



Stress Detection via Keyboard Typing Behaviors by Using Smartphone Sensors and Machine Learning Techniques

Ensar Arif Sağbaş¹ · Serdar Korukoglu¹ · Serkan Balli²

Received: 7 November 2019 / Accepted: 23 January 2020 / Published online: 17 February 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Stress is one of the biggest problems in modern society. It may not be possible for people to perceive if they are under high stress or not. It is important to detect stress early and unobtrusively. In this context, stress detection can be considered as a classification problem. In this study, it was investigated the effects of stress by using accelerometer and gyroscope sensor data of the writing behavior on a smartphone touchscreen panel. For this purpose, smartphone data including two states (stress and calm) were collected from 46 participants. The obtained sensor signals were divided into 5, 10 and 15 s interval windows to create three different data sets and 112 different features were defined from the raw data. To obtain more effective feature subsets, these features were ranked by using Gain Ratio feature selection algorithm. Afterwards, writing behaviors were classified by C4.5 Decision Trees, Bayesian Networks and k-Nearest Neighbor methods. As a result of the experiments, 74.26%, 67.86%, and 87.56% accuracy classification results were obtained respectively.

Keywords Stress detection · Smartphone · Accelerometer · Gyroscope · Machine learning · Classification

Introduction

Stress is a mental state in which everyone experiences in their daily lives [1]. It is a rescue mechanism of the body at critical periods. However, after exceeding a specific level, stress is no longer healthy. Contrarily, it begins to harm an individual's health, emotional state, productivity and quality of life. If the individual becomes highly stressed, this can cause serious health problems [2]. However, there are several difficulties in monitoring stress. Gjoreski et al. [3] identified three topics

that make it difficult to monitor stress of the people. The first one is that stress is a subjective condition. While a stimulus triggers stress in one person, it may not in another person. The second one is the difficulty of defining the ground truth. For this reason, monitoring of physiological data is performed or the self-assessment method is used. The last one is that stress can not be directly monitored. While physiological data can be monitored directly by sensors, behavioral and affective data can not be directly monitored.

Over the past two decades, researchers have found an important relationship between an individual's physical health and emotional state [4]. They use the physical and physiological symptoms of a person to detect existing emotions. Feelings such as happy, angry, fear, sadness, disgust and surprise are the most important basic emotional states [5]. Stress has been added to this set of basic feelings that can be detected recently. Stress is an important problem in modern society. Early detection of stress reduces damage and precludes it from becoming ingrained. The harms of stress on human health are known by researchers, and a considerable effort has been made recently to develop an automated stress detection system using smart devices and various computational algorithms. Automated stress detection systems can be used in various areas such as vehicle drivers, workplaces, passengers with the phobia, and patients [2].

This article is part of the Topical Collection on *Image & Signal Processing*

✉ Serkan Balli
serkan@mu.edu.tr

Ensar Arif Sağbaş
ensar.arif.sagbas@ege.edu.tr

Serdar Korukoglu
serdar.korukoglu@ege.edu.tr

¹ Faculty of Engineering, Department of Computer Engineering, Ege University, İzmir 35100, Turkey

² Faculty of Technology, Department of Information Systems Engineering, Muğla Sıtkı Koçman University, Muğla 48000, Turkey

In this study, by using the data of smartphone's touchscreen panel, accelerometer and gyroscope sensors, identification of whether the user stressful or not is realized through the writing behaviors on the smartphone's keyboard. To the best of authors' knowledge, this is the first stress detection study performed by examining with the machine learning methods of smartphone's motion sensors and keyboard usage behaviors. In this respect, this research would lead to new studies in this field. In addition, new and original data sets will be gained into the literature and new feature vectors will be extracted.

In the second part of the study, the previous studies will be mentioned. The created data sets, feature extraction and selection stages and smartphone sensors will be explained in Chapter 3. In chapter 4, the machine learning methods used will be mentioned and the experimental results will be discussed. Finally, the study will be concluded in Chapter 5.

Related work

When the literature is examined, it is found that stress detection studies use various data sources. The first one of these data sources is physiological data. Gjoreski et al. [6] proposed a 3-stage stress detection method with a wrist-wearing device. Minguillon et al. [7] proposed a portable system for real-time stress detection based on multiple bio-signals. Gjoreski et al. [3] developed a method for stress detection that can accurately, consistently and unobtrusively monitor psychological stress in real life. Padmaja et al. [8] presented an effective method for determining stress levels using data from a physical activity monitor. Pandey [9] used the heart rate as one of the parameters to predict stress. Choi et al. [10] developed a wearable sensor platform to monitor a series of physiological correlations of mental stress. Zenonos et al. [11] focused on working environments and explored the possibility of using smartphones and wearable devices for mood recognition. Mozos et al. [12] realized machine learning based stress detection by using physiological and social response information. Egilmez et al. [13] analyzed the effects of different body sensing platforms and their wrist-worn systems in stress estimation. Navea et al. [14] proposed a method for determining stress when the person is in mobile communication by using the galvanic skin response sensor.

The second one of used data sources in stress detection studies is behavioral data. Sysoev et al. [15] aimed to determine the level of stress to the greatest extent possible by analyzing behavioral and contextual data with only a smartphone. Lu et al. [16] proposed StressSense application to recognize stress from human voice by using smartphones. Wang et al. [17] evaluated the effects of workload on stress, sleep, activity, mood, socialization, mental health, and academic performance of students in a

single class using an Android phone with StudentLife application. Bogomolov et al. [18] proposed an alternative approach to recognize daily stress by the user's cell phone activity, additional indicators such as weather conditions, and behavioral measures derived from personality traits. Bauer and Lokowicz [19] investigated whether the differences between stressful and stress-free periods can easily be found in the information on a smartphone, such as Bluetooth devices and phone calls seen during the day. Cho et al. [20] proposed DeepBreath, a deep learning model that automatically recognizes people's psychological stress levels from breathing habits. Han et al. [21] proposed a psychological stress perception algorithm based on deep learning using speech signals. Kostopoulos et al. [22] presented a system aimed at detecting stress by analyzing users' behavior with their smartphones. The system offered by Gimpel et al. [23] used 36 hardware and software sensors to detect perceived stress levels of users. In the study conducted by Raichur et al. [24], videos were taken without interfering with the real-time user and it analyzed the facial expression and determined a person's emotional state. Vildjionaite et al. [25] proposed a new method of unsupervised stress detection using only smartphone data.

In addition, there are studies in which physiological and behavioral data are used together. Maier et al. [26] described the development of a mobile solution based on smartphones and sensors for early recognition of stress. Their solution was based on real-time capture and analysis of vital data such as heart rate variability, as well as analysis of contextual data such as activity, location and time. Muaremi et al. [27] provided a solution to assess people's stress experiences using features derived from smartphones and wearable chest belt. Sano and Picard [28] aimed to find physiological and behavioral markers for stress.

