Smartphone-sensor-based human activities classification for forensics: a machine learning approach

Nchouwat Ndumgouo Ibrahim Moubarak[1*], Njutapmvoui Mbah Mohamed Omar[1] and Vepouyoum Njouokouo Youssef[1]

*Correspondence:
nchouwat@iut-dhaka.edu

[1] Department of Electrical and Electronic Engineering, Islamic University of Technology, Gazipur, Dhaka 1704, Bangladesh

Abstract

The accurate classification of human activities in crime scenes during forensics (criminalistics) is of utmost importance in classifying suspicious and unlawful activities, easing their acceptability and interpretability by judges during legal procedures in courts or by other non-experts in the field of forensics. This paper implements machine learning (ML) algorithms: support vector machine (SVM) and decision tree (DT), to demonstrate with a high accuracy, how data emanating from smartphones' sensors reveal and isolate relevant information about static and dynamic human activities in criminalistics. Smartphones' data from five different sensors (accelerometer, gravity, orientation, Gyroscope and light), related to ten recurrent crime scenes activities, grouped into three classes of events (normal, felony and none-felony events) are classified by the proposed algorithms, with novelty being the classification decisions based on the entire period of the events and not instantaneous decision makings. Three independent data-subsets were made, with permutations done between them and at each time, two sets used for training and the third set used for testing. Time- and frequency-domain features were initially used separately and then combined for the model training and testing. The best average training accuracies of 100% and 97.8% were obtained for the DT and SVM, respectively, and the testing accuracies of 89.1% were obtained for both algorithms. We therefore believe that these results will serve as a solid persuasive and convincing argument to judges and non-experts of the field of forensics to accept and easily interpret computer-aided classification of suspicious activities emanating from criminalistic studies.

Keywords: Machine learning, SVM, Decision tree, Classification, Human activities, Forensics

Introduction

In the present era, mobile phones are undoubtedly the most omnipresent devices being used by almost everyone. The latest generations of mobile devices provide intelligent help to users rather than just for information transfer [1]. Smartphones have features such as cameras, GPS, and web browsers and embedded sensors such as accelerometers, gyroscope and magnetometer used for human motional activity detection/recognition and measurements [2].

Moubarak *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:33

Page 2 of 16

Human activity recognition (HAR) with visual concept in Fig. 1 is the application of the data collected from sensors, usually embedded in smartphones to accurately recognize and classify the states of a person's activities, independent of external noises [1]. Smartphone data are used to reveal information contextually and situationally on different human activities. Most of the sensors are specific; hence, they are only applied in situations where they yield their best performances. Based on this, an accelerometer and gravity sensor will perform optimally in measuring acceleration and gravity, respectively. Prior to model training, data are initially collected by smartphone sensors and stored temporally in a database of a server. This is then followed by an approach based on learning features by codebook. This generates feature vectors by encoding data coming from the smartphone's sensors by parameter tuning [3] of the different groups of human activity recognition. Figure 1 is the concept diagram that illustrates the steps involved in the use of smartphones in HAR classification.

Human activity recognition (HAR) can be grouped into four major groups namely: group activities, real-time activities, individual activities and daily living activities. These activities can be categorized as static or dynamic in nature. Static activities account for 18% of the total HAR while dynamic activities account for 24%. When an individual performing the activity is fixed with respect to an observer, the activity is said to be static, but when the individual moves consistently, the activity is said to be dynamic [4].

The diversified nature of human activities makes them not readily recognizable, hence the necessity of applying machine learning algorithms to easily recognize and monitor them accurately [5]. Apart from sectors like business, healthcare, safety, transportation, where human activities detection and classification are becoming widespread, forensics is a potential field where smartphone data could be used. This is due to an increase of at least 20% in crimes after 2021 [6, 7]. Forensics is the application of scientific approaches, procedures or expertise to investigate or probe into evidence from crime scenes that can
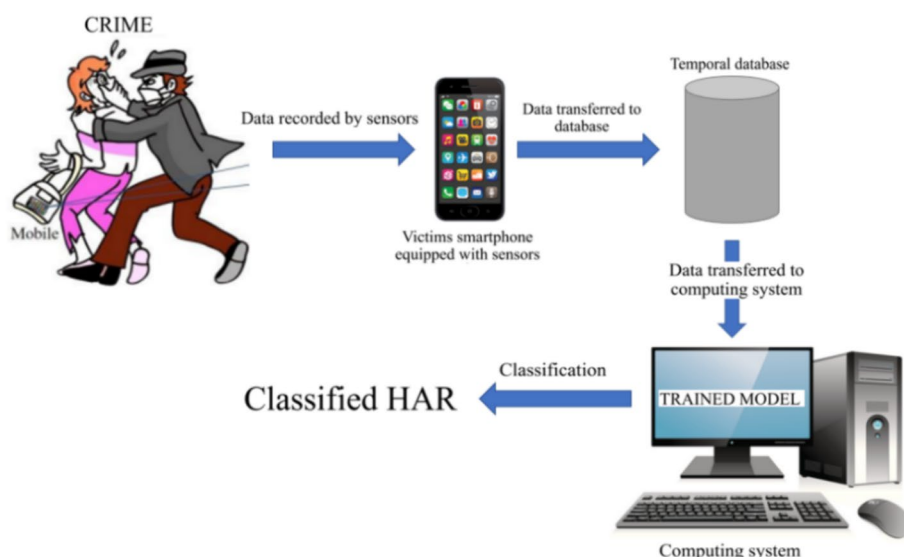
Fig. 1 The concept diagram for HAR. The figure shows an example of a crime scene where a woman is being attacked by an armed rubber. The smartphone having the sensors is placed in her handbag. Data are collected during the crime, stored and later used to predict the type of activity that was involved

be presented before a judge in a criminal procedure court [8]. Though some areas of criminalistics involve straight forward scientific procedures, many challenges remain and include the reliability and acceptability of the results obtained by a particular forensics method and how these results could be easily interpreted and understood by a judge or by a non-expert of the field.

