



# Real-time prediction of smoking activity using machine learning based multi-class classification model

Saurabh Singh Thakur<sup>1</sup>  · Pradeep Poddar<sup>2</sup> · Ram Babu Roy<sup>1</sup>

Received: 12 November 2020 / Revised: 18 August 2021 / Accepted: 18 January 2022 /  
Published online: 25 February 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Smoking cessation efforts can be greatly influenced by providing just-in-time intervention to individuals who are trying to quit smoking. Detecting smoking activity accurately among the confounding activities of daily living (ADLs) being monitored by the wearable device is a challenging and intriguing research problem. This study aims to develop a machine learning based modeling framework to identify the smoking activity among the confounding ADLs in real-time using the streaming data from the wrist-wearable IMU (6-axis inertial measurement unit) sensor. A low-cost wrist-wearable device has been designed and developed to collect raw sensor data from subjects for the activities. A sliding window mechanism has been used to process the streaming raw sensor data and extract several time-domain, frequency-domain, and descriptive features. Hyperparameter tuning and feature selection have been done to identify best hyperparameters and features respectively. Subsequently, multi-class classification models are developed and validated using in-sample and out-of-sample testing. The developed models obtained predictive accuracy (area under receiver operating curve) up to 98.7% for predicting the smoking activity. The findings of this study will lead to a novel application of wearable devices to accurately detect smoking activity in real-time. It will further help the healthcare professionals in monitoring their patients who are smokers by providing just-in-time intervention to help them quit smoking. The application of this framework can be extended to more preventive healthcare use-cases and detection of other activities of interest.

**Keywords** Preventive healthcare · mHealth · Smoking cessation · Wearable sensors · Predictive modeling · Personalized healthcare · Multimedia applications · IoT

---

✉ Saurabh Singh Thakur  
saurabhjan07@gmail.com

Pradeep Poddar  
pradeppoddar.iitkgp@gmail.com

Ram Babu Roy  
rambabu@see.iitkgp.ac.in

Extended author information available on the last page of the article

# 1 Introduction

## 1.1 Background and motivation

The detrimental impact of addictive behavior on health is widely studied and is a well-known fact [28]. Tobacco smoking is one of the most widespread but modifiable risk factors for various non-communicable diseases like cardiovascular diseases (CVD), stroke, chronic obstructive pulmonary disease (COPD), and cancer [36]. The habit of tobacco smoking in the form of cigarettes, cigars, etc. is highly addictive and at the same time, it is injurious to one's health. Smoking habit is difficult to quit as abstinent smokers often start with a sudden craving for smoking which results in lapse and then permanent relapse. Despite several efforts, the cessation rate of smoking is not encouraging enough. It becomes more challenging to improve upon cessation rates due to a lack of effective methods for delivering the just-in-time intervention to the smokers who are willing to quit smoking [53].

There are several apps for helping those who want to quit smoking. Such applications are often called persuasive applications and the technique is termed as persuasive technology in the literature [11, 16]. However, these apps are mostly based on self-reporting. A study reported that adherence to app features influences the quit rate. The app provides various features such as a personalized quit plan, tracking progress, and interactive-ness to facilitate quitting [49]. The main road-block in such types of apps is adherence to using the app without losing the motivation [1, 31, 49]. An application MapMySmoke used motivational messaging to promote smoking cessation [54] where the patient is required to manually log the smoking and craving events and the application detects the location of the patients using geo-positioning technology. Based on these logs and location information, the doctor of the patient provides a personalized quit attempt plan. However, only 50% of participants logged their smoking details regularly resulting in poor quit rate. Another study on smoking cessation applications revealed that instead of using Behavior Change Technology (BCT) based features, applications rather focuses on ease-of-use features [63]. Therefore, using self-reporting based or educational applications has a limited impact on improving the quit rate or providing a personalized quit plan.

Detecting the smoking log in near real-time is an ongoing and challenging research problem and it can overcome the limitation of manually entering the smoking logs. Further, it makes mobile app-based intervention possible to motivate the user to quit smoking just-in-time. A recent development in mHealth [2] and persuasive technologies [65] provides an opportunity to intervene effectively and help in smoking cessation. The developments in the area of IoT, sensors, machine learning, mobile computing enable ubiquitous monitoring of ADLs with the help of body wearable sensor devices in an unobtrusive manner. Computational intelligence can help in identifying these ADLs like smoking, walking, running, and sleeping in real-time. With the help of smartphones and mobile app, the behavioral feedback can be communicated back to the user instantly. Furthermore, mobile applications (apps) can also be used to deliver the interventions for behavioral modifications.

Keeping in mind the general psychology of smoking behavior, a solution is envisaged in which a near real-time intervention could help in smoking cessation [55, 60] by motivating users for behavior change [8, 37, 45, 61]. Body wearables like smartwatches, bands, etc. could track physical activity, sleeping behavior in real-time. These wearable devices consist of sensors like accelerometer, gyroscope, magnetometer, etc. which provides the raw data signals generated due to body movement. Further, computational intelligence helps in processing raw sensor data for prediction, classification, or recognition of activities like running, walking, sitting, and laying [37]. Data analytics helps in building the personalized behavioral profile of a user with the help of mobile apps. This whole process helps in motivating users to adopt healthy behavior, changing lifestyles, and thus improving one's health. Recognition of cigarette smoking is one such activity which could be included as one of the features in such smart wearable devices and application. There are some studies conducted in this direction which attempts to detect smoking activity with the help of sensor data [53, 55, 59]. Such studies are limited to smoking activity and do not account for other ADLs limiting its practical application. Further, these studies are at the preliminary level and are limited to the prediction of smoking episodes, detecting smoking gestures, and first lapse detection. Other limitations are the obtrusive use of wearable sensors.

These studies do not propose a holistic solution for smoking cessation due to its limitation of post-processing of sensor data. In a study conducted by Thakur et.al., a mobile app-based model of smoking cessation was conceptualized. The study had shown that the smoking activity signals are periodic, and therefore utilizing this periodicity of the signals, a recognition system could be developed, which is capable of detecting the smoking activity among ADLs in near-real-time [60]. In this study, the authors conceptualized a holistic solution where the ADLs can be measured using the IMU sensor and recognized in real-time using machine learning models. Further, interventions can be delivered in real-time using a smartphone app.

## 1.2 Research objectives

In this study, we aim to develop an activity recognition system for recognizing activities of daily-living including smoking in near real-time. Motivated by a lack of studies on real-time prediction of smoking and other daily living activities, this paper has been developed to address following research objectives which in turn will lead to the development of automated prediction of smoking activity system in near-real time:

- 1) *To derive features from the raw sensor signals of streaming nature to build predictive models for predicting the smoking activity along with the activities of daily living.*
- 2) *To build a generalized predictive modeling framework suitable for real-time predictions of activities using IMU sensor data.*
- 3) *To identify important features variables for the prediction of the activities.*
- 4) *To assess the generalizability of the modeling framework for predicting activities of interest in preventive healthcare settings.*

The rest of this paper is organized as follows. Section 2 discusses the relevant literature in sensor-based detection of smoking activity. In Section 3, details of the experimental setup, data acquisition, and solution approach are given. In section 4, we present the findings of the experiments carried out, followed by a brief discussion of

the findings and comparison with baseline studies. Preventive healthcare implications of the proposed modeling framework are presented in Section 5 and the conclusions of this study are given in Section 6.

## 2 Literature review

### 2.1 Review of methods proposed in the literature for promoting smoking cessation

Various methods for encouraging smoking cessation have been discussed in the literature. The methods with reasonable efficacies reported in the literature are pharmacological treatments [27] like nicotine replacement therapy (NRT), combination therapy, non-nicotine medications and non-pharmacological treatments like training, behavior therapy, counseling, e-cigarettes, etc. [23]. According to a study conducted by Heydari et al. (2015) on combination therapy, which involved NRT along with the counseling is reported to be the most preferred way of treatment [24]. Various studies reported that the rate of smoking cessation is higher in the adult population as compared to older age groups [33, 34]. A recent study revealed that the smoking cessation rate could be improved if physicians put more effort towards counseling patients in outpatient services and encourage follow-up visits [26]. Despite several methods and expenditures, the rate of smoking cessation is not satisfactory and requires significant efforts to bridge the gaps. The factors like the inability to track patient's progress, reliance on patient's self-reporting, poor engagement, and lack of emotional involvement are said to be the major hurdles towards smoking cessation [5, 25, 32, 66, 69].