This study deals with stress detection through keyboard dynamics. There are various stress detection applications related to keyboard dynamics in the literature. Kim and Choi [29] investigated human behavior related to the touch interface on a smartphone as a way to understand users' emotional states. Lee et al. [30] aimed to recognize the user's emotions by unobtrusively collecting and analyzing user-generated data from different types of sensors on the smartphone. In the study conducted by Gao et al. [31], during an iPod game, finger-strokes were extracted and distinctive forces were analyzed. Lau [32] focused to detect stress using dynamics based on keystrokes (analysis of a user's writing rhythms) and to detect changes in stress with these rhythms. Ghosh et al. [33] worked on automatic emotion detection by modeling spelling characteristics and permanence of emotions together. Ghosh et al. [34] used the text input scheme to monitor multiple emotions. Exposito et al. [35] investigated the use of pressure sensing when typing to detect the stress of smartphone users.

Materials and methods

Data collection

Sample data were collected from the motion sensors and touchscreen panel to determine the stressful or non-stressful situation according to the behavior of the smartphone keyboard typing. The motion sensors used in the study and their brief descriptions are as follows:

Accelerometer This sensor measures the acceleration applied to the device. The accelerometer reports the acceleration values of the X, Y and Z axes of the device shown in Fig. 1a. This value is expressed in G. G is equal to the gravitational force exerted by the gravitational field ($9.81 \text{ m} / \text{s}^2$). The calculated values include the gravitational force in addition to the linear acceleration of the device [36].

Gyroscope The gyroscope sensor gives the angular velocity of the smartphone on the X, Y and Z axes. Axes trajectories are shown in Fig. 1b. The raw data collected from the gyroscope sensor reports the rotation of the smartphone about the three physical axes in rad/s (radians/s) [36, 37].

For this purpose, an application that can work on mobile phones with Android operating system was developed. This application consists of 4 stages. These stages are:

1. Data collection phase for non-stressful state (CALM)
2. Stressor task
3. Data collection in case of stress (STRESS)
4. Ground truth survey

In the 1st and 3rd stages of the mobile application, sensor data were collected. The developed application is set to collect 20 samples per second. The collected data is stored in the internal memory of the smartphone in a CSV format. The information gaining process is shown in Fig. 2.

Fig. 1 Accelerometer and Gyroscope sensor axes [38]

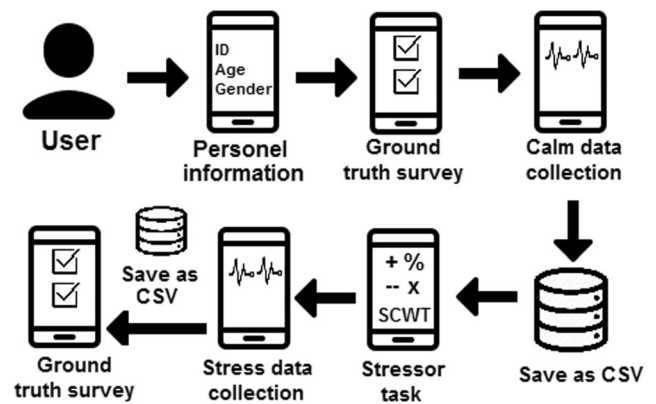
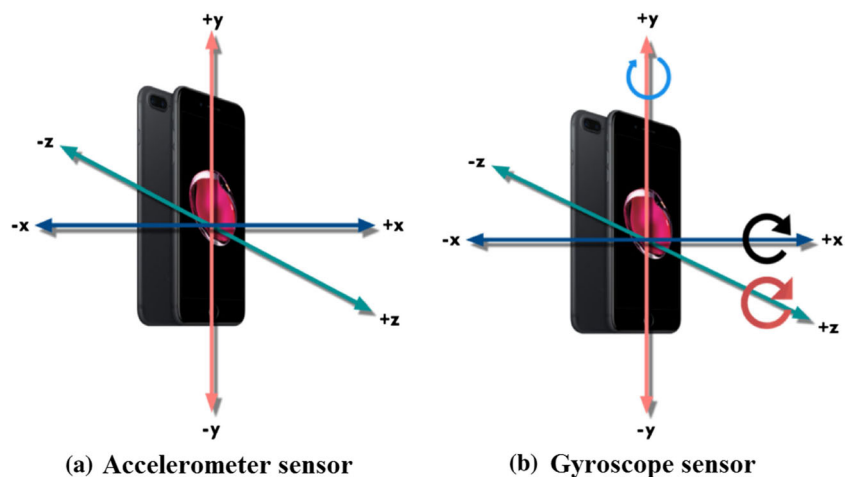


Fig. 2 Data collection stages

Before proceeding to the data collection phase, the user is recorded with a nickname or a number given by the tester and age and gender information are taken. Screenshot of this information is shown in Fig. 3.

At the beginning of the data collection phase, the participant is asked to write only the desired texts using the smartphone’s keyboard. The participant does not know that stress-related information is collected until the test is completed.

Non-stressful state (CALM) data collection phase

At this stage, the participants are asked to write the text displayed on the screen without any time limitation. Each text has a particular writing period, but the user is not informed about this period. 5 different texts of various lengths, including the letters on the keyboard, are entered into the system by the participants. Meanwhile, the smartphone accelerometer and gyroscope sensors collect data at 20 samples per second and store the data in the internal memory of the phone with the label “calm”. The screenshot of collecting non-stressful data is shown in Fig. 4.

As soon as the user begins to write the text, the sensors are activated. If the entire sentence is entered correctly or the

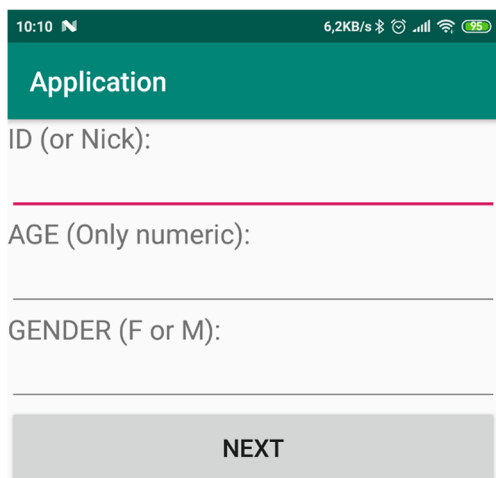


Fig. 3 Screenshot where participant information is saved

time’s up, the keyboard automatically closes and the next sentence is displayed.

Stressor task

Specific tasks are applied to the participant to measure the differences in interactions between non-stressful and stressful states. Several articles in the literature address the problem of triggering stress in humans [39, 40]. The main stressor methods used by Ciman et al. [1] are listed as follows:

- **Cognitive stress factor:** Memory and mathematical tests, e.g. beginning from a large prime or odd number and falling 13, 7 or 17 with mind calculation.
- **Social pressure:** Evaluation of an individual’s performance, especially by an external person, for example, public speaking.