Nowadays, companies, healthcare, education, transport and many other sectors store information electronically. This causes the e-discovery challenge to become expensive and complicated [9]. These have been causing tremendous problems and huge challenges to lawyers in large scale litigations as they entail the management of a huge volume of data (documents). The review of these legal documentations, the acceptance and interpretability of forensic results have been (and are) burdensome for these legal experts. To mitigate these problems, computer-aided techniques have been introduced to process these documents faster, with machine learning being the backbone of this novelty in the legal community [9]. Interpretation of stored smartphone sensor data, collected during the occurrence of a crime, could provide additional evidence needed for a successful prosecution.

The focus of this work is the development of a predictive machine learning model that will accurately classify the patterns of ten simulated crime scene scenarios based on HAR, grouped into three classes recognizable by criminal law. The goal is to precisely classify these human activities and behaviors related to crime scenes. The predictive models will be evaluated in terms of reliability, acceptability and their quick and easy interpretability by the judges and other non-experts. This is because the application of ML in the legal community has caught the attention of several researchers in the past decade, with significant state of the art advances done in the area of human activities and behavior classification. This has eased some of the challenges faced by experts in the legal community.

Related works

Table 1 summarizes the closest works of machine learning related to this paper. Other related works include (1) a comprehensive review study done by Muhammad et al. [4], (2) the state-of-the-art results obtained by Saponas et al. [8] and Moco et al. [10], focusing on the motion-related activities like standing, running and walking. They use data from a smartphone, emanating from a set of users within a particular time frame. (3)

Table 1 Related works

Work	Dataset					Algorithm	Accuracy (%)
	Activity	Number of classes	Number of sensor data	Features domain	Size per class		
Köping et al. [3]	Static/ dynamic	11	8	Time	9 s data at 1500 time points	SVM	87.1
Sukhada et al. [6]	Static	2 (binary classification)	10	Time	5994	1. Modified-KNN	99
						2. Subspace discriminate	83.4
Saini et al. [24]	Dynamic	24	Not mentioned	Time	2220–6660	(BLSTM-NN)	68.9

Khan et al. [11], used human motion-based data from mobile sensors (accelerometer, gyroscope and magnetometer) to distinguish between "normal" and "brick" walking, where brick walking refers to lame or dangling walking patterns. (4) Mylonas et al. [12], summarized the use of smartphone data in forensic and their legal implications. (5) Keeling et al. [9] introduced the predictive coding or technology assisted review (TAR) approach that gives an insight into how machine learning could help ease the burden on lawyers. The research showed that no machine learning algorithm was superior to another in the legal field by undertaking 34,000 experiments in view of searching for the best machine learning algorithm and (6) Jahangiri et al. [13] implemented KNN and SVM to identify transportation mode by using mobile sensor data. Most of these works do not focus on real-time smartphone mobile sensor data during the pursuit of a crime, are based on instantaneous classification decision making and do not focus on the criminal laws criteria of crime events, hence the necessity of this work to portray real HAR events in criminalistics.

Sukhada et al. [6] is the closest work to this paper related to HAR in digital forensics. The work focused on binary classification (suspicious or non-suspicious activities) and used a black-box technique in features extraction. Time-domain features were used to train a KNN and Subspace discriminate model for the classification of activities in digital forensics. This paper takes into account the static and dynamic natures of multiple activities in classification, involving ten different types of activities.

Novelty

The contribution of this work is the accurate classification of real-time patterns for human activities on crime scenes based on ten different crime scenes activities grouped into three classes as per judiciary considerations, with classification decision taken based on the entire period of the activities and not time instances of the activities. This is the first paper focusing on this procedure. The association of an ensemble of static and dynamic activities as well as their intensities are considered in contrast to [7, 8]. The latter are focused solely on motion-based activities (walking, climbing, sitting, etc.). A combination of data from five different sensors (accelerometers, gyroscope, gravity, light intensity and orientation) are used in contrast to only three sensor data used by the previous works. The accurate real-time static and dynamic human activity recognition and classification, which help to precisely materialize a crime scene behavior during forensics is achieved. These complement the works done in [1–8, 10–12] which are solely focused on static activities like slow movement, standing still and falling.

Protocol specification

In order to mature the novelty of this work as mentioned above, the following questions were to be answered;

- Question one (Q1); Can several crime scene activities be mimicked, for data emanating from them to be recorded by smartphone sensors, that would be used to train predictive models?
- Question two (Q2): Can data from static and dynamic crime scene activities be used simultaneously to train a predictive model?

- Question three (Q3): Can these crime scene activities be successfully classified using several sensor data?
- Question four (Q4): Does the application of time and frequency-domain features improve the classification accuracy of these activities?