### 2.2 Review of smartphone-based applications for promoting smoking cessation

With the advancements in information and communication technologies (ICT's) the mode of delivering healthcare services are also changing rapidly [58]. It opens doors for the physicians and the patients to exchange information quickly and transparently, crossing geographical limits. Effective use of technology can cater to alleviating the hurdles in delivering effective treatment to the patients [58, 61]. Recently, a significant number of studies conducted on quit smoking, which uses mobile applications or wearable devices. The method of targeting behavior change using mobile applications is called as persuasive technology [51, 65]. A study suggests that computer-based intervention (CBI) is more effective as compared to in-person brief intervention (IBI) in cases of substance use. CBI has advantages like cost, reliability, honest self-reporting from a patient, and regular feedback from the healthcare provider [56]. Besides it also aids self-monitoring, progress tracking, and daily reminders [22]. Most of the mobile applications available on google play store which offers smoking cessation or quit smoking features are ineffective. Factors like losing interest in using apps, unable to provide incentive, lack of evidence-based clinical practices, more reliance on user-based reporting are some of the reasons that causes poor adherence to use such apps and resulting in futile efforts. [1, 17]. Therefore, there is a need to devise mechanisms to make such applications more engaging to improve adherence to the interventions delivered through these applications. Apps with features like gaming can also help in increasing its usage and adherence [31].

Applications using evidence-based clinical practice and behavior change theories are more effective for engaging patients in the long term. Long term engagement of patients with such

apps has shown improved results in quit smoking [22, 49, 63]. Due to the self-reporting feature in these apps, patients gradually lose interest in using the application and results in poor adherence [51]. These applications lack the required intelligence, static, and unable to provide tailored and personalized feedback [25].

### 2.3 Review of literature on the detection of smoking activity using wearable sensor devices

An automated detection or logging of smoking events can help in providing personalized and tailored feedback. It is one of the challenging research problems of the field of human activity recognition using sensors. Various approaches has been discussed for human activity recognition in real-time [20, 68]. To detect smoking automatically requires using some type of body wearable sensors. These sensors can be an accelerometer, gyroscope, respiratory sensors, smart lighter, and smoke detection sensors. The key issues while using these wearable sensor devices are unobtrusive sensing [70] and ubiquitous sensing [38]. It is observed that wrist wearable IMU (Inertial Measurement Unit) sensors are mostly used in different studies to detect smoking activity [53, 57, 59, 60, 64]. In some research, a respiration sensor, heart rate sensor, smoke sensor, or smart lighter has also been used in addition to the IMU sensor [51, 53, 57, 60].

In a study, a 6-axis (accelerometer and gyroscope) IMU sensors to be worn on both wrists and body wearable electrocardiograph (ECG) sensor suite has been used for recording respiration data. Their model did well when used IMU and ECG data together, but the performance was poor when these sensors were used in silos. Further, the body wearable suit is quite obtrusive for practical applications [53]. In a similar study, four 6-axis IMU sensor attached on the wrist, lower elbow, upper elbow, and lower arm [46] has been used. The objective of this study was to classify smoking and non-smoking events and to determine the optimal number of sensors required and their position on the body to detect smoking events. The predictive accuracy of the classification models developed in this study could be improved, and further, they had used 4 sensors at different body positions which is not feasible in practical applications. Building highly accurate predictive models without using obtrusive sensing could be a viable option for practical usage and we worked with the same objective in our study.

The IMU sensor data is streaming in nature and carries unique signatures of different activities involving the motor movement of the human body. However, earlier studies have not exploited this nature of sensor data. Therefore, earlier proposed approaches were lacking in performance limiting its practical applicability. It is more prevalent and comfortable to wear sensors in the form of a wrist band or smartwatches. Therefore, hand movement becomes extremely important in identifying activities of daily living. Activities like smoking, walking, running, eating, drinking, talking, etc. involves a significant amount of hand movement. IMU sensor data needs to be translated into a suitable form to detect or classify these activities which brings the importance of feature extraction. The size of the time-window that should be considered to derive the feature vector from the streaming data is equally important. To find an optimal time-window size requires experimentation with the segmentation of streaming sensor data in different size time- windows. In our study, we have used a 50% overlapping window of 1, 3, 5 s respectively to create feature vectors. The raw data corresponding to each window is processed to derive the associated feature

vector. The feature vector we extracted consist of various time domain, frequency domain, and descriptive features. The data we used are all properly labeled and, therefore, we build several supervised machine learning-based models. In subsequent sections, we shall explain the complete methodology and results.

### 3 Methodology

This study aims to develop a modeling framework for activity recognition with a special focus on smoking activity. Accordingly, a framework is developed which is as shown in Fig. 1. The framework consists of 3 layers which are data preparation, model development, and model application. In the first layer, raw data from the IMU sensor is processed, transformed and features are extracted. The feature vectors are then labeled to generate the final input data which becomes the output for the model development layer. In this layer a classification model is developed, tuned, best features are identified, and the model is evaluated and validated. The best model identified in this layer is served in the model application layer where new data streaming from the sensor is used to make predictions of the activities in real-time. In the following sub-sections, we will discuss the methodology and solution approach in detail.

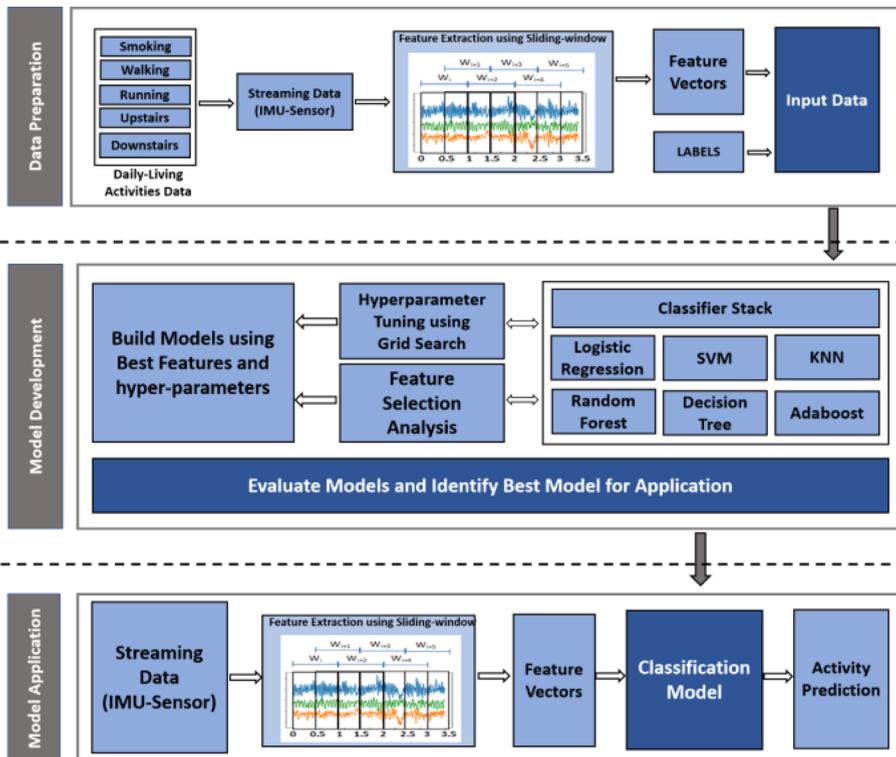


Fig. 1 Framework for developing a real-time activity prediction model

### 3.1 Wearable wrist band development

A prototype of a wearable wrist band has been developed for experimental purposes, as shown in Fig. 2, to capture the hand motion patterns of the activities. The band was developed by assembling MPU-9250 (a 6 axial inertial sensor) sensor unit with an Arduino nano micro-controller (ATmega328). The data logger unit with RTC DS3231 was used to capture the sensor data. A program was written and uploaded in the micro-controller to write the data from the accelerometer and gyroscope sensors in the SD card via a data logger unit along with time-stamps whenever the circuit is turned on. The frequency of the data points is 14 Hz. The pin diagram of the circuit module is as shown in Fig. 3.