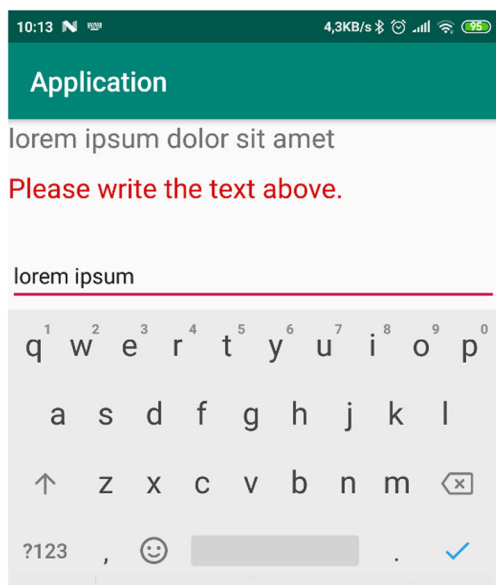


Fig. 4 Screenshot of non-stressful data collection

- **Timing pressure:** Give a time limit to realize the mission.
- **Random events:** Creation of random events that may disturb the main user’s mission, i.e. unexpected results, simulation of errors, and so on.

A stage with all the stressors mentioned above is added to the application. At this stage, the participants are asked to perform mathematical and arithmetic operations from the mind. It is desirable to subtract odd numbers such as 21, 13, 7 from a random 4-digit number and enter the result obtained. Each result found correct is +1 and each result found incorrect is -1 point. An annoying sound is played if the answer is wrong. The participant is asked to reach 7 points within 60 s (Cognitive stress factor & Timing pressure). The screenshot of the mobile application for the stressor task is presented in Fig. 5.

Then, the Stroop Color-Word Test (SCWT) is then performed. SCWT is one of the most commonly used and oldest stress induction tests [41]. Different variations of the test are established by changing the number of sub-tasks, number and type of stimuli, scoring procedures or task times [42]. In the regular version of the test, the participant is asked to read the name of the colored words. This sub-task is named “word reading”. Following sub-task is to name the color of the ink [2].

At the SCWT stage of the study, color texts in different color fonts that change every 2 s are asked to confirm the font colors by pressing the color buttons on the screen. When the time given to the participant expires, the second phase of the SCWT begins. At this stage, the participant has to confirm the written color on the screen by pressing the buttons. The

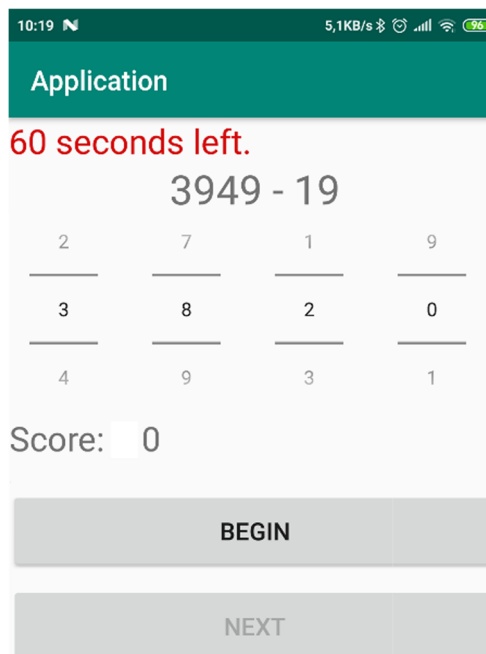


Fig. 5 Screenshot of stressor task

participants are asked to score more than 25 points in total. The highest total score is 30 (Timing pressure & Random events). An example screenshot of the SCWT test is shown in Fig. 6.

Data collection in case of stress (STRESS)

In the third and final stage of the data collection, participants are asked to write the texts they wrote during the “calm” phase under extra timing pressure. Although there is a time limit for participants to write the texts during the “calm” stage, the participants are not aware of this. At this stage, the time remaining in red color is shown just above the text to attract the attention of the participants (Fig. 7). In addition to the time pressure, when this stage is started, tension music starts to play in the background. Furthermore, what the user wrote is followed by the people around her/him (Social pressure). In case of mistakes, the participant is notified about the errors. At this stage, the developed application records the information with the “stress” label.

Ground truth survey

The stress level of the participant before and after the experiment is obtained by self-assessment. The questionnaire in Fig. 8 is presented to the participant in order to express how she/he felt before the experiment started and at the end of the experiment. Participants report their mental status with this 5-point Likert scale [43]. Five stars for all emotions states that the feeling is very intense.

The scores obtained as a result of the surveys are recorded in the internal memory of the smartphone. The first thing analyzed is that the stressor task achieves its goal. In other words, it increases the perceived stress of the participants.

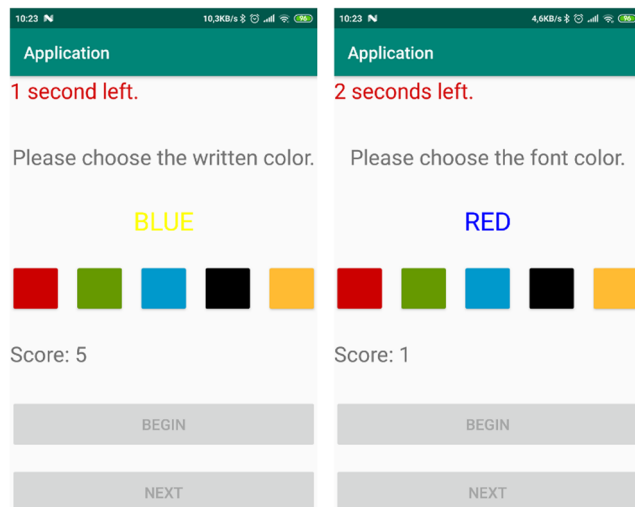


Fig. 6 SCWT test sample

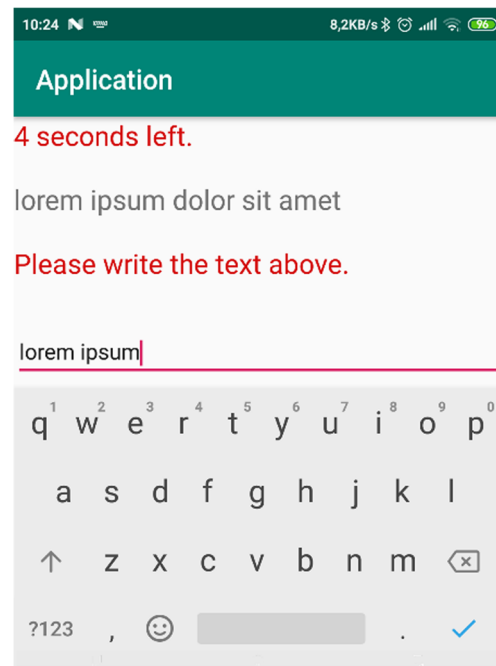


Fig. 7 Stressfulness data collection screen

In addition, it is investigated whether there is a significant difference between the work done in calm and stress situations. Participants are required to complete a questionnaire at the beginning and end of the application. The average values and standard deviations of all participants are reported in Table 1.