The attempts made to answer these questions are elaborated in the rest of this paper adopting the work flow chart in Fig. 2. For the rest of the paper, (Q1), (Q2), (Q3) or (Q4) will be mentioned where the corresponding indexed question is addressed and a comprehensive pipeline of the work done will be describe for its easy replication or use, in the classification of activities with different datasets.

Paper organization

The organization of this paper is nurtured as follows: "Introduction" section exposes an insight into the recent trends made on the classification of human activities using data from smartphone sensors and a nutshell description of forensic science. "Materials and methods" section gives an overview of the proposed models. It details the materials and the methods applied in this work. This is followed by "Results and discussions" section, that presents the results obtained and a comprehensive discussion levied on the results. "Conclusion and perspectives" section rounds up this work with a conclusion encompassing the approach by which this work could be improved in future research.
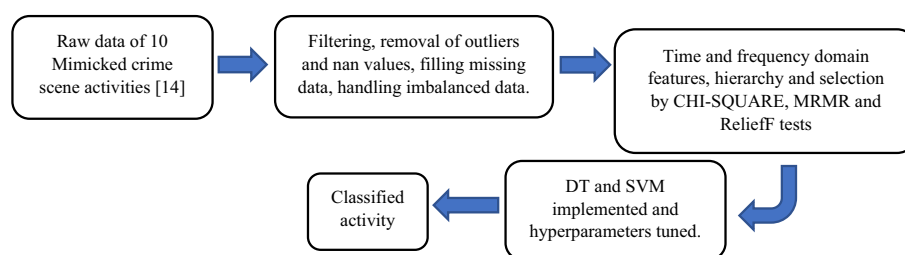
Materials and methods

Data description

To obtain a genuine outcome in this project, a privately collected dataset from [14] was used. The data were recorded from five sensors in a smartphone (Q3). These sensors were used to separately record five different parameters related to the behavior of a subject in a mimicked crime scene (Q1). The parameters are acceleration, gravity, orientation, data from the gyroscope and light intensity. These parameters were collected for ten activities, with each activity corresponding to data from a child or an adult, performing different activities during the day or recreating the scenario of the particular crime scene.

Data recording

Two or three independent repetitions were conducted for each activity. This was aimed at varying the data sets, to prevent any eventual model overfitting and to make

Fig. 2 The flow chart of the work done

Moubarak *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:33

Page 6 of 16

the trained model robust. We later grouped them into three classes. These classes were classified in accordance with the criminal law in [15] to help judges take quick decisions. Class1 can be viewed as an escape from a "felony" attack, class2 as a normal activity and class3 as a "felony" attack (the most serious crimes that can be lethal). The recording of the data was done through the "Sensor Logger" App using a VIVO Y50 (VIVO 1935) phone, version 10, a processor of 2 GHz and a RAM of 8 GB. The sampling rate for all the sensors is 100 Hz, except for the light sensor that was collected without sampling rate. Accelerations were recorded at ± 6 g and the gyroscope calibrated at ± 2000 deg/s. Segmentation was done with an optimized window size of 40 ms. This window size did not lead to the reduction of the size of the dataset and ensured the capturing of signals resulting from micro activities like ECG. To avoid data loss at the edges, a 50% overlap was implemented as in [15]. The data for the five signals were recorded for approximately 10.5 s per activity. Each activity was realized by mimicking the behavior of a real crime scene and by firmly attaching the smartphone on the arm of the subject and the classification decision was done at the end of 10.5 s corresponding to the entire duration of the activity being considered as per this work. Table 2 shows the different activities and the classes constituted from the activities, the number of repetitions made during recording and their classification as static or dynamic (Q2). The table also shows how the different repetitions of the activities were distributed between the subsets to be used for training and testing of the proposed models. Two of these subsets will be used for training and the other will be left out and used for testing. Permutation was done between the subsets and cross-validation done. The "x" on the subsets means no data entry for that set corresponding to the concerned activity. This distribution strategy was adopted to reduce the effect of data imbalance between the sets.

Table 2 Classes of activities and their descriptions

Number	Activity	Description	Number of repetitions	Distribution			Class
				Set 1	Set 2	Set 3	
1	A teenager running	Dynamic	3	1	1	1	Running away from a felony attack
2	A man hastily starting a car	Dynamic	2	1	x	1	
3	A woman running	Dynamic	2	1	1	x	
4	A Women sitting in moving vehicle	Static	3	1	1	1	Sitting (normal event)
5	A man sitting in speedy car	Static	3	1	1	1	
6	Women calling for help during the kidnapping	Dynamic	2	1	1	x	Felony attack
7	A man being dragged into the car by kidnappers	Dynamic	3	1	1	1	
8	Kids fighting	Dynamic	3	1	1	1	
9	A woman forcefully pushed in the back seat of a vehicle by a man entering the vehicle	Dynamic	2	1	x	1	
10	Woman struggling against attack	Dynamic	2	X	1	1	