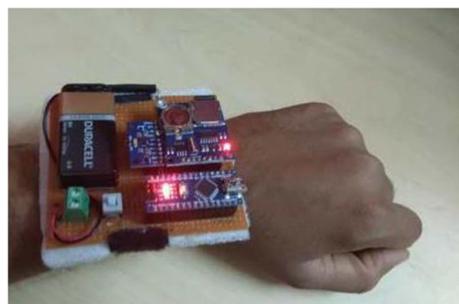
### 3.2 Data collection

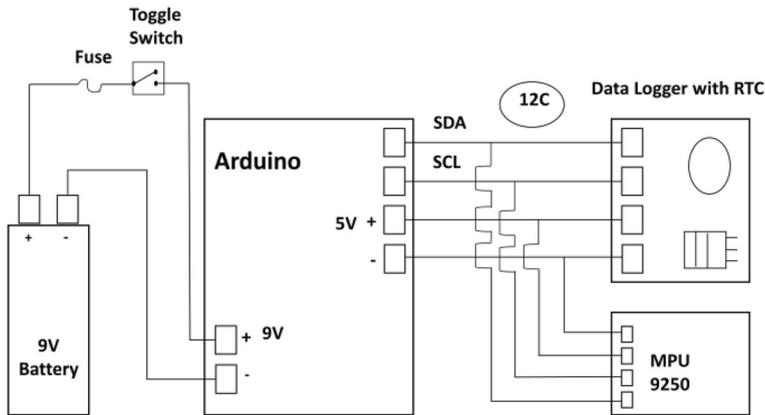
Data was collected for activities like walking, running, walking-upstairs, walking-downstairs, and smoking. During these activities' participants were asked to wear the wearable wrist band. To conduct this study, ethical clearance was taken from the Institute Ethical Committee of Indian Institute of Technology Kharagpur, India. The data was collected from a total number of 13 male participants in March 2019. Participation in the study was voluntary and informed consent was taken before starting the data collection. Out of 13 participants, 7 individuals aged between 22 and 24 years participated in providing data for the physical activities. The other 6 participants, who were regular cigarette smokers aged between 30 and 36 years, participated in providing data for their smoking activity. The raw sensor data was annotated with their corresponding activities namely smoking, walking, running, walking-upstairs, and walking-downstairs. The baseline characteristics of the study data are presented in Table 1.

### 3.3 Feature extraction

Feature extraction has been carried out on the raw sensor signals from the accelerometer and gyroscope. The time-domain features, frequency domain features, and descriptive features were extracted from each axis x, y, and z of both sensors based on similar work carried out in earlier studies [3, 14, 35, 42, 57]. The sensor data is of streaming nature and an appropriate method was required through which features can be extracted in real-time. Therefore, the concept of the sliding window is used for feature extraction. This window keeps collecting the streaming sensor data depending upon its size and carries out the feature extraction from the raw data. Also, we have kept an overlap of 50% from the previous window for each

**Fig. 2** Image of the wrist-wearable band used to capture hand motion data





**Fig. 3** Circuit diagram of the wearable band

subsequent window. It has been done to ensure that the characteristics of every trivial gesture present in any activity are captured completely in the feature vector that is generated from the sensor data of that sliding window. In this work, we tried and tested with different window sizes of 1-s, 3-s, and 5-s to identify which is an optimal window size for creating features. The sensor data were sampled at 14 Hz which means that the length of the data segment is,  $N = 14$ , for each epoch of 1 s. Therefore, the data size corresponding to the 3 windows turns out to be as,  $N = 14, 42, \text{ and } 70$ .

A total number of 8 time-domain features were extracted for each axis of both sensors. Therefore, the total number of time-domain features for each window is 48. These are mean, standard deviation, peak-to-peak, root mean square, skewness, mean absolute deviation, kurtosis, median, and inter-quartile-range. Three frequency domain features were extracted from each window, and these features are maximum amplitude, spectral energy, and frequency-domain entropy. Total of 18 features was generated in the frequency domain considering all the axis. At first, the Fast Fourier Transform (FFT) is used to transform the data from the time domain to the frequency domain. Then the frequency domain features are extracted for every window of data. Pearson Correlation coefficient has been used as the descriptive feature. It is measured using Pearson's correlation coefficient between the signals of all the axes of the IMU sensor. A total of 15 correlation coefficients were computed from the binary combination of the 6 axes of the IMU sensor. The detailed explanation of the sliding window and features extracted in this study is given in Appendix A.

**Table 1** Baseline characteristics of the study data

Activity	No. of participants	No. of episodes	Total duration (in minutes)
Smoking	6	32	137.71
Walking	7	19	91.32
Running	7	18	40.65
Walking-upstairs	7	19	27.27
Walking-downstairs	7	19	21.38

### 3.4 Development of machine learning based multi-class classification models

In this study, a supervised multi-class classification model has been developed which attempts to assign a class label to a new input feature vector. Such a model can be described as a training data set  $(x_1, y_1)$ , where  $x_1 \in R_n$  is the  $i_{th}$  input feature vector and  $y_1 \in (1, \dots, K)$  is the  $i_{th}$  class label. The objective is to find a learning model  $f$  such that  $f(x_1) = y_1$  for new unseen input feature vectors [6]. The predictive strength of several classification models has been analyzed which facilitates choosing the best model for the application. The stack of classification models consists of Logistic regression (LR) [12, 67], k-nearest neighbour (KNN) [4, 10], adaptive boosting (AdaBoost) [18, 19], random forest (RF) [9], support vector machines (SVM) [13] and a decision tree classifier (DT) [52]. We used the Python distribution Anaconda of version 5.1.0 and various libraries of scikit-learn 0.19.1 to build and evaluate these classifiers [39].

In-sample and out-of-sample validation had been carried out to measure the predictive efficacy of the models and also to check its generalizability on new unseen data [7, 50]. In the in-sample validation approach, the training samples are used to make predictions using the developed model. In-sample validation provides an upper bound of predictive accuracy of the developed model. The out-of-sample validation test data sample which has not been used during training and is unseen to the model is used to make predictions using the developed model. Out-of-sample validation ensures the generalizability of the model. Further, grid search using 5-fold cross-validation had been carried out to identify the best hyper-parameters of the models which were later used during model development. The area under the ROC curve (AUC) is a standard metric used to evaluate a classification task [21, 41]. AUC is used to evaluate the performance of the various classification models being developed. As discussed in the previous section, we had used a different window size of 1 s, 3 s, and 5 s; therefore, we developed the same models three times using different window sizes and compared their outcome with each other. It informed us of the optimal window size to consider while generating features from the raw sensor data.

### 3.5 Feature selection analysis

Sequential backward feature selection (SBFS) which is a wrapper-based method had been used to select the best features from the pool of all features [43]. This method belongs to the class of greedy search algorithms. In SBFS, a d-dimensional feature space is reduced to a k-dimensional feature space where  $k < d$ . SBFS algorithm is implemented using a python library *mlxtend* [47]. Feature selection analysis had been carried out to identify the most relevant features, to mitigate the curse of dimensionality which occurs for high-dimensional feature spaces with a limited number of samples, to remove the noise or irrelevant features and to reduce the computational time complexity.