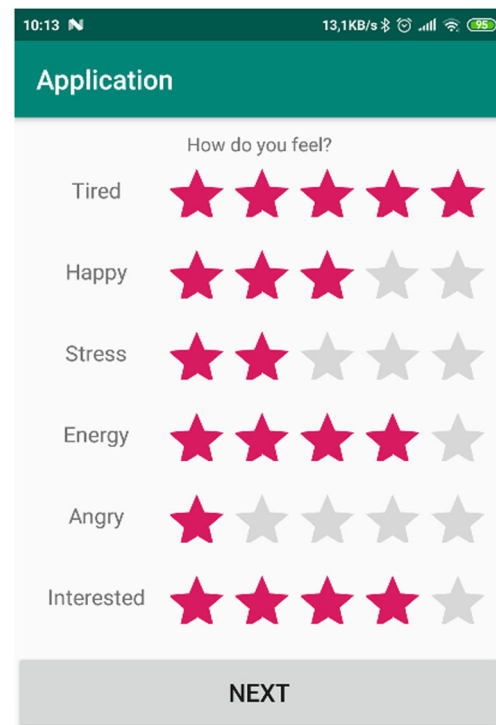


Fig. 8 The questionnaire presented to the participant

Table 1 Average values \pm standard deviations of the questionnaire presented to the participants at the beginning and at the end of the application

	Tired	Happy	Stress	Energy	Angry	Interested
At the beginning of the application	2.50 \pm 1.46	3.50 \pm 1.41	1.88 \pm 0.78	3.68 \pm 1.15	1.56 \pm 0.70	3.44 \pm 1.22
At the end of the application	2.69 \pm 1.21	2.94 \pm 1.30	3.38 \pm 1.05	3.62 \pm 1.17	2.62 \pm 1.41	3.31 \pm 1.53

Participants

An Android application is developed to collect data. If users' own phones have the desired sensors to collect data, it is provided to use their own phones. For users who do not have a phone with an Android operating system or whose phone is not eligible for data collection, one of the smartphones Samsung Android S6, Xiaomi Mi A2 Lite or Xiaomi Redmi Note 4 is provided. Orientation is limited in portrait mode. In addition, phone holding habits (such as one-handed or two-handed holding) are interfered because they are part of the writing behavior. However, since the new generation phones have a large and widescreen, it is observed that none of the participants held the phone with one hand. However, although very rare, the presence of participants using a single finger keyboard is noteworthy. The phones' default typing keyboards are used and automatic corrections and suggestions are turned off. For the purpose of preventing bias, participants are not stated that data is collected for stress detection. When the data collection process is completed, the purpose of the study is explained in detail. Information on the study population is presented in Table 2 and the age distribution of the participants is shown in Fig. 9.

Creating dataset

The data collected within the scope of the study are labeled with the developed mobile application and the self-evaluation applied to the participants. Raw data from all participants are combined in a single file. Then, this information is divided into 5, 10 and 15 s interval windows and three different data sets are created. Since the data collection mobile application is set up to collect 20 samples of data per second, the data compiled into a single file is set to be 100, 200 and 300 ($5 \times 20 = 100$, $10 \times 20 = 200$, $15 \times 20 = 300$) common multiple, and the excess data is cleared. In the remainder of the study, the data sets are expressed as DS-A (5 s), DS-B (10 s), and DS-C

(15 s). The numerical information about the data sets is presented in Table 3.

mean: number of.

Feature extraction

In addition to the smartphone sensors during the data collection phase, the number of times the user touched the screen and the number of times the backspace key pressed while writing the text are also recorded. The number of taps and deletions on the screen are added to the data sets as features. In addition, the age and gender of the participants are used as features.

For the data obtained from the sensor signals, 14 statistical measures are used in the feature extraction stage. In addition, a total of 18 features are extracted with zero crossings, mean energy, mean curve length, and mean teager energy. Brief descriptions of these features are given in Table 4.

Gain ratio feature selection algorithm

Gain Ratio is a different version of Information Gain (IG) [44] that reduces bias [45]. Gain Ratio aims to prevent the increase in the number of nodes. This is important when data is spread evenly and small when all data on a single branch [46].

A decision tree is a simple structure in which non-terminal nodes symbolize tests on one or more features and the terminal nodes reflect the final results. IG measurement is used to choose the test feature in each node of the decision tree and prefers to select attributes with multiple values. The decision tree uses information acquisition, known as the rate of gain, to overcome bias [47]. And it also uses the rate of gain that applies normalization for the acquisition of information using a value defined as in Eq. 19.

$$SplitInfo_A(S) = - \sum_{i=1}^v (|S_i|/|S|) \log_2(|S_i|/|S|) \quad (19)$$

Table 2 Information on the study population

Variable	Value
Population size	46
Age (minimum, maximum, average, standard deviation)	18, 39, 24.35, 6.10
Number of female participants	11
Number of male participants	35

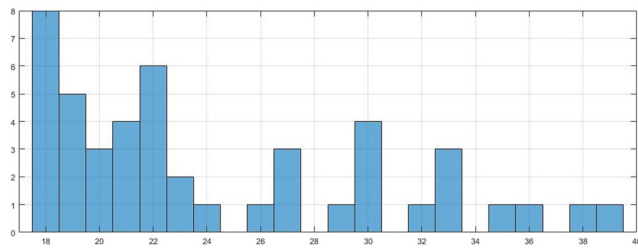


Fig. 9 Age distribution of participants

Gain ratio subordinates knowledge gain values to division information and normalizes them [48]. The gain ratio is defined as shown in Eq. 20.

$$Gain\ Ratio(A) = Gain(A) / SplitInfo_A(S) \tag{20}$$

Gain ratio feature selection algorithm is applied for each data set and the ranks of the features are calculated. New feature subsets are created using ranked features.

Tenfold cross validation

Ten-fold cross-validation technique is used to evaluate the classifier models. Cross-validation provides that each folding accurately represents class values on each fold. In this way, it helps to decrease variance in the prediction. Throughout this process, the samples are randomly divided into 10 (equal size) sub-samples. One of 10 sub-sample is held for testing and the remaining 9 sub-samples are used for training. This procedure is repeated 10 times in total. In this case, each of the 10 sub-samples is used one time as verification data. Then, obtained 10 results are averaged and a sole estimate is generated. This has the advantage that all samples are used for both validation and training [49, 50].

Experimental results and discussion

In the study, obtained data from smartphone’s touchscreen and motion sensors are evaluated with k-nearest neighbors (kNN), Bayesian Networks (BN) and C4.5 Decision Trees methods. Weka toolkit [51] and Java programming language are used. When comparing the outputs of the models, various performance evaluation measures are taken. These are classification performance including Precision, Recall, Classification accuracy (CA), Root Mean Square Error (RMSE), and F-Measure. These are given in Eq. 21–25, respectively. Where CE is the

Table 3 Numerical information of data sets

Dataset	# Calm	# Stress	Total
DS-A	3866	3171	7037
DS-B	1888	1519	3407
DS-C	1218	969	2187

number of correctly estimated samples and N is the total number of samples.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{21}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{22}$$

$$CA = \frac{100 * CE}{N} \tag{23}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - x_{true}}{x_{true}} \right)^2} \tag{24}$$

$$F\text{-Measure} = \frac{2 * Precision * Recall}{Precision + Recall} \tag{25}$$

Experiments

When the related works are examined, it is seen that stress detection is handled by various machine learning methods. In this study, kNN, Decision Trees and Bayesian Networks methods which are commonly seen in the literature are applied.

kNN

kNN is a supervised classification algorithm used in the classification process. This algorithm works with the logic of deciding which class a new sample belongs to by looking at a neighbor as many k as defined by the user [52]. While measuring the proximity value in this study, criteria such as Manhattan, Euclidean and Minkowski are used. Euclidean distance criterion is chosen as the most commonly used Euclidean distance criterion [38]. Furthermore, the problem is tested with different k neighborhood values and the most appropriate k value is found to be 1. Visualization of the features that have the highest three ranks for DS-C is given in Fig. 10. Figure 10 shows that the data is decomposed to some extent, even in only three dimensions.