Data preprocessing

The data used in this work were raw in nature and hence might be corrupted by artifacts. These artifacts could be noise, degradation of the signal (that leads to missing values), imperfection of in measurement due to the degradation of the smartphone sensors, etc. These artifacts are not part of the patterns of the recorded signals and if not removed, will cause the overfitting of the model. This is because the algorithm will be finding a target function that will fit all the data points (including the noise). Hence the resulting function will work well with the training data but will fail miserably with the test data. Therefore, the conditioning of the data was primordial prior to the training of the algorithm to classify simulated human activities by the machine learning algorithm proposed in this work. MATLAB-2022(a) was used for both data preprocessing and feature computation. The classification task was performed using MATLAB. To ensure the high performance of the proposed algorithm, the artifacts that corrupted the data were handled by; (1) linear interpolation, to replace missing data, (2) filing the imbalance of the observations in activities 2 and 5 with data from previous observations as in [16] and distributing the activities between the classes as shown on Table 2. This was to make the observations from all the activities same. The data were initially plotted to pinpoint the outliers, the latter that were simply deleted from each class of data and which constituted less than 0.1% of the data. Low frequency artifacts from usual body motions and possible noise emanating from the accelerometer were attenuated by the implementation of a second order Butterworth band pass filter, with cutoff-frequencies of 0.05 Hz and 0.2 kHz [17]. This frequency range was taken in order not to miss any contribution of ECG or EEG signals that may be generated during the recording of the human activity data. A notch filter centered at 60 Hz was implemented to filter out any noise coming from the power source and processing devices.

Features extraction

The features considered in this research are time- and frequency-domain features (Q4) as presented in Table 3. Prior to the features' computation, segmentation of the data into smaller segments to reduce computation time was done [18]. A sliding window technique was used with an optimized window size of 40 ms selected for the time-domain feature extraction. For each of the 40 ms window size, the following procedure was adopted;

Table 3 Features extracted

Sensors		Features per window		Total features
Name	No. of axes	Time domain	Frequency domain	
Acceleration	3	Mean, standard deviation, variance, harmonic mean, zero-crossing rate, kurtosis, skewness, rms, sum of the absolute values, mean absolute value (10)	Median, sum of absolute values, energy spectral density, total energy, harmonic mean, RMS values, minimum value of spectral coefficients, maximum value of spectral coefficients (8)	170 + 136
Gravity	3			
Gyroscope	3			
Orientation	7			
Light	1			
Total	17	(10*17) = 170	(8*17) = 136	306

Moubarak *et al. Journal of Electrical Systems and Inf Technol* (2024) 11:33

Page 8 of 16

1. Ten time-domain features were extracted, and
2. Eight frequency-domain features extracted based on the FFT.

This procedure was implemented to prevent foreign frequency components in another window from appearing in the current window being considered. A 50% window overlap was implemented for both time and frequency domains in computing the features. This was to prevent signal being lost at the edges of the windows. The mean, standard deviation, variance, harmonic mean, zero-crossing rate, kurtosis, skewness, RMS, sum of the absolute values and mean absolute value were computed for each axis of the sensors in the time domain while the median, sum of absolute values, energy spectral density, total energy, harmonic mean and RMS values [19–21] were calculated for the frequency domain. The minimum value and maximum value of the spectra coefficient in frequency domain were also analyzed. The window size and overlap chosen for feature computation did not lead to a significant reduction in the row sizes of the features, so the whole length of 10.5 s were used for the data in both training and testing. This can be understood in the sense that an activity must be initiated and completed (or run for at least a considerable duration, as per this work) for it to be accurately classified. So, using instantaneous or point data to classify an activity will not reflect the reality.

Properly normalizing (standardizing) and selecting features prior to the training of any model will help the training to converge faster and prevent the overfitting of the trained model. Based on these, the features were normalized by the Z-norm technique. ($Z = (v - m)/std$, where Z is the normalized datapoint, v is the raw datapoint, and m and std are the mean and standard deviation of the feature data, respectively). This step was to convert the data points to a range between 0 and 1(This reduced the computation time and to made the algorithm converge sooner), to solve the problem of biasing of the data and to prevent the overfitting of the trained model (this may result if the hypothesis function is simpler than the target function).Due to the complexity and the highly uncorrelated nature of the features, statistical approaches using CHI_SQUARE, MRMR and ReliefF techniques were used for feature hierarchy. The best twenty time and frequency-domain features were selected to train the SVM and DT while the best forty features were used when the combination of time and frequency domains were considered. These features were the subsets of features intersecting in all the three statistical tests.

Implementation of the algorithm

Classification algorithms have been found to yield high accuracies in human activities detection [19, 21]. Relying on this, a multi-sensor classification of human activities is performed in this work (Q3). To accomplish this task, the decision tree and support vector machine classification algorithms are implemented. Classification accuracy and N-way ANOVA test (Table 5) evaluation techniques were used in this work. The former is a type of evaluation technique associating both precision and recall and is very appropriate in evaluating human activity classifications [22]. Two of the three repetitions from each class were used as training sets and the other one as testing set. For validation between the repetitions, permutations were done between the sets for training and testing as shown on Table 2 and the average accuracy computed as shown on Table 4. Classification decisions were made

Table 4 (a) Training and testing accuracies with time-domain features, (b) training and testing accuracies with frequency-domain features, (c) training and testing accuracies with time–frequency-domain features

Experiment No	Time-domain feature sets					
	Training accuracies			Testing accuracies		
	Training sets	SVM	Decision tree	Testing set	SVM	Decision tree
(a)						
1	1 and 2	79.8	100	3	78.9	100
2	1 and 3	76.9	96.8	2	76.8	100
3	2 and 3	100	100	1	67.3	67.3
Average		85.5	98.9		74.3	89.1
Experiment No	**Frequency-domain feature sets**					
	Training accuracies			Testing accuracies		
	Training sets	SVM	Decision tree	Testing set	SVM	Decision tree
(b)						
1	1 and 2	100	100	3	89.1	100
2	1 and 3	96.8	100	2	100	100
3	2 and 3	100	100	1	67.3	67.3
Average		98.9	100		85.4	89.1
Experiment No	**Time–frequency-domain feature sets**					
	Training accuracies			Testing accuracies		
	Training sets	SVM	Decision tree	Testing set	SVM	Decision tree
(c)						
1	1 and 2	95.7	100	3	100	100
2	1 and 3	97.8	100	2	100	100
3	2 and 3	100	100	1	67.3	67.3
Average		97.8	100		89.1	89.1

