi.org/10.1038/s41597-021-00843-9
31. Ozdemir MA, Kisa DH, Guren O, Akan A (2022) Dataset for multi-channel surface electromyography (sEMG) signals of hand gestures. Data Br 41. https://doi.org/10.1016/j.dib.2022.107921
32. Sharma K, Castellini C, van den Broek EL et al (2019) A dataset of continuous affect annotations and physiological signals for emotion analysis. Sci Data 6:1–13. https://doi.org/10.1038/s41597-019-0209-0
33. Ojetola O, Gaura E, Brusey J (2015) Data set for fall events and daily activities from inertial sensors. Proc 6th ACM Multimed Syst Conf MMSys 243–248. https://doi.org/10.1145/2713168.2713198
34. DEAP: A dataset for emotion analysis using physiological and audiovisual signals. https://www.eecs.qmul.ac.uk/mmv/datasets/deap/. Accessed 10 June 2023

35. Sucerquia A, López JD, Vargas-Bonilla JF (2017) SisFall: a fall and movement dataset. Sens (Switzerland) 17. https://doi.org/10.3390/s17010198
36. Sikder N, Nahid AA (2021) KU-HAR: an open dataset for heterogeneous human activity recognition. Pattern Recognit Lett 146:46–54. https://doi.org/10.1016/j.patrec.2021.02.024
37. Mosquera-Lopez C, Wan E, Shastry M et al (2021) Automated detection of real-world falls: modeled from people with multiple sclerosis. IEEE J Biomed Heal Inf 25:1975–1984. https://doi.org/10.1109/JBHI.2020.3041035
38. Miranda-Correa JA, Abadi MK, Sebe N, Patras I (2021) AMIGOS: a dataset for affect, personality and mood research on individuals and groups. IEEE Trans Affect Comput 12:479–493. https://doi.org/10.1109/TAFFC.2018.2884461
39. Phinyomark A, Phukpattaranont P, Limsakul C (2012) Feature reduction and selection for EMG signal classification. Expert Syst Appl 39:7420–7431. https://doi.org/10.1016/j.eswa.2012.01.102
40. Roccetti M, Delnevo G, Casini L, Salomoni P (2020) A cautionary tale for machine learning design: why we still need human-assisted big data analysis. Mob Netw Appl 25:1075–1083. https://doi.org/10.1007/s11036-020-01530-6
41. Aslam N, Alzamzami O, Xia K et al (2023) Improving the review classification of Google apps using combined feature embedding and deep convolutional neural network model. J Ambient Intell Humaniz Comput. https://doi.org/10.1007/S12652-023-04529-5
42. Bagui S, Li K (2021) Resampling imbalanced data for network intrusion detection datasets. J Big Data. https://doi.org/10.1186/S40537-020-00390-X
43. Ramadhan NG (2021) Comparative analysis of ADASYN-SVM and SMOTE-SVM methods on the detection of type 2 diabetes mellitus. Sci J Inf 8:276–282. https://doi.org/10.15294/sji.v8i2.32484
44. Awal MA, Masud M, Hossain MS et al (2021) A novel bayesian optimization-based machine learning framework for COVID-19 detection from inpatient facility data. IEEE Access 9:10263–10281. https://doi.org/10.1109/ACCESS.2021.3050852
45. Li Y, Wang Y, Li T et al (2021) SP-SMOTE: a novel space partitioning based synthetic minority oversampling technique. Knowl-Based Syst 228:107269. https://doi.org/10.1016/j.knosys.2021.107269
46. Kim YT, Kim DK, Kim H, Kim DJ (2020) A comparison of oversampling methods for constructing a prognostic model in the patient with heart failure. Int Conf ICT Converg 379–383. https://doi.org/10.1109/ICTC49870.2020.9289522
47. Anđelić N, Lorencin I, Baressi Šegota S, Car Z (2023) The development of symbolic expressions for the detection of Hepatitis C patients and the disease progression from blood parameters using genetic programming-symbolic classification algorithm. Appl Sci 13:1–33. https://doi.org/10.3390/app13010574
48. McCallan N, Davidson S, Ng KY et al (2023) Rebalancing techniques for asynchronously distributed EEG data to improve automatic seizure type classification. 2023 57th Annu Conf Inf Sci Syst CISS 1–6. https://doi.org/10.1109/CISS56502.2023.10089669
49. Kaur R, Sharma R, Dhaliwal MK (2023) Evaluating Performance of SMOTE and ADASYN to classify falls and activities of daily living. In: Proceedings of the 12th International Conference on Soft Computing for Problem Solving(SocProS 2023) Moving Toward Society 5.0 Department of Applied Mathematics and Scientific Computing IIT Roorkee and Liverpool Hope University, UK
50. Bhoi SK, Panda SK, Patra B et al (2018) FallDS-IoT: a fall detection system for elderly healthcare based on IoT data analytics. Proc – 2018 Int Conf Inf Technol ICIT 155–160. https://doi.org/10.1109/ICIT.2018.00041