Bayesian networks (BN)

The variables of the BN nodes are graphical models used to decide under uncertainty in which the arrows represent the connection between these nodes. In a Bayesian network, nodes represent random variables obtained from the environment and are connected to each other by directional arrows. These arrows represent the dependency between nodes. The strength of the connection between two discrete nodes is measured by the conditional probability between those two nodes. Firstly, the interest between the variables is determined. This

Table 4 Brief description of the features

Feature name	Formula	
Minimum value (MinV)	$MinV = \min[x_i], i = 1, \dots, n$	(1)
Maximum value (MaxV)	$MaxV = \max[x_i], i = 1, \dots, n$	(2)
Standard deviation (S)	$S = \sqrt{\sum_{i=1}^n \frac{(x_i - AM)^2}{n-1}}$	(3)
Arithmetic mean (AM)	$AM = \frac{1}{n} \sum_{i=1}^n x_i$	(4)
Absolute arithmetic mean (AAM)	$AAM = AM $	(5)
Geometric mean (GM)	$GM = \sqrt[n]{x_1 * x_2 * \dots * x_n}$	(6)
Harmonic mean (HM)	$HM = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$	(7)
Sum	$Sum = x_1 + x_2 + \dots + x_n$	(8)
Q ₁	$Q_1 = x_{(n)[0.25]}$	(9)
Median	$Median = x_{(n)[0.5]}$	(10)
Q ₃	$Q_3 = x_{(n)[0.75]}$	(11)
Variance (S ²)	$S^2 = \sum_{i=1}^n \frac{(x_i - AM)^2}{n-1}$	(12)
Skewness (SK)	$SK = \sqrt{\frac{\sum_{i=1}^n \frac{(x_i - AM)^3}{(n-1) * S^3}}{\sum_{i=1}^n \frac{(x_i - AM)^2}{(n-1) * S^2}}}$	(13)
Kurtosis (K)	$K = \sqrt{\frac{\sum_{i=1}^n \frac{(x_i - AM)^4}{(n-1) * S^4}}{\sum_{i=1}^n \frac{(x_i - AM)^2}{(n-1) * S^2}}}$	(14)
Zero Crossings (ZC)	$x_{i-1} < 0 \text{ ve } x_i > 0 \text{ OR } x_{i-1} > 0 \text{ ve } x_i < 0 \text{ OR } x_{i-1} \neq 0 \text{ ve } x_i = 0$	(15)
Mean Energy (ME)	$ME = \frac{1}{n} \sum_{i=1}^n x_i^2$	(16)
Mean Curve Length (MCL)	$MCL = \frac{1}{n} \sum_{i=2}^n x_i - x_{i-1} $	(17)
Mean Teager Energy (MTE)	$MTE = \frac{1}{n} \sum_{i=3}^n (x_{i-1}^2 - x_i x_{i-2})$	(18)

determines what the nodes represent and which values they take. The topology or structure of the network deals with the qualitative relationship between variables. In particular, if two nodes affect each other or cause one to occur, these two nodes are shown in direct connection [53, 54]. One of the most useful applications of the Bayesian rule is the Naive Bayes

classifier. The Naive Bayes model can be represented as a Bayesian network that encodes conditional independence between attributes and the class variable. In this study, a Bayesian network structure with independent attributes was built as seen in Fig. 11. In Fig. 11, ZC represents the zero-crossing values, MCL represents the mean curve length values, MTE represents the mean teager energy values and finally Std. Dev. represents the standard deviation values. The title of the boxes indicates the source sensor of the values.

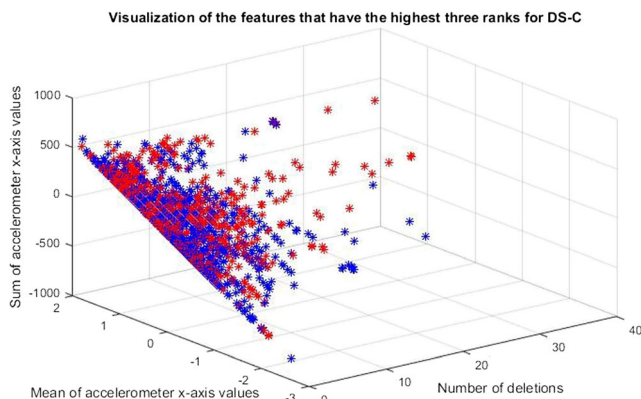


Fig. 10 Visualization of the features that have the highest three ranks for DS-C

C4.5

The C4.5 algorithm constructs the decision tree from a training set using the concept of knowledge entropy. The training set is classified as examples of $S = S_1, S_2, \dots, S_n$ data. Each S_j sample consists of a p-dimensional vector $(X_{1,i}, X_{2,i}, \dots, X_{p,i})$. Here X_j represents the properties or properties of the examples. On each node of the tree, C4.5 selects the attribute of the enriched data in the subsets. The division criterion is the normalized information gain. The attribute with the highest normalized information gain is chosen for the decision. The C4.5 algorithm is then

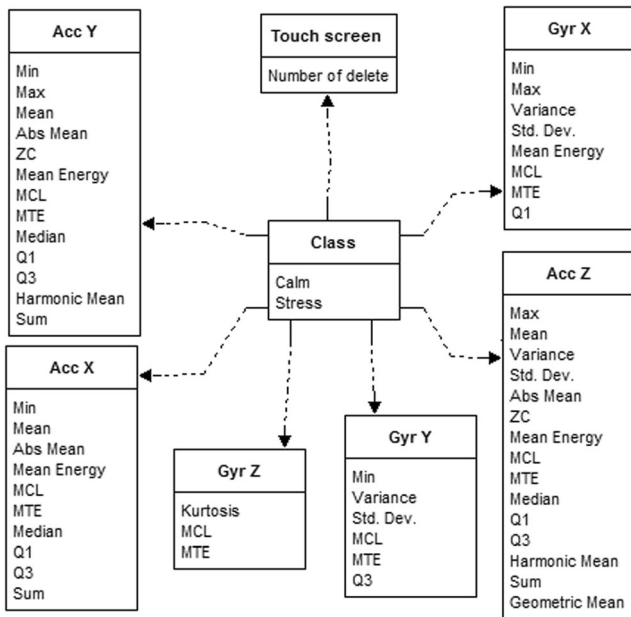


Fig. 11 Bayesian network structure

withdrawn from smaller sub-lists [55, 56]. In the structure, the median of the tree is 140 and the number of leaves is 279.