Table 5 ANOVA test results

Source of variation	Sum of squares	Degree of freedom	Mean square	F-score (*F*)	Probability
Columns pf of 40 features	5541	40	615.71	69.76	9.23038 Exp (− 109)
Error	14,298.8	1620	8.826		
Total	19,840.2	1629			

at the end of 10.5 s (the entire duration of the activity). The accurate performances of the proposed models and the reduction of overfitting were ensured by applying a tenfold cross-validation on each of the repetitions used for training [23]. This has been found to be the accurate method of preventing the overfitting of the model.

Results and discussion
Results
Having addressed the four questions that we set as the basis of this work; the following subsection elaborates the outcome that was achieved. Initially, the performances of the

sensors for (acceleration, gyroscope, light, orientation, gravity) were analyzed separately. The models were trained by 20 best feature columns out of the total 306 original features columns. The emanating average accuracy after training the time–frequency features with SVM is 97.8% (which is the sum of the accurately classified activities in this work), with the testing accuracy of 89.1%. Table 5 shows the ANOVA test results. It can clearly be seen that the probability value on the table is too small (close to zero). This legitimates the rejection of the null hypothesis, which states that the results obtained may be by chance. The F value indicates clearly that the between group variation of the final features are very high compared to the within group variation hence ensuring the effective separation of the classes. To ensure that the accuracy obtained by the SVM is unbiased, the decision tree algorithm was also implemented with training and testing accuracies of 100% and 89.1 obtained, respectively. The confusion matrixes of these results are shown in Fig. 3. Two misclassifications are observed on class 3 being classified as class 2 (with positive predictive value of 90% and false discovery rate of 10%), and all the other classifications are correctly predicted (PPV of 100% and FDR of 0%).

Separate results obtained with the time-domain features, frequency-domain features and the combination of features of both domains are presented in Table 4a–c, respectively. Statistical tests as proposed in [24–27] for twenty best intersecting features from CHI_SQUARE, MRMR and ReliefF tests were considered for all the domains except for the combination of both domains where forty best features were considered. For each feature domain two sets of data from each class were used for training and one for testing with a tenfold cross-validation and optimization by grid search. The decision tree shows consistencies in training and testing accuracies with all the feature domains in contrary to the SVM as can be seen on Table 4, showing that the decision tree is a better algorithm for this work. The training and testing accuracies slightly higher with the combination of the time- and frequency-domain features (Q4) than when the domains are considered separately. The confusion matrices in Fig. 3 present the classification matrices for both classifiers.

Analyses
The objectives of this work were anchored on examining the following;

1. Whether several crime scene activities could be mimicked, for data emanating from them to be collected by mobile sensor, that would be used to train predictive models.
2. Whether data from static and dynamic crime scene activities could be used simultaneously to train a predictive model.
3. Whether these crime scene activities could be successfully classified using more than three sensor data without lost in accuracy and
4. Whether the application of time- and frequency-domain features would improve the classification accuracy of these activities.

These objectives were addresses, respectively, as follows;

1. Three classes constituted made from ten activities have been successfully mimicked, and the smartphone mobile sensor data collected from them have been successfully

Moubarak *et al. Journal of Electrical Systems and Inf Technol*   (2024) 11:33

Page 11 of 16
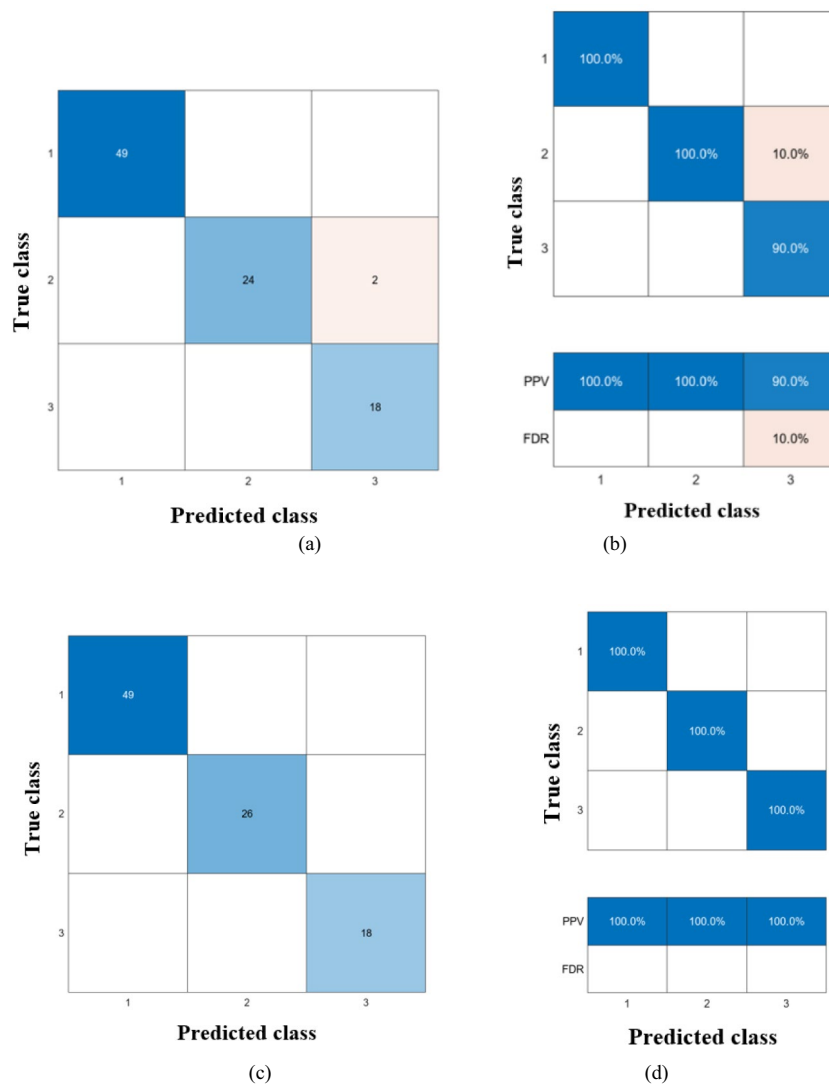
Fig. 3 Confusion matrixes of the decision tree and SVM classification results from the using the time–frequency features with sets 1 and 3 as training sets and set 2 as testing set, **a**, **c** Classification of observations for the DT and SVM, respectively, **b**, **d** the positive predictive values and false discovery rates for the DT and SVM, respectively