## 4 Experimental results and discussions

### 4.1 Data visualization and analysis of raw sensor signals

A brief analysis of the raw patterns of the activities has been carried out with the help of data visualization. Line plots of the raw sensor data have been plotted for 60 s to understand the

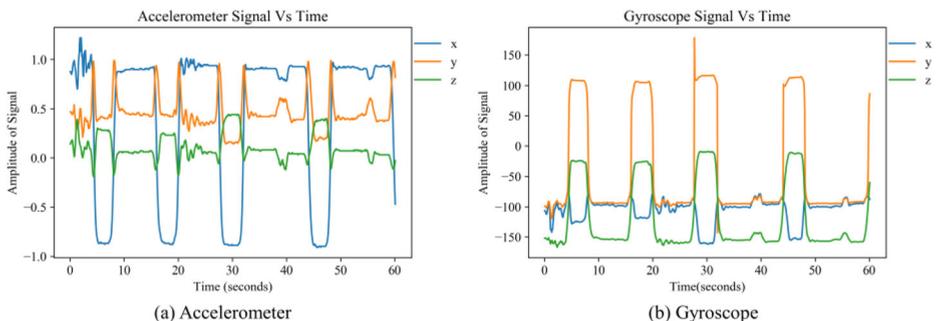
signal signature for different activities. The patterns of signal for each activity have been analyzed. Signal patterns of smoking activity are as shown in Fig. 4a, b. It has been observed that signals of the smoking activity are periodic, and the peaks appear at a regular interval. It reflects the hand movement pattern during smoking in which few gestures repeat at regular intervals. These gestures are hand-rise, puffing, and hands-down. Walking activity is shown in Fig. 5a, b which shows high acceleration, these signals are also periodic but the repetition is highly frequent, which resembles faster hand movement during walking as compared to smoking. Running activity data shown to have the highest acceleration as compared to the other activities being considered. It can be observed from the plots shown in Fig. 6a, b that the signals follow a sinusoidal trend. Running and walking activity signals share a similar signature to some extent, with different intensities. The intensity of running activity signals is very high as compared to that of walking activity signals. Figure 7a, b represents walking-downstairs, and Fig. 8a, b represents walking-upstairs are very similar but the first one is more frequent than the second because the spacing between the peaks of downstairs activity is less than walking-upstairs. It is observed from the raw patterns of the different activities that each activity bears a unique signature, especially smoking, walking, and running. Upstairs and downstairs activity shares close similarities. A natural similarity exists between walking, upstairs, and downstairs activity as they are walking activity only with trivial variations. The difference in the raw patterns of the activities is characterized with the help of feature extraction exercise. These features are utilized in building the predictive models which are explained in the further section.

## 4.2 Results of multi-class classification models

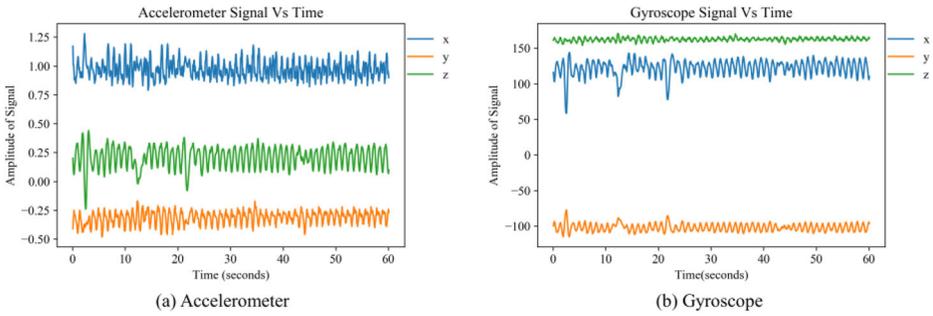
A champion-challenger method [29] is presented in the modeling framework to select the best model. Champion-challenger is a type of hybrid ensemble approach where several models compete with each other and the model with the highest predictive accuracy is chosen for the actual application. Hyperparameter tuning, and feature selection analysis has also been carried out to identify the best combination of hyperparameters and best features.

### 4.2.1 Hyperparameter tuning

At first, the best hyper-parameters are identified for each model using 5-fold cross-validation (CV) based grid search [40] approach. These hyper-parameters and their values were used while building the models throughout the experiments and are given in Table 2.



**Fig. 4** Raw signal patterns of smoking activity. **a** Accelerometer, **b** Gyroscope



**Fig. 5** Raw signal patterns of walking activity. **a** Accelerometer, **b** Gyroscope

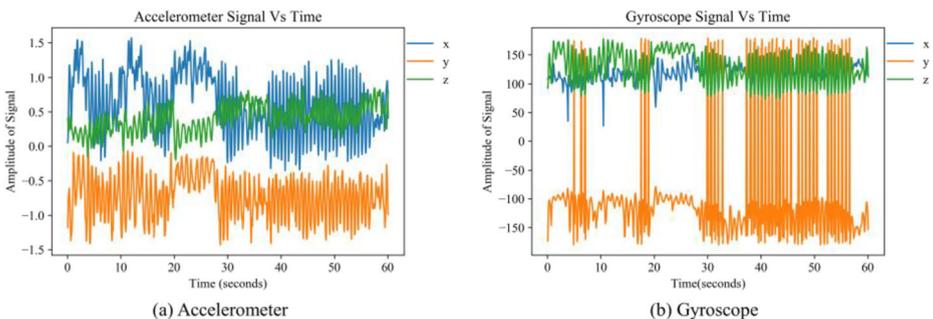
### 4.2.2 Feature selection analysis

The most prominent features were identified using the feature selection algorithm, as discussed in the previous section. The algorithm used in this study for feature selection is sequential backward feature selection (SBFS). SBFS method gives the best subset of features as output. Then the classification model is build using this selected subset of features. The output of feature selection is explained with the help of Fig. 9. The output plots resemble the elbow-curve, and the elbow-joint in the curve can be considered as the point where we get the optimal number of features required to build a model with the most optimized predictive accuracy. It can be observed from the figure, that the optimal predictive accuracies are being started receiving from the feature subset of size 12 features.

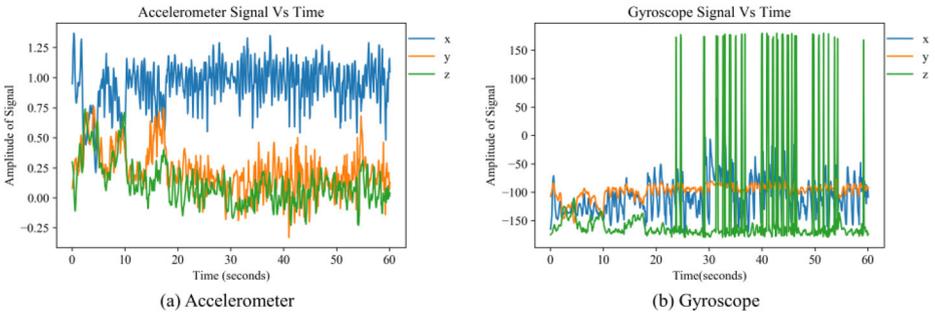
However, the highest accuracy for a window of 1 s is received from a subset of 44 features with predictive accuracy being at 88.8%. Similarly, for a 3-s window, the highest accuracy is achieved at 91.9% with a subset of 56 features and for a 5-s window, the highest accuracy is achieved at 93.1% with a subset of 27 features. Once the reduced feature subset is obtained after running the SBFS algorithm, we build a classification model for the classification of activities.

### 4.2.3 Feature importance analysis using heat-map

Next, we analyzed which are the most important features. We initially had a set of 87 features. SBFS method gave us 87 sets of features and the size of each set is,  $S \in \{1, 2, 3, \dots, 87\}$ . We used this data to get the total number of occurrences of each feature in these sets. The



**Fig. 6** Raw signal patterns of running activity. **a** Accelerometer, **b** Gyroscope



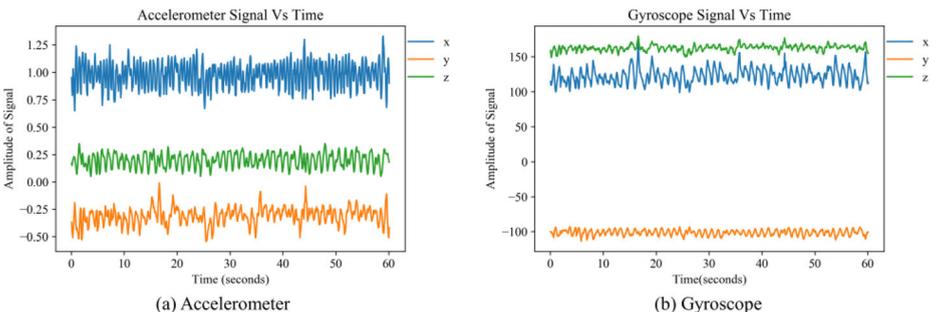
**Fig. 7** Raw signal patterns of walking-downstairs activity. **a** Accelerometer, **b** Gyroscope

count of total occurrences is then normalized in the range (0, 1). The resultant values are the importance score of each feature. A heat map is drawn using this information and is shown in Fig. 10. From the heat map, it is evident that the accelerometer features are more important than the gyroscope, further within the descriptive features which are the correlation of different axis of accelerometer and gyroscope also shows that the correlation between signals of different accelerometer axis and between accelerometer and gyroscope axis have more importance towards the predictive accuracy of the activities.