Results

In the study, three different classification procedures are applied with raw sensor data obtained from the smartphone. Whether the user of the smartphone is stressful or not is revealed with high precision. Variation in classification accuracies for three data sets are shown in Fig. 12.

When Fig. 12 is examined, DS-B and DS-C classification achievements are similar. However, the DS-A with a window interval of 5 s is insufficient for stress detection. The highest accuracy is obtained from the DS-C by using kNN method. To achieve this success, 43 features with the highest ranks are used. The most successful classification results obtained by data sets are presented in Table 5.

When Table 5 is examined, it is seen that the most successful results are obtained with DS-C with window intervals of 15 s. The most successful of the three machine learning methods used was kNN with 87.56% accuracy rate. This method is followed by C4.5 with 74.26% and BN with 67.86% accuracy rates. Similar to accuracy; the highest Recall, Precision and F-Measure values are also obtained from kNN method. As a trustworthy evaluation parameter in machine learning, RMSE is a measure of the distinctions between estimated and observed values. When the RMSE value approaches zero, there is an increase in the prediction capability of the model [57]. RMSE values are calculated as 0.54, 0.45 and 0.31 for BN, C4.5 and kNN methods, respectively. The confusion matrices of the most successful results obtained from the three methods are presented in Table 6.

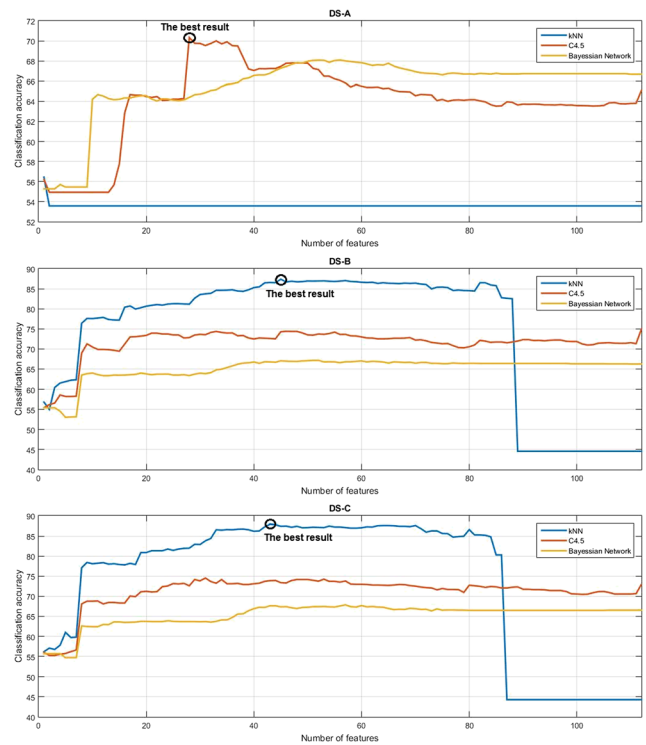


Fig. 12 Variation in classification accuracies

When the confusion matrix of kNN method is examined, it is seen that 11.8% of calm samples are classified as stress. The wrong classification of stress samples is 13%. On average, 87.56% accuracy is achieved. The F-measure is the harmonic mean of the Recall and Precision values. This value is among 0 and 1. In an accomplished classification, the F-measure is awaited to be close to 1 [58]. A balanced data set is used in the study. The F-measure value is calculated as 0.876. In the C4.5 method, 23% of calm samples are stress and 29% of stress

Table 5 Classification results

		DATASET A					
		CA	RMSE	Precision	Recall	F-Measure	# feature
BN		67.86	0.5405	0,682	0,679	0,679	56
kNN		56.59	0.4941	0,585	0,566	0,465	1
C4.5		70.54	0.4697	0,706	0,705	0,706	28
		DATASET B					
		CA	RMSE	Precision	Recall	F-Measure	# feature
BN		66.22	0.5469	0,668	0,662	0,663	51
kNN		87.38	0.3551	0,874	0,874	0,874	45
C4.5		74.20	0.4932	0,742	0,742	0,742	112
		DATASET C					
		CA	RMSE	Precision	Recall	F-Measure	# feature
BN		67.26	0.5474	0,678	0,673	0,674	57
kNN		87.56	0.3102	0,876	0,876	0,876	43
C4.5		74.26	0.4591	0,743	0,743	0,743	31

Table 6 Confusion matrices

	kNN DS-C		C4.5 DS-C		BN DS-A	
	calm	stress	calm	stress	calm	stress
calm	1074	144	936	282	2598	1268
stress	128	841	281	688	994	2177

samples are mixed with calm samples. In Bayesian Networks, these rates are 32.7% and 31%. Faulty classification in the stress class is similarly higher in DS-B and DS-C than DS-A.

Discussion and comparison with other studies

It is not possible to compare the results in this study directly with other studies because all previous studies have been tried with different data sets and their approaches have been evaluated in different ways. Therefore, only stress detection studies which employ data obtained from the smartphone are considered for discussion. Comparison of stress detection studies carried out by smartphone is listed in Table 7. In the study conducted by Ghosh et al. [33, 34], writing behavior was examined and 78% and 84% accuracy were achieved in the classification of four different mood states. Kim and Choi [29] categorized seven different emotions with 0.82 F-score using the information obtained from accelerometer, gyroscope and the touchscreen panel. Lee et al. [30] classified the seven emotions with 67.5% accuracy using information such as location, time, ambient brightness, and weather, in addition to touch screen information. Ciman et al. [1] performed two classes of stress detection with 0.92 F-score with

information such as tap, swipe, scroll and text input on the phone screen. In the studies conducted by Sano and Picard [28], Bogomolov et al. [18] and Vildjionaite et al. [25] 87.5%, 72.28%, and 70% accuracy were obtained in the classifications based on application usage logs, respectively. In the studies such as Syoev et al. [15], Wang et al. [17], Muaremi et al. [27], Gjoreski et al. [6], activities identified during the day were used as features in addition to the application usage logs. Lu et al. [16] performed two classes of stress detection with sound signals with 81% accuracy. This study was carried out using only smartphone data and high accuracy (87.56% CA and 0.876 F-score) classification was performed in 2 classes. In this study, daily smartphone usage of the user was not analyzed, only writing behaviors were examined. Compared to other previous studies, the highest accuracy was achieved in this study among stress detection studies which employ smartphone sensor data.

This study employs an unused sensor and feature set in the literature and performs 2 classes of stress detection with a success rate of 87.56%. Since the sensor information is collected only during the writing process with the keyboard, no battery problem also occurs. In addition, there is no need for a long period of time to reach a definitive decision for stress detection.