used to train two predictive models. This serves as a genuine argument for the veracity of this work as the results are comparable to others done with real-time data in [4]

2. As elaborated in Table 1, two out of the ten activities were static. This had little bias impact on the accuracy because the bias was largely mitigated by the distribution technique adopted between the three classes. This increased the variability of the data.

3. Contrarily to most of the related works, five different sensor data have been used in this work for the model training with and acceptable accuracy.

4. Combining the time- and frequency-domain features slightly increases the average accuracies of the models as shown in Table 4c.

A comparison of the models proposed in this work was made with related works as shown on Table 1. Based on the results on the table, it can be argued that:

1. The accuracy obtained in this work is lower than that obtained by Sukhada et al. [6] since the work focuses solely on binary classification (suspicious or non-suspicious activities) while this work took into account the static and dynamic natures of multiple activities in classification, involving ten activities grouped into three classes.
2. The accuracy obtained is clearly higher than all the other accuracies, even though Köping et al. [3] used up to eight different sensor data to classify five activities.
3. Overall, we verified the four claims that were set for this work by addressing them simultaneously, which were not done by the works stated for comparison as shown on Table 1.

The rule of thumb was to test the proposed model with data used in the literature mentioned on the comparison table especially that in [6]. But the existence of the disparities in the number of classes, the number of sensor data as well as the activity types classified, did not agree nor constitute a common framework for it to be done. A possible future work will be to extend the binary classification technique mentioned in [6] to multi-class classification.

Parameter tuning

In order to obtain the optimal accuracy results presented above, the following parameter tunings were done:

1. *Segmentation window size* The window sizes of 20 ms, 30 ms, 40 ms, 50 ms and 60 ms were considered. The subsets of features retained from these windows were trained at similar conditions. The best accuracies were obtained at 40 ms window size in the time domain. The corresponding Fourier coefficients in the frequency domain were then considered. Figure 4a depicts both the training and testing accuracies against the different window sizes, considering sets 1 and 3 as training sets and set 2 as testing set. Underfitting occurred at window sizes of 20 ms and 60 ms while overfitting occurs at window sizes of 30 ms, 40 ms and 50 ms. 40 ms was chosen due to the fact that it has higher accuracies, and the values were closer together.
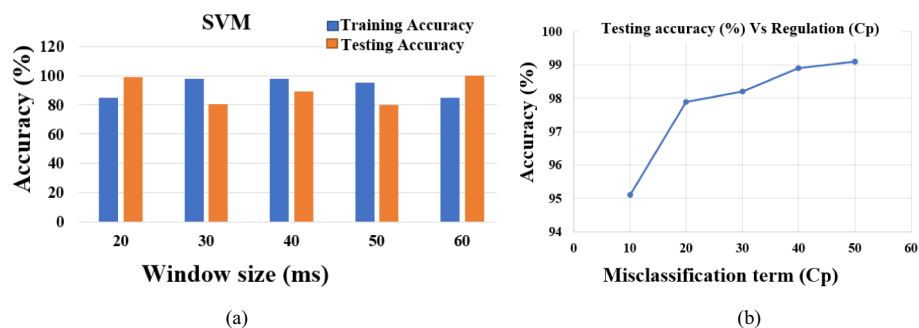
Fig. 4 Optimization of the SVM; **a** Plot of training and testing accuracies against window size and **b** Plot of testing accuracy against the error term

2. *Hyperparameters* The optimized values for the hyper parameters of the DT and SVM models were done by the grid-search method with ten divisions at the window size of 40 ms. These hyper parameters are listed in Table 6.