**4.2.4 Classification model development and validation**

Classification models were build using feature data set generated by each sliding window size of 1 s, 3 s, and 5 s on the model-input data with all the features and the model-input data with only the best features identified after feature selection analysis. In total 6 experiments were carried out to develop these models which are as mentioned below:

1. Experiment 1: Model development had been done using all the features on the dataset generated using a 1-s sliding-window-based feature extraction process.
2. Experiment 2: Model development had been done using all the features on the dataset generated using a 3-s sliding-window-based feature extraction process.
3. Experiment 3: Model development had been done using all the features on the dataset generated using a 5-s sliding-window-based feature extraction process.



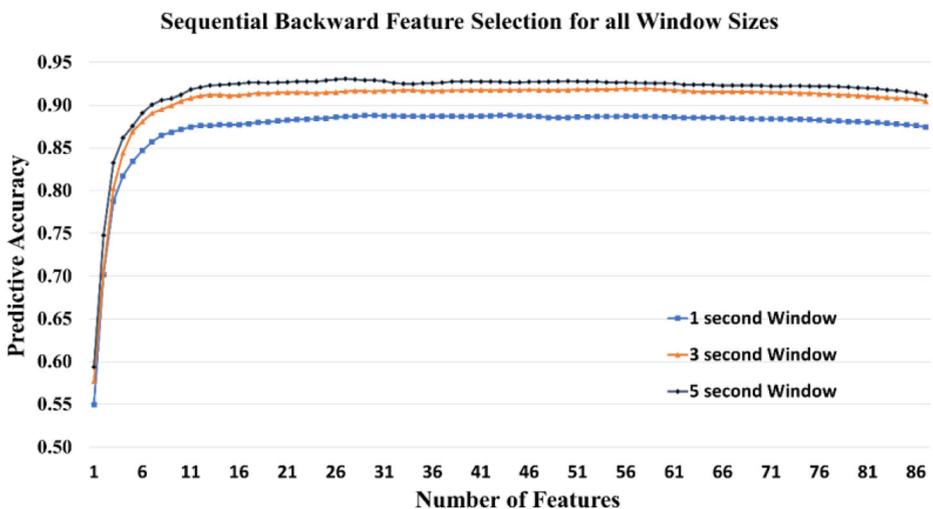
**Fig. 8** Raw signal patterns of walking-upstairs activity. **a** Accelerometer, **b** Gyroscope

**Table 2** List of hyper-parameters along with their best values to achieve the best estimator during modeling

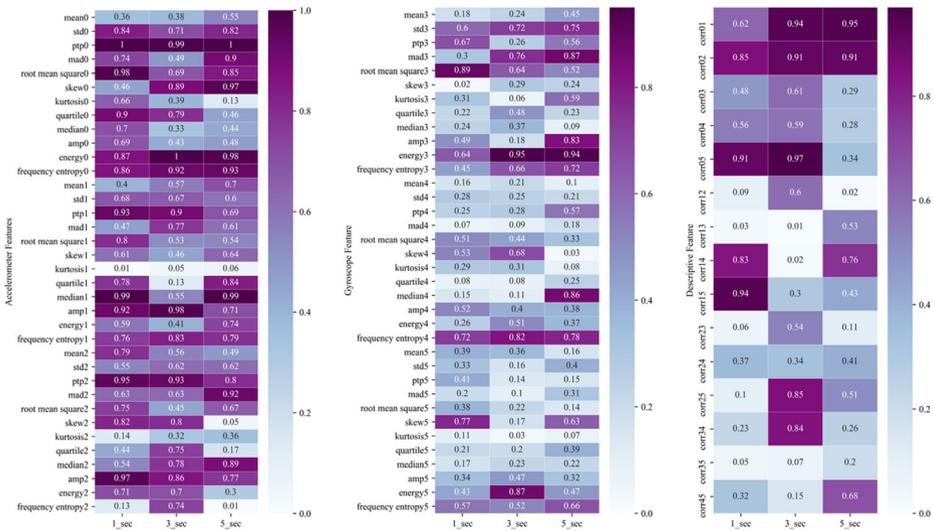
Classification model	Parameter of best estimator	Value	Accuracy of best estimator
Adaboost	no. of estimators	1000	78.39
	learning rate	1	
RF	max depth	30	94.13
	no. of Estimators	1000	
SVM	C	10	94.09
	gamma	0.01	
	kernel	rbf	
KNN	metric	Manhattan	93.15
	no. of neighbors	1	
	weights	Uniform	
Logistic Regression	C	20	89.35
	penalty	l2	
DT	criterion	entropy	87.46
	max depth	21	

4. Experiment 4: Model development had been done using the best features subset derived using SBFS on the dataset generated using a 1-s sliding-window-based feature extraction process.
5. Experiment 5: Model development had been done using the best features subset derived using SBFS on the dataset generated using a 3-s sliding-window-based feature extraction process.
6. Experiment 6: Model development had been done using the best features subset derived using SBFS on the dataset generated using a 5-s sliding-window-based feature extraction process.

The detailed validation results are presented in Appendix B. For better readability, model validation results for only smoking activity are shown below in Tables 3, 4, 5, 6, 7, and 8 respectively for each of the above experiments. The findings from these experiments are



**Fig. 9** Output of feature selection algorithm for all the windows



**Fig. 10** Analysis of the importance of all the features using a heat map

manifold. The predictive accuracy of classifiers for smoking, running, and walking activities was equally comparable. Decision tree, KNN, and Random forest overfitted as their in-sample AUC is 100% and it dropped significantly in out-of-sample validation. Of which decision tree is more biased and overfitted as its AUC is dropping close to 3.5%. The overall performance is highest in the SVM classification model in a 3-s all feature model developed in experiment 2 with an AUC of 98.799%. The classification models like Adaboost, Logistic Regression, and SVM are generalizable as the difference between AUC values of in-sample and out-of-sample validation is not very significant implying that models are performing well on unseen data. The performance of classifiers is distinguishably better in 3-s and 5-s windows as compared to 1-s window for all the activities. The overall AUC is less in case of walking-upstairs and walking-downstairs activities as compared to the AUC of the other activities. This could be due to less sample size available for training and validation of these two activities. The predictive accuracy in experiments 4, 5, and 6 are equally comparable or better with their counter-parts in experiments 1, 2, 3 respectively. It implies that feature selection led to the improvement of the model both in terms of predictive performance and prediction time optimization. A sliding window of 3 s is most optimal as it gives the highest predictive accuracy and at the same time, the first response time of such a prediction model in practical applications will be less than a sliding window of 5-s.

**Table 3** In-sample and out-of-sample validation of classification model developed in experiment-1 for smoking activity

Classification model	AUC (in-sample testing)	AUC (out-of-sample testing)
Adaboost	0.97672	0.97099
Decision Tree	1	0.95815
KNN	1	0.97286
Logistic Regression	0.9681	0.96533
SVM	0.99536	0.9808
Random Forest	1	0.97616

**Table 4** In-sample and out-of-sample validation of classification model developed in experiment-2 for smoking activity

Classification model	AUC (in-sample testing)	AUC (out-of-sample testing)
Adaboost	0.96791	0.96164
Decision Tree	1	0.95996
KNN	1	0.98185
Logistic Regression	0.98191	0.97556
SVM	0.99904	0.98799
Random Forest	1	0.98321

### 4.3 Discussions

The experiments were carried out to identify an optimal approach for real-time prediction of smoking activity using supervised machine learning-based multi-class classification methods. Various time domain and frequency domain features were derived from the raw sensor data. A sliding window mechanism has been used to address the streaming nature of sensor signals. Determining an optimal window size was critical to develop an accurate prediction model. Therefore, sliding windows of different sizes (1-s, 3-s, and 5-s) were used to derive the feature variables from the raw sensor data. Subsequently, experimentation has been carried out with different window sizes to identify which is most important to build a classification model. To summarize, our objectives for this study were: To engineer an exhaustive feature set from IMU sensor raw data and develop a multi-class classification model using this feature set as an input; To identify an optimal sliding window size to gain the maximum predictive accuracy; To compare different classification algorithms and identify the best classification method for this problem; To conduct a feature selection analysis to identify the most important features.