Table 7 Comparison of stress detection studies carried out by smartphone

Ref.	Year	Authors	# Class	Performance measure	Success rate	Data source	Method
[1]	2015	Ciman et al.	2	F-score	0.92	Swipe, scroll and text input	Decision Tree
[6]	2015	Gjoreski et al.	3	Accuracy	60	Accelerometer, sound, GPS, Wi-Fi, call logs and light	Random Forest
[15]	2015	Syoev et al.	2	Accuracy	77.5	Sound, light, gyroscope, accelerometer, screen on/off	Simple Logistic
[16]	2012	Lu et al.	2	Accuracy	81	Sound	Gaussian Mixture Models
[18]	2014	Bogomolov et al.	2	Accuracy	72.28	Call and SMS logs, bluetooth and weather conditions	Random Forest
[25]	2018	Vildjionaite et al.	7	Accuracy	70	Phone usage data	Hidden Markov Model
[27]	2013	Muaremi et al.	3	Accuracy	61	Heart rate, sound, accelerometer, GPS, Applications (call, address book, calendar, battery)	Multinomial Logistic Regression
[28]	2013	Sano and Picard	2	Accuracy	87.5	Accelerometer, skin conductance and mobile phone usage	PCA + kNN
[29]	2012	Kim and Choi	7	F-score	0.82	Accelerometer, touch panel and gyroscope	Decision Tree
[30]	2012	Lee et al.	7	Accuracy	67.52	Touchscreen, location, time, weather conditions, ambient brightness	Bayesian Network
[33]	2019	Ghosh et al.	4	Accuracy	78	Typing characteristics	Random Forest
[34]	2017	Ghosh et al.	4	Accuracy	84	Typing characteristics	Random Forest
		This study	2	Accuracy and F-score	87.56, 0.876	Accelerometer, gyroscope and touch screen	kNN

Conclusion

In this study, keyboard typing behaviors and user stressfulness were determined by using the touchscreen panel, gyroscope and accelerometer sensors of smartphones. Owing to a simplistic mobile interface, sensor data was collected from smartphones with the Android operating system and a unique data set was formed. Gain ratio feature selection and cross-validation techniques were used to evaluate the accuracy of classification. It was observed that standard machine learning methods like Bayesian networks, kNN, and C4.5 decision trees achieved successful results in stress detection. The most successful classification was obtained by kNN method. The results obtained in this study showed that it is possible to determine whether the user is under stress or not by using motion sensor data obtained from the smartphone. In future studies, the stress detection application with the smartphone can be improved in several ways: (i) other motion and position sensors of the smartphones can be used. (ii) new and effective feature extraction algorithms may be implemented. (iii) devices with various internal sensors, such as a smartwatch, which the user carries with him during the day can be utilized. (iv) more efficient feature subsets can be extracted with various dimension reduction and feature selection algorithms.

Acknowledgements We would like to thank the personnel and undergraduate students of the Computer Engineering Department of Ege University for volunteering to participate in the experiment. Raw sensor data are available at: <https://tinyurl.com/2019-stress-detection-dataset>.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

- Ciman, M., Wac, K., & Gaggi, O... iSenseStress: Assessing stress through human-smartphone interaction analysis. In proceedings of the 9th international conference on pervasive computing Technologies for Healthcare. 84-91, 2015.
- Can, Y. S., Amrich, B., & Ersoy, C., Stress detection in daily life scenarios using smart phones and wearable sensors: A survey. *J. Biomed. Inform.*, 103139, 2019.
- Gjoreski, M., Luštrek, M., Gams, M., and Gjoreski, H., Monitoring stress with a wrist device using context. *J. Biomed. Inform.* 73:159–170, 2017.
- Picard, R. W., Automating the recognition of stress and emotion: From lab to real-world impact. *IEEE MultiMedia* 23(3):3–7, 2016.
- Stress is Killing You. http://www.who.int/occupational_health/topics/stressatwp/en/ Accessed: 06.11.2019.
- Gjoreski, M., Gjoreski, H., Luštrek, M., & Gams, M., Continuous stress detection using a wrist device: In laboratory and real life. In proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing: Adjunct. 1185-1193, 2016.
- Minguillon, J., Perez, E., Lopez-Gordo, M., Pelayo, F., and Sanchez-Carrion, M., Portable system for real-time detection of stress level. *Sensors* 18(8):2504, 2018.
- Padmaja, B., Prasad, V. R., and Sunitha, K. V., A machine learning approach for stress detection using a wireless physical activity tracker. *Int. J. Mach. Learn. Comput* 8:33–38, 2018.
- Pandey, P. S., Machine learning and IoT for prediction and detection of stress. In: In 2017 17th international conference on computational science and its applications (ICCSA), 2017, July, 1–5.
- Choi, J., Ahmed, B., and Gutierrez-Osuna, R., Development and evaluation of an ambulatory stress monitor based on wearable sensors. *IEEE Trans. Inf. Technol. Biomed.* 16(2):279–286, 2011.
- Zenonos, A., Khan, A., Kalogridis, G., Vatsikas, S., Lewis, T., and Sooriyabandara, M., HealthyOffice: Mood recognition at work using smartphones and wearable sensors. In: In 2016 IEEE international conference on pervasive computing and communication workshops, 2016, 1–6.
- Mozos, O. M., Sandulescu, V., Andrews, S., Ellis, D., Bellotto, N., Dobrescu, R., and Ferrandez, J. M., Stress detection using wearable physiological and sociometric sensors. *Int. J. Neural Syst.* 27(02): 1650041, 2017.
- Egilmez B, Poyraz E, Zhou W, Memik G, Dinda P and Alshurafa N., UStress: Understanding college student subjective stress using wrist-based passive sensing, 2017 IEEE international conference on pervasive computing and communications workshops (PerCom workshops) paper 7, 2017.
- Navea, R. F., Buenvenida, P. J., & Cruz, C. D., Stress detection using galvanic skin response: An android application. In journal of physics: Conference series (Vol. 1372, no. 1, p. 012001). IOP publishing, 2019, November.
- Sysoev, M., Kos, A., and Pogačnik, M., Noninvasive stress recognition considering the current activity. *Pers. Ubiquit. Comput.* 19(7):1045–1052, 2015.
- Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., ... & Choudhury, T., Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 351–360, 2012.
- Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., ... & Campbell, A. T., StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing, 3–14, 2014.
- Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., & Pentland, A. S., Daily stress recognition from mobile phone data, weather conditions and individual traits. In proceedings of the 22nd ACM international conference on multimedia, 477-486, 2014.
- Bauer, G., and Lukowicz, P., Can smartphones detect stress-related changes in the behaviour of individuals? In: In 2012 IEEE international conference on pervasive computing and communications workshops, 2012, 423–426.
- Cho, Y., Bianchi-Berthouze, N., and Julier, S. J., DeepBreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings. In: In 2017 seventh international conference on affective computing and intelligent interaction (ACII), 2017, 456–463.
- Han, H., Byun, K., & Kang, H. G., A deep learning-based stress detection algorithm with speech signal. In proceedings of the 2018