3. To analyze and confirm the regularization (Cp) effects on the accuracy as depicted in Fig. 4b, all the other parameters were kept constant and Cp was varied. Cp is a very important hyperparameter due to the fact that it is the one that gives information on misclassification and model fitting (overfitting/underfitting). This Cp value was then varied from 10 to 50 at a step of 5. The value of 20 was found for the error (misclassification) term because the accuracy it yielded converged with the training accuracy and gave a good trade off at the decision boundary while the other values were rejected because they were prone to overfitting

Evaluation of performance and robustness

We evaluated the performance and robustness of the proposed model by a fair comparison with the results from Sukhada et al. [6]. This is because it is the closest work dealing with HAR in forensics. We used their dataset as independent data to evaluate and validate our results. We computed the frequency-domain features of their data. Only one experiment was done in time domain, frequency domain and time–frequency domain because the data was not recorded in triplicate (the case of our data), Slide degradation in accuracies were observed. These could be because of the fundamental difference in the nature of their dataset with ours. Table 7 presents the results obtained.

Discussion

As a grand premier, several static and dynamic real-life crime scenes for human behaviors and activities have been mimicked and accurately classified with good accuracy. By the time this work was done, and to the best of our knowledge, it was the only work known to have been done using this dataset. The data emanated from five different sensors and the features used for classification were both time- and frequency-domain features. The aim of this research was focused on the experiments based on the four

Table 6 (a) Tuned hyperparameters for the decision tree, (b) tuned hyper parameters for the SVM

Hyperparameter	Value
(a)	
Tree	Fine tree
Maximum depth	3
Maximum leaf nodes	30
Number of splits	100
Split type	Gini's diversity index
(b)	
Kernel	Rbf
Gamma	30
Constrain level	1
Multi-class	One to One
Regularization (Cp)	20

Table 7 (a) Evaluation training and testing accuracies with time-domain features, (b) evaluation training and testing accuracies with frequency-domain features, (c) evaluation training and testing accuracies with time–frequency-domain features

Experiment No	Time-domain feature sets	
	Training accuracies	Testing accuracies
	SVM	SVM
(a)		
1	85.0	73
Experiment No	**Frequency-domain feature sets**	
	Training accuracies	Testing accuracies
	SVM	SVM
(b)		
1	97.1	84.7
Experiment No	**Time/frequency-domain feature sets**	
	Training accuracies	Testing accuracies
	SVM	SVM
(c)		
1	97.2	88.3

questions set at the opening of this paper. These questions have been addressed with the use of a dataset on which no prior paper has been published to date. With this dataset, a better accuracy was obtained compared to the closest work to this one done by Sukhada et al. [6]. The paper focused mainly on binary classification of suspicious against non-suspicious activities for digital forensics.

Nevertheless, the accuracy obtained needs improvement. This accuracy would have been higher but for the fact that, (1) We did not collect the data by ourselves to be sure of its genuineness at 100%. (2) Using the feature selection methods in the time domain as analyzed in [28, 29] would have enhanced the results obtained in the time domain. (3) The effect of the segmentation window size [30] on the accuracy was analyzed only in the time domain while this effect was assumed for the frequency domain, (4) the smartphone was placed on the arm of the subject under test during data collection, which is not the optimal positioning according to the studies undertaken by Nweke et al. [31]. They showed that the optimal positioning of the phone is on the chest. (5) The imbalance of the different classes of the dataset also contributed to the biased classification of the classes with higher observations (repetitions) to the detriment of the ones with smaller observations, (6) No information on the adversity of the weather was mentioned, since the data were from mimicked activities and (7) It was noticed that the data from all the sensors were not synchronized in time. A better performance would have been achieved if the synchronization was done before being used for classification as suggested by Jahangiri et al. [13]. These factors might have compromised the quality of the data by one way or the other, hence not properly reflecting the reality. Attempts to handle these problems constitute the basis of future work on this dataset.

Conclusion and perspectives

Ten recurrent human activities in suspicious or crime scenes have been classified in this work. SVM and decision tree algorithms have been implemented with average accuracies of 97.8% and 100% obtained with SVM and DT, respectively, for training and 89.1% for both algorithms in testing. A complete pipeline for the addressing of our four objectives has been met (but not fully for objective Q2 due to imbalanced data) associated with a trained model encompassing a complete model selection and regularization, feature hierarchy and selection. The accuracies obtained have complimented the works mentioned in the literature of this work, most of which are based on bagging and majority voting. The trained models could be used in forensics science to accurately classify the activities that could have been done by a suspect or done on a victim. This accuracy would be a convincing factor for machine-aided human activity classification to be accepted in legal procedures by Judges. Prior to its improvement, this work will serve as a starting point in the analysis of a complex dataset emanating from HAR and the training of more complex models for classification. An extension of this work would be to place smartphones on several locations of the body of the subjects during activity mimicking and to use data from all of them (and synchronize them) to have more precise and unbiased data. A multi-sensor stacking ensemble algorithm could then be used in model training to obtain better results.

Declarations

Competing interests
The authors declare that they have no competing interests.