To develop an accurate classification model, the quality of data becomes very important on which it is to be trained. To ensure that, data of physical activities like walking, running, walking-upstairs, walking-downstairs along with smoking activity is gathered in free-living conditions from the subjects for a total duration of 318.33 min. We used this data for the development of the classification model. The authors were interested in evaluating different questions, as discussed in the previous paragraph. The first question is to find the best classification method for activity detection. The experimental findings suggest that the SVM and logistic regression classification models were the most generalized learner because the difference between the predictive accuracy of their in-sample and out-of-sample validation is very less. SVM further achieved the highest predictive accuracy in almost all the experiments.

**Table 5** In-sample and out-of-sample validation of classification model developed in experiment-3 for smoking activity

Classification model	AUC (in-sample testing)	AUC (out-of-sample testing)
Adaboost	0.97622	0.96708
Decision Tree	1	0.96947
KNN	1	0.98531
Logistic Regression	0.98691	0.9859
SVM	0.99983	0.98757
Random Forest	1	0.97956

**Table 6** In-sample and out-of-sample validation of classification model developed in experiment-4 for smoking activity

Classification model	AUC (in-sample testing)	AUC (out-of-sample testing)
Adaboost	0.97301	0.97113
Decision Tree	1	0.96
KNN	1	0.97486
Logistic Regression	0.95707	0.95703
SVM	0.98906	0.98132
Random Forest	1	0.97614

KNN, decision tree, and random forest model were suffered by the overfitting and bias. The application of such models is not very safe in real-life scenarios.

#### 4.4 Comparative analysis with the baseline studies

The results of this study are further compared with the baseline studies. The comparison was primarily based on the 4 criteria; the performance of the algorithms/classifiers, obtrusivity of the sensors, consideration of confounding activities in the model, and focus on the real-time prediction. These dimensions signify the practical applicability of the model. The comparison with the baseline studies is presented in Table 9. The comparison suggests that the SVM classification model is most appropriate to classify activities using IMU sensor data. In terms of the performance of the model, this study has achieved the highest predictive accuracy for smoking. Some of the baseline studies have also achieved comparable accuracy but they have used more sensors at different body positions which increases the obtrusiveness and impacts the practicality. The obtrusivity is defined in terms of mild, high, and very high in the table. Mild obtrusive means the use of sensor wearable at a single body position and mostly on the wrist. High means using more than one type of sensor wearable or at 2 different body positions whereas very high obtrusivity is classified when more than one type of wearable is used at 2 or more body positions. The other dimension of comparison was the consideration of confounding activities while predicting smoking activity. Apart from this study, only two other studies (Varkey et al., 2012 and Senyurek et al., 2019) have considered predicting smoking activity among confounding activities in their models. But, these two studies have high obtrusivity and predictive accuracy is significantly less than this study. The comparison in Table 9 over the chosen criteria indicates that the current study is best positioned for practical applications among the prominent baseline studies.

**Table 7** In-sample and out-of-sample validation of classification model developed in experiment-5 for smoking activity

Classification model	AUC (in-sample testing)	AUC (out-of-sample testing)
Adaboost	0.98051	0.97285
Decision Tree	1	0.96801
KNN	1	0.98115
Logistic Regression	0.97689	0.9711
SVM	0.99799	0.98652
Random Forest	1	0.98291

**Table 8** In-sample and out-of-sample validation of classification model developed in experiment-6 for smoking activity

Classification model	AUC (in-sample testing)	AUC (out-of-sample testing)
Adaboost	0.95987	0.95864
Decision Tree	1	0.96662
KNN	1	0.98428
Logistic Regression	0.96674	0.96349
SVM	0.99626	0.98619
Random Forest	1	0.98432

## 5 Preventive healthcare implications of the proposed modeling framework

The study was conducted to explore the practical applicability of the IMU sensor in real-time prediction the daily-living activities with a special focus on smoking. The application of IMU is practical in identifying the activities only if they are used unobtrusively. Literature suggests that use of smartwatches and wearable bands are increasing significantly for preventive healthcare purpose [48]. The existing wearable band or smartwatches are limited to predicting only physical activity like running, walking, etc. The objective of this study was to extend the features of such devices to predict more sophisticated activities like smoking and beyond. To make the application practical, the activity recognition should be in near real-time, using such methods and models which are lightweight, fast, and highly accurate. In this study, an IMU sensor-based generalized modeling framework is proposed which can help in delivering preventive healthcare more smartly. The efficacy of this framework is tested on the real-time predictions of daily living activities. The further applications of the proposed IMU sensor-based activity-recognition modeling framework in other healthcare settings include:

### 5.1 Ensuring conformance to the standard operating procedures (SOPs) in hospital settings

The proposed framework can be extended to develop models for predicting safety-critical activities in hospital settings e.g. frequent washing/sanitization of hands by healthcare professionals in the view of the prevailing COVID-19 pandemic.

### 5.2 Geriatric care management

Usually, youth are busy in their day-to-day activities and old people are staying at home. The proposed framework is going to provide a mechanism to monitor the well-being of old people staying alone at home [15, 30].

### 5.3 Management of chronic diseases

The level of daily physical activity of patients can be tracked and suitable intervention can be planned. This framework can also help in developing mechanisms to ensure adherence to

**Table 9** A comparative analysis of baseline studies with the current study based on pre-selected objective criteria's

Comparison criteria	Baseline studies comparison with this study					
	Varkey et al. (2012) [64]	Scholl et al. (2012) [55]	Tang et al. (2014) [59]	Saleheen et al. (2015) [53]	Senyurek et al. (2019) [57]	This study
Objective	To recognize a range of activities and fine movements within a specific activity.	To develop a basic smoking detection method.	To automatically detect puffing and smoking	To detect the first smoking lapse of abstinent smokers	To detect smoking events	To detect smoking activity among ADLs in real-time
Evaluation Metric	Accuracy	Precision, and Recall	F1-score	Recall	Accuracy, F1-score	AUC
Duration of data	n.a.	n.a.	11.8 h	40 h	1h30m – 2 h (lab) 24 h (free-living)	5h18m
No of Sensor	2	1	4	2	2	1
Body Positions of Sensors	Wrist, Ankle	Wrist	Wrist, arm, and ankle	Wrist, Respiration (body-suit)	Wrist, instrumented lighter	Wrist
Classifier/Model	SVM	Gaussian Mixture	Random forest	SVM	SVM	Ensemble
Performance	Overall: 97.2% Smoking: 94.4%	Precision: 51.2% Recall: 53.7%	Puffing: 70% Smoking: 79%	First lapse: 96.9%	Smoking: 98%	Overall: 92.9% Smoking: 98.7%
Obtusivity	High	Mild	Very high	Very high	High	Mild
Confounding activities	Yes	No	No	No	Yes	Yes
Real-time prediction	No	No	No	No	No	Yes

doctor's prescription by the patients suffering from chronic diseases like diabetes, cardiovascular disease, chronic obstructive pulmonary disease [62].

#### **5.4 Lifestyle management**

A suitable recommendation can be planned based on the activity log of the people being monitored. A recent study had enlisted various future work in their review research on physical activity recognition and monitoring (PARM) [44].