- workshop on audio-visual scene understanding for immersive multimedia, 11-15, 2018.
22. Kostopoulos, P., Kyritsis, A. I., Deriaz, M., and Konstantas, D., Stress detection using smart phone data. In: *eHealth 360°*. Cham: Springer, 2017, 340–351.
 23. Gimpel, H., Regal, C., & Schmidt, M. (2015). myStress: Unobtrusive smartphone-based stress detection. In ECIS.
 24. Raichur, N., Lonakadi, N., and Mural, P., Detection of stress using image processing and machine learning techniques. *International Journal of Engineering and Technology* 9(3):1–8, 2017.
 25. Vildjiounaite, E., Kallio, J., Kyllönen, V., Nieminen, M., Määttä, L., Lindholm, M. et al., Unobtrusive stress detection on the basis of smartphone usage data. *Personal and Ubiquitous Computing* 22(4):671–688, 2018.
 26. Maier, E., Reimer, U., Laurenzi, E., Ridinger, M., and Ulmer, T., A mobile solution for stress recognition and prevention. In *Proc. Int'l Conf. Health Informatics (HealthInf)*:428–433, 2014.
 27. Muaremi, A., Arrich, B., and Tröster, G., Towards measuring stress with smartphones and wearable devices during workday and sleep. *BioNanoScience* 3(2):172–183, 2013.
 28. Sano, A., & Picard, R. W., Stress recognition using wearable sensors and mobile phones. In 2013 Humaine association conference on affective computing and intelligent interaction, 671-676, 2013.
 29. Kim, H. J., & Choi, Y. S., Exploring emotional preference for smartphone applications. In 2012 IEEE consumer communications and networking conference (CCNC), 245-249, 2012.
 30. Lee, H., Choi, Y. S., Lee, S., & Park, I. P., Towards unobtrusive emotion recognition for affective social communication. In 2012 IEEE Consumer Communications and Networking Conference (CCNC), 260-264, 2012.
 31. Gao, Y., Bianchi-Berthouze, N., and Meng, H., What does touch tell us about emotions in touchscreen-based gameplay? *ACM Transactions on Computer-Human Interaction (TOCHI)* 19(4):31, 2012.
 32. Lau, S. H., Stress detection for keystroke dynamics. Doctoral dissertation: Carnegie Mellon University, 2018.
 33. Ghosh, S., Ganguly, N., Mitra, B., & De, P., Tapsense: Combining self-report patterns and typing characteristics for smartphone based emotion detection. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (p. 2), 2017.
 34. Ghosh, S., Sahu, S., Ganguly, N., Mitra, B., and De, P., EmoKey: An emotion-aware smartphone keyboard for mental health monitoring. In: In 2019 11th international conference on communication systems & networks (COMSNETS), 2019, 496–499.
 35. Exposito, M., Hernandez, J., & Picard, R. W., Affective keys: Towards unobtrusive stress sensing of smartphone users. In proceedings of the 20th international conference on human-computer interaction with Mobile devices and services adjunct (pp. 139-145), 2018, September.
 36. Sağbaş, E. A., & Ballı, S., Usage of the smartphone sensors and accessing raw sensor data. In proceedings of the 17th conference of Academic Computing:158–164, Eskişehir, Turkey, 2015, February.
 37. Peker, M., Ballı, S., & Sağbaş, E. A., Predicting human actions using a hybrid of ReliefF feature selection and kernel-based extreme learning machine. In handbook of research on predictive modeling and optimization methods in science and engineering. 379-397, 2018.
 38. Yüksel, A. S., Senel, F. A., and Cankaya, I. A., Classification of soft keyboard typing behaviors using Mobile device sensors with machine learning. *Arab. J. Sci. Eng.* 44(4):3929–3942, 2019.
 39. L. Bernardi, J. Wdowczyk-Szulc, C. Valenti, S. Castoldi, C. Passino, G. Spadacini, and P. Sleight., Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability. *J. Am. Coll. Cardiol.* 1462–1469, 2000.
 40. Dickerson, S. S., and Kemeny, M. E., Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological bulletin.*:355–391, 2004.
 41. Stroop, J. R., Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18(6):643, 1935.
 42. Lezak, M.D., *Neuropsychological assessment*, Oxford University Press, USA, 2004.
 43. Likert, R., A technique for the measurements of attitudes. *Archives of psychology* 55, 1932.
 44. Hall, M. A., and Holmes, G., Benchmarking attribute selection techniques for discrete class data mining. *IEEE Trans. Knowl. Data Eng.* 15(6):1437–1447, 2003.
 45. Priyadarsini, R. P., Valarmathi, M. L., and Sivakumari, S., Gain ratio based feature selection method for privacy preservation. *ICTACT Journal on soft computing* 1(4):201–205, 2011.
 46. Trabelsi, M., Meddouri, N., and Maddouri, M., A new feature selection method for nominal classifier based on formal concept analysis. *Procedia Computer Science* 112:186–194, 2017.
 47. Karegowda, A. G., Manjunath, A. S., and Jayaram, M. A., Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management* 2(2):271–277, 2010.
 48. Yazıcı B, Yaşlı F, Gürleyik HY, Yurgut UO., Aktas MS, Kalıpsız O. Veri Madenciliğinde Özellik Seçim Tekniklerinin Bankacılık Verisine Uygulanması Üzerine Araştırma ve Karşılaştırmalı Uygulama. In Proceedings of the 9th Turkish National Software Engineering Symposium, 1–11.
 49. Witten, I. H., Frank, E., and Hall, M. A., *Data mining: Practical machine learning tools and techniques*. 3rd edition. Burlington: Morgan Kaufmann, 2011.
 50. Yüksel, A. S., Şenel, F. A., and Çankaya, İ. A., Classification of writing behaviors using mobile device sensors. *Dicle University Journal of Engineering* 9(1):133–142, 2018.
 51. Witten, I. H., and Frank, E., Data mining: Practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record* 31(1):76–77, 2002.
 52. Amin, H. U., Malik, A. S., Ahmad, R. F., Badruddin, N., Kamel, N., Hussain, M., and Chooi, W.-T., Feature extraction and classification for EEG signals using wavelet transform and machine learning techniques. *Australas. Phys. Eng. Sci. Med.* 38(1):139–149, 2015.
 53. Korb, K. B., and Nicholson, A. E., *Bayesian artificial intelligence*. 2nd ed. Boca Raton: FL, USA, CRC Press, 2011.
 54. Sağbaş, E. A., and Ballı, S., Transportation mode detection by using smartphone sensors and machine learning. *Pamukkale University Journal of Engineering Sciences* 22(5):376–383, 2016.
 55. Feng T., Timmermans H.J.P., Comparative evaluation of algorithms for GPS data imputation. 13th WCTR, 1-11, 2010
 56. Ballı, S., and Sağbaş, E. A., Classification of human motions with Smartwatch sensors. *Süleyman Demirel University Journal of Natural and Applied Sciences* 21(3):980–990, 2017.
 57. Peker, M., A new approach for automatic sleep scoring: Combining Taguchi based complex-valued neural network and complex wavelet transform. *Comput. Methods Programs Biomed.* 129:203–216, 2016.
 58. Ballı, S., Sağbaş, E. A., and Peker, M., Human activity recognition from smart watch sensor data using a hybrid of principal component analysis and random forest algorithm. *Meas. Control.* 52(1–2):37–45, 2019.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.