References

1. Liu Q, Zhou Z, Shakya SR, Uduthalapally P, Qiao M, Sung AH (2018) Smartphone sensor-based activity recognition by using machine learning and deep learning algorithms. Int J Mach Learn Comput 8(2):121-126
2. Lane ND, Miluzzo E, Lu H, Peebles D, Choudhury T, Campbell AT (2010) A survey ofmobile phone sensing. IEEE Commun Mag 48:140–150
3. Köping L, Shirahama K, Grzegorzek M (2018) A general framework for sensor-based human activity recognition. Comput Biol Med 95:248–260
4. Arshad MH, Bilal M, Gani A (2022) Human activity recognition: review, taxonomy and open challenges. Sensors-MDPI 22:6463
5. Brezmes T, Gorricho JL, Cotrina J (2009) Activity recognition from accelerometer data on mobile phones. In: IWANN '09: proceedings of the 10th international work conference on artificial neural networks, pp 796–799
6. Aloni S, Shekhawat D (2022) Detection of suspicious activity using mobile sensor Data and Modified Sub-space K-NN for criminal investigations. In: YMER-ISSN 0044-0477, vol 21, no 8
7. Gyorbiro N, Fabian A, Homanyi G (2009) An activity recognition system for mobile phones. In: MONET

8. Saponas TS, Lester J, Froehlich JE et al (2008) Ilearn on the Iphone: real-time human activity classification on commodity mobile phones. CSE Technical Report
9. Robert K, Rishi C, Nathaniel H, Zhang J, Zhao H (2020) Using machine learning on legal matters: paying attention to the data behind the curtain. Law J Lib 11:9–37
10. Moco NF (2014) Mobile forensics: a smartphone-based activity logger. In: 21st International conference on telecommunications (ICT). https://doi.org/10.1109/ict.2014.6845159
11. Khan YA, Imaduddin S, Prabhat R, Wajid M (2022) Classification of human motion activities using mobile phone sensors and deep learning model. In: 8th International conference on advanced computing and communication systems (ICACCS), vol 1, pp 1381–1386. IEEE
12. Mylonas A, Meletiadis V, Tsoumas B, Mitrou L, Gritzalis D (2012) Smartphone forensics: a proactive investigation scheme for evidence acquisition. In: IFIP international information security conference, pp 249–260. Springer
13. Jahangiri A, Rakha HA (2015) Applying machine learning techniques to transportation mode recognition using mobile phone sensor data. In: IEEE transactions on intelligent transportation systems, 2406/2417
14. https://ieee-dataport.org/documents/mobile-sensor-dataset-human-activity-classification-used-forensics
15. Nandy A, Saha J, Chowdhury C (2020) Novel features for intensive human activity recognition based on wearable and smartphone sensors. Microsyst Technol 26:1889-1903
16. Ordonez FJ, Roggen D (2016) Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors (Basel) 16:115
17. Imtiaz MH, Senyurek VY, Belsare P, Tiffany S, Sazonov E (2019) Objective detection of cigarette smoking from physiological sensor signals. IEEE
18. Leutheuser H, Schuldhaus D, Eskofier BM (2013) Hierarchical, multi-sensor-based classification of daily life activities: comparison with state-of-the-art algorithms using a benchmark dataset. PLoS One 8:e75196
19. Shoaib M, Bosch S, Incel OD, Scholten H, Havinga PJ (2016) Complex human activity recognition using smartphone and wrist-worn motion sensors. Sensors (Basel) 16:426
20. Figo D, Diniz PC, Ferreira DR, Cardoso JM (2010) Preprocessing techniques for context recognition from accelerometer data. Pers Ubiquitous Comput 14:645–662
21. Zdravevski E, Lameski P, Trajkovik V, Kulakov A, Chorbev I, Goleva R et al (2017) Improving activity recognition accuracy in ambient-assisted living systems by automated feature engineering. IEEE Access 5:5262–5280
22. Chowdhury AK, Tjondronegoro D, Chandran V, Trost SG (2017) Ensemble methods for classification of physical activities from wrist accelerometry. Med Sci Sports Exerc 49:1965
23. Garcia-Ceja E, Galván-Tejada CE, Brena R (2018) Multiview stacking for activity recognition with sound and accelerometer data. Inf Fusion 40:45–56
24. Saini R, Kumar P, Roy P, Dogra DP (2018) A novel framework of continuous human-activity recognition using Kinect. Neurocomputing 311:99–111
25. Rahman AA, Siraji MI, Khalid LI, Faisal F, Nishat MM, Islam MR, Moubarak NN (2022) Detection of mental state from EEG signal data: an investigation with machine learning classifiers. In: 14th International conference on KST, pp 152–156. IEEE
26. Apon HJ, Abid MS, Morshed KA, Nishat MM, Faisal F, Moubarak NNI (2021) Power system harmonics estimation using hybrid Archimedes optimization algorithm-based least square method. In: 13th International conference on ICTS, pp 312–317. IEEE
27. Kabir MR, Muhaimin MM, Mahir MA, Nishat MM, Faisal F, Moubarak NNI (2021) Procuring MFCCs from Crema-D dataset for sentiment analysis using deep learning models with hyperparameter tuning. In: International Conference on RAAICON, pp 50–55. IEEE
28. Igene L, Alim A, Imtiaz MH, Schuckers S (2023) A machine learning model for early prediction of Parkinson's disease from wearable sensors. In: 13th Annual computing and communication workshop and conference (CCWC). IEEE
29. Islam R, Sultana A, Islam MR (2024) A comprehensive review for chronic disease prediction using machine learning algorithms. J Electr Syst Inf Technol 11:27. https://doi.org/10.1186/s43067-024-00150-4
30. Mohammed CN, Ahmed AM (2024) A semantic-based model with a hybrid feature engineering process for accurate spam detection. J Electr Syst Inf Technol 11:26. https://doi.org/10.1186/s43067-024-00151-3
31. Nweke HF, Teh YW, Alo UR, Mujtaba G (2018) Analysis of multi-sensor fusion for mobile and wearable sensor based human activity recognition. In: ICDPA

Publisher's Note