#### **5.5 Monitoring the activity of persons working in a hazardous environment to ensure safety**

Only authorized persons are allowed to work in a hazardous environment, there may be some inadvertent lapses in following safety protocols. This framework will help in monitoring the activity of workers in such an environment. The framework could also be extended to include various other suitable sensor devices and making a relevant prediction about the health and safety of such personnel.

### **6 Conclusion and future directions**

The main objective of this study was to develop a modeling framework to predict the smoking activity among the activities of daily-living in real-time. The main contributions of this study are:

#### **6.1 Development of a low-cost wrist wearable IMU sensor device-**

We first developed a wrist wearable device using a 6-axis IMU sensor. We used this device to collect data for activities like walking, running, walking up-stairs, walking down-stairs, and smoking in a free-living environment.

#### **6.2 Developed the methodology of pre-processing the IMU sensor data for the real-time applications-**

The sensor generates the data which is streaming in nature, so we proposed the utilization of a sliding-window mechanism to create features out of the streaming sensor data. This study is first of its kind which has studied the optimal size of the sliding window. Different size of sliding window has experimented and an optimal size was identified to be used for feature generation.

#### **6.3 Developed a generalized machine learning-based predictive modeling framework-**

A supervised machine learning-based modeling framework has been designed to build the classification models for the recognition of smoking and physical activities. The framework also performs the hyperparameter tuning and feature selection analysis to identify the best set of hyperparameters and important features. Subsequently, those hyperparameters and

important features are used to build the prediction models. The developed models are validated using in-sample and out of sample testing to assess their performance and the best model is chosen for the application. In this study, the SVM model achieved the best performance with an AUC of 98.78%.

Hence, the present work lays the foundation for future research of real-time detection of addictive activities using wearables and developing systems for just-in-time intervention to promote cessation of such addictive behaviors. The developed framework can also be used for different healthcare applications and use-cases. In this study, an attempt has been made to predict smoking activity among the activities of daily living to assess the practical applicability of the system. However, there are few limitations of this work despite the relative merits of the proposed model is compared to the models available in the literature.

This paper considered only a few physical activities apart from the smoking activity for classification. More activities like sleeping, eating, drinking, talking, and any other activity of interest can be incorporated in future research work. It will enable the real-life application of this model with more profoundness. The developed modeling framework can be deployed using a mobile application and its performance in a real-life scenario is yet to be assessed. Using the same mobile application, healthcare services can be designed to provide real-time interventions for smoking cessation. Further, a study can be conducted to investigate the efficacy of such interventions. The performance of supervised machine learning models also depends on the choice of the features which are hand-crafted and thus defines an upper bound on the performance of the machine learning models. Deep learning models can draw the features from raw data automatically. The performance of the real-time activity detection system could be further improved both in terms of predictive accuracy and prediction time using deep learning-based classification models like recurrent neural network (RNN), and LSTM. This study was limited by the smaller dataset whereas a deep learning model requires large data for training. Further experimentation with the large dataset could be done to test a deep learning model for the real-time prediction of the smoking activity. There are various other methods for human activity recognition (HAR) based on video data, and other environmental sensor data like infra-red sensors. It will be interesting to compare HAR methods based on wearable sensor devices versus video or environmental sensor-based methods. Further, the computational time complexity analysis should also be carried out to understand the suitability of these methods.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11042-022-12349-6>.

## References

1. Abrams LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J (2013) A content analysis of popular smartphone apps for smoking cessation. *Am J Prev Med* 45(6):732–736
2. Adibi S (2015) *Mobile health a technology road map*, 5th edn. Springer International Publishing
3. Akash K, Hu W-L, Jain N, Reid T (2018) A classification model for sensing human trust in machines using EEG and GSR. *ACM Trans Interact Intell Syst* 8(4):1–20
4. Altman NS (1992) An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat* 46(3): 175–185
5. Al-Ubaydli O, List JA, LoRe D, Suskind D (2017) Scaling for economists: lessons from the non-adherence problem in the medical literature. *Econ Perspect* 31(4):125–144
6. Aly M (2005) Survey on multiclass classification methods

7. Anguita D, Ghio A, Oneto L, Ridella S (2012) In-sample and out-of-sample model selection and error estimation for support vector machines. *IEEE Trans Neural Networks Learn Syst* 23(9):1390–1406
8. Atallah L, Lo B, King R, Yang G-Z (Aug. 2011) Sensor positioning for activity recognition using wearable accelerometers. *IEEE Trans Biomed Circuits Syst* 5(4):320–329
9. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
10. Burba F, Ferraty F, Vieu P (2009)  $K$ -nearest neighbour method in functional nonparametric regression. *J Nonparametr Stat* 21(4):453–469
11. Chatterjee S, Price A (2009) Healthy living with persuasive technologies: framework, issues, and challenges. *J Am Med Informatics Assoc* 16(2):171–178
12. Cox DR (1958) The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)* 20. WileyRoyal Statistical Society:215–242
13. Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines : and other kernel-based learning methods. Cambridge University Press
14. Erdaş B, Atasoy I, Açıci K, Oğul H (2016) Integrating features for accelerometer-based activity recognition. *Proc Comput Sci* 58:522–527
15. Eyobu OS, Kim YW, Cha D, Han DS (2018) A real-time sleeping position recognition system using IMU sensor motion data. In: 2018 IEEE International Conference on Consumer Electronics, ICCE 2018, vol 2018, pp 1–2
16. Fogg BJ (1999) Persuasive technologies. *Commun ACM* 42(5):26–29
17. Formagini TDB, Ervilha RR, Machado NM, de Andrade BABB, Gomide HP, Ronzani TM (2017) A review of smartphone apps for smoking cessation available in Portuguese. *Cad Saude Publica* 33(2)
18. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55(1):119–139
19. Freund Y, Schapire RE (1999) A short introduction to boosting. *J Japanese Soc Artif Intell* 14(5):771–780
20. Gao Z, Xuan HZ, Zhang H, Wan S, Choo KKR (2019) Adaptive fusion and category-level dictionary learning model for multiview human action recognition. *IEEE Internet Things J* 6(6):9280–9293
21. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1):29–36
22. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED (2017) A systematic review of smartphone applications for smoking cessation. *Translational Behavioral Medicine* 7(2) Springer New York LLC: 292–299
23. Heydari G et al (2014) A comparative study on tobacco cessation methods: A quantitative systematic review. *International Journal of Preventive Medicine* 5(6) Isfahan University of Medical Sciences:673–678
24. Heydari G et al (2015) Assessment of different quit smoking methods selected by patients in tobacco cessation centers in Iran. *Int. J. Prev. Med* 2015
25. Hoepfner BB, Hoepfner SS, Seaboyer L, Schick MR, Wu GWY, Bergman BG, Kelly JF (2016) How smart are smartphone apps for smoking cessation? A content analysis. *Nicotine Tob Res* 18(5):1025–1031
26. Huang WH, Hsu HY, Chang BCC, Chang FC (2018) Factors correlated with success rate of outpatient smoking cessation services in Taiwan. *Int J Environ Res Public Health* 15(6)
27. Jain R, Majumder P, Gupta T, Bandiera SM (2013) Pharmacological Intervention of Nicotine Dependence. *Biomed Res Int* 2013
28. Jha P, Peto R (2014) Global effects of smoking, of quitting, and of taxing tobacco. *N Engl J Med* 370(1): 60–68
29. Kim E, Lee J, Shin H, Yang H, Cho S, Nam SK, Song Y, Yoon JA, Kim JI (2019) Champion-challenger analysis for credit card fraud detection: hybrid ensemble and deep learning. *Expert Syst Appl* 128:214–224
30. Kosse NM, Brands K, Bauer JM, Hortobagyi T, Lamothe CJC (2013) Sensor technologies aiming at fall prevention in institutionalized old adults: A synthesis of current knowledge. *International Journal of Medical Informatics* 82(9) Elsevier:743–752
31. Luna-Perejon F, Malwade S, Styliadis C, Civit J, Cascado-Caballero D, Konstantinidis E, Abdul SS, Bamidis PD, Civit A, Li YC(J) (2019) Evaluation of user satisfaction and usability of a mobile app for smoking cessation. *Comput Methods Prog Biomed* 182:105042
32. McClure JB, Hartzler AL, Catz SL (2016) Design considerations for smoking cessation apps: feedback from nicotine dependence treatment providers and smokers. *JMIR mHealth uHealth* 4(1):e17
33. Méndez D, Tam J, Giovino GA, Tsodikov A, Warner KE (2016) Has Smoking Cessation Increased? An Examination of the US Adult Smoking Cessation Rate 1990–2014. *Nicotine Tob Res*:ntw239
34. Messer K, Trinidad DR, Al-Delaimy WK, Pierce JP (Feb. 2008) Smoking cessation rates in the United States: a comparison of young adult and older smokers. *Am J Public Health* 98(2):317–322
35. Miao F, He Y, Liu J, Li Y, Ayoola I (2015) Identifying typical physical activity on smartphone with varying positions and orientations. *Biomed. Eng. Online* 14(32)

36. O'Donoghue G, Cunningham C, Murphy F, Woods C, Aagaard-Hansen J (2014) Assessment and management of risk factors for the prevention of lifestyle-related disease: a cross-sectional survey of current activities, barriers and perceived training needs of primary care physiotherapists in the Republic of Ireland. *Physiotherapy* 100(2):116–122
37. Pärkkä J, Ermes M, Korpiä P, Mäntyjärvi J, Peltola J, Korhonen I (2006) Activity Classification Using Realistic Data From Wearable Sensors. *IEEE Trans. Inf. Technol. Biomed* 10(1)
38. Paulovich FV, De Oliveira MCF, Oliveira ON (2018) A future with ubiquitous sensing and intelligent systems. *ACS Sensors* 3(8):1433–1438
39. Pedregosa F et al (2011) Scikit-learn: machine learning in Python
40. Pedregosa F et al (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830
41. Powers DMW (2011) Evaluation: from Precision, Recall and f-measure to roc, informedness, markedness & correlation. *Int J Mach Learn Technol* 2(1):37–63
42. Preece SJ, Goulermas JY, Kenney LPJ, Howard D (2009) A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Trans Biomed Eng* 56(3):871–879
43. Pudil P, Novovičová J, Kittler J (Nov. 1994) Floating search methods in feature selection. *Pattern Recogn Lett* 15(11):1119–1125
44. Qi J, Yang P, Waraich A, Deng Z, Zhao Y, Yang Y (2018) Examining sensor-based physical activity recognition and monitoring for healthcare using Internet of Things: A systematic review. *Journal of Biomedical Informatics* 87. Academic Press Inc:138–153
45. Qi J, Yang P, Fan D, Deng Z A Survey of Physical Activity Monitoring and Assessment Using Internet of Things Technology. In: 2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, vol 2015, pp 2353–2358
46. Raiff BR, Karataş Ç, McClure EA, Pompili D, Walls TA (2014) Laboratory validation of inertial body sensors to detect cigarette smoking arm movements. *Electron.* 3(1):87–110
47. Raschka S (2018) MLxtend: providing machine learning and data science utilities and extensions to Python's scientific computing stack software • review • repository • archive
48. Reeder B, David A (2016) Health at hand: a systematic review of smart watch uses for health and wellness. *J Biomed Inform* 63:269–276
49. Regmi K, Kassim N, Ahmad N, Tuah NA (2017) Effectiveness of Mobile Apps for Smoking Cessation: A Review. *Tob. Prev. Cessat* 3(April)
50. Rose S (2018) Machine Learning for Prediction in Electronic Health Data. *JAMA Network Open* 1(4) NLM (Medline):e181404
51. Rossel PO, Paredes L, Bascur A, Martínez-Carrasco C, Herskovic V (2019) SAS4P: Providing automatic smoking detection for a persuasive smoking cessation application. *Int J Distrib Sens Networks* 15(11): 155014771988816
52. Safavian SR, Landgrebe D (1991) A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern* 21(3):660–674
53. Saleheen N et al (2015) puffMarker: A Multi-Sensor Approach for Pinpointing the Timing of First Lapse in Smoking Cessation. In: Proc. ... ACM Int. Conf. Ubiquitous Comput. . UbiComp, vol. 2015, pp 999–1010
54. Schick RS, Kelsey TW, Marston J, Samson K, Humphris GW (2018) MapMySmoke: feasibility of a new quit cigarette smoking mobile phone application using integrated geo-positioning technology, and motivational messaging within a primary care setting. *Pilot feasibility Stud* 4:19
55. Scholl PM, van Laerhoven K (2012) A Feasibility Study of Wrist-Worn Accelerometer Based Detection of Smoking Habits. In: 2012 Sixth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, pp 886–891
56. Schwartz RP, Gryczynski J, Mitchell SG, Gonzales A, Moseley A, Peterson TR, Ondersma SJ, O'Grady KE (2014) Computerized versus in-person brief intervention for drug misuse: a randomized clinical trial. *Addiction* 109(7):1091–1098
57. Senyurek V et al (2019) Cigarette smoking detection with an inertial sensor and a smart lighter. *Sensors* 19(3):570
58. Silva BMC, Rodrigues JJPC, de la Torre Diez I, Lopez-Coronado M, Saleem K (2015) Mobile-health: a review of current state in 2015. *J Biomed Inform* 56:265–272
59. Tang Q, Vidrine D, Crowder E, Intille S (2014) Automated Detection of Puffing and Smoking with Wrist Accelerometers. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare, pp 80–87
60. Thakur SS, Roy RB (2018) A mobile app based smoking cessation assistance using automated detection of smoking activity. In: Proceedings of the ACM India Joint International Conference on Data Science and Management of Data - CoDS-COMAD '18, pp 352–355

61. Thakur SS, Roy RB (2018) Smartphone-Based Ubiquitous Data Sensing and Analysis for Personalized Preventive Care: A Conceptual Framework. In: Verma NK, Ghosh AK (eds) *Advances in Intelligent Systems and Computing* vol. 798. Springer, Singapore, pp 119–132
62. A. Triantafyllidis, H. Kondylakis, K. Votis, D. Tzovaras, N. Maglaveras, and K. Rahimi, “Features, outcomes, and challenges in mobile health interventions for patients living with chronic diseases: A review of systematic reviews,” *International Journal of Medical Informatics*, vol. 132. Elsevier Ireland Ltd, p. 103984, 2019.
63. Ubhi HK, Kotz D, Michie S, van Schayck OCP, Sheard D, Selladurai A, West R (2016) Comparative analysis of smoking cessation smartphone applications available in 2012 versus 2014. *Addict Behav* 58: 175–181
64. Varkey JP, Pompili D, Walls TA (2012) Human motion recognition using a wireless sensor-based wearable system. *Pers Ubiquitous Comput* 16(7):897–910
65. Verbeek P-P (2012) *Persuasive Technology*. In: *Encyclopedia of Applied Ethics*. Elsevier, pp 431–437
66. Villalobos-Zúñiga G, Cherubini M (2020) Apps that motivate: a taxonomy of app features based on self-determination theory. *Int J Hum Comput Stud*:102449
67. Walker SH, Duncan DB (1967) Estimation of the Probability of an Event as a Function of Several Independent Variables. *Biometrika* 54(1/2):167
68. Wan S, Qi L, Xu X, Tong C, Gu Z (2019) Deep learning models for real-time human activity recognition with smartphones. *Mob Networks Appl* 2019 252 25(2):743–755
69. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y (2012) Mobile phone-based interventions for smoking cessation. In: *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd
70. Zheng YL, Ding XR, Poon CCY, Lo BPL, Zhang H, Zhou XL, Yang GZ, Zhao N, Zhang YT (2014) Unobtrusive sensing and wearable devices for health informatics. *IEEE Trans Biomed Eng* 61(5):1538–1554

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Saurabh Singh Thakur<sup>1</sup> · Pradeep Poddar<sup>2</sup> · Ram Babu Roy<sup>1</sup>

<sup>1</sup> Rajendra Mishra School of Engineering Entrepreneurship, Indian Institute of Technology, Kharagpur, India

<sup>2</sup> Department of Metallurgical and Materials Engineering, Indian Institute of Technology Kharagpur, Kharagpur, India