International Journal of Medical Informatics 185 (2024) 105401

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

StresSense: Real-Time detection of stress-displaying behaviors

Nida Saddaf Khan [a,e], Saleeta Qadir [b,e], Gulnaz Anjum [c,*], Nasir Uddin [d]

[a] CITRIC Health Data Science Centre, Medical College, Agha Khan University, Stadium Road, P.O. Box 3500, Karachi 74800, Pakistan
[b] National High-Performance Computing Center, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Schloßplatz 4, 91054 Erlangen, Germany
[c] Department of Psychology, University of Oslo, Forskningsveien 3A, Harald Schjelderups hus, 0373 Oslo, Norway
[d] School of Computer Science, National University of Computer and Emerging Sciences, Karachi Campus, Pakistan
[e] Telecommunication Research Lab (TRL), School of Mathematics and Computer Science, Institute of Business Administration, Karachi, Pakistan

A R T I C L E I N F O

Keywords:
Deep Neural Networks (DNN)
Stress
Boredom
human activity recognition (HAR)
Mental health

A B S T R A C T

Background: Wrist-worn gadgets like smartphones are ideal for unobtrusively gathering user data, in various fields such as health and fitness monitoring, communication, and productivity enhancement. They seamlessly integrate into users' daily lives, providing valuable insights and features without the need for constant attention or disruption. In sensitive domains like mental health, these devices provide user-friendly, privacy-protected means of diagnosis and treatment, offering a secure and cost-effective avenue for seeking help.
Objectives: This study addresses the limitations of traditional mental health assessment techniques, such as intrusive sensing and subjective self-reporting, by harnessing the unobtrusive data collection capabilities of smartphones. Equipped with accelerometers and other sensors, these devices offer a novel approach to mental health research. Our objective was to develop methods for real-time detection of stress and boredom behavior markers using smart devices and machine learning algorithms.
Methodology: By leveraging data from accelerometers (A), gyroscopes (G), and magnetometers (M), we compiled a dataset indicative of stress-related behaviors and trained various machine-learning models for predictive accuracy. The methodology involved collecting data from motion sensors (A, G, and M) on the dominant arm's wrist-worn smartphone, followed by data preprocessing, transformation from time series format, and training a Deep Neural Network (DNN) model for activity recognition.
Findings: Remarkably, the DNN achieved an accuracy of 93.50% on test data, outperforming traditional and ensemble machine learning methods across different window sizes, and demonstrated real-time accuracy of 77.78%, validating its practical application.
Conclusion: In conclusion, this research presents a novel dataset for detecting stress and boredom behaviors using smartphones, reducing reliance on costly devices and offering a more objective assessment. It also proposes a DNN-based method for wrist-worn devices to accurately identify complex activities associated with stress and boredom, with benefits in terms of privacy and user convenience. This advancement represents a significant contribution to the field of mental health research, providing a less intrusive and more user-friendly approach to monitoring mental well-being.

1. Introduction

Activity recognition has become an integral part of human behavior analysis since it allows us to store, process, and analyze people's behaviors. Human activity recognition (HAR) can be divided into two broad types: video-based systems and sensor-based systems [1]. Video-based systems rely on images and videos, while sensor-based systems rely on the body or environment sensors. This paper focuses on sensor-based systems such as wearable devices like Inertial Measurement Units (IMUs) or smartwatches and fitness trackers as they are non-intrusive and widely available. The ubiquitous nature of wearable devices allows for data collection, irrespective of location or social context.

The advancements in the field of artificial intelligence and HAR have led to myriad applications such as smart homes [54], healthcare [3], and improved manufacturing [4]. Research in human emotions based on activity recognition for emotion detection often requires various sources

* Corresponding author at: Department of Psychology, University of Oslo, Norway.
E-mail addresses: nida.saddaf@aku.edu (N. Saddaf Khan), saleeta.qadir@fau.de (S. Qadir), gulnaz.anjum@psykologi.uio.no (G. Anjum), nasir.uddin@nu.edu.pk (N. Uddin).

https://doi.org/10.1016/j.ijmedinf.2024.105401
Received 9 October 2023; Received in revised form 29 February 2024; Accepted 2 March 2024
Available online 7 March 2024

of data, including physiological signals (e.g., Electroencephalogram (EEG) and Galvanic Skin Response (GSR)), environmental data (e.g., audio, location, and weather), and videos (e.g., facial expressions, motions, and gestures). Many classical machine learning (ML) algorithms have been used to detect stress through physiological sensors alone and boredom through facial expressions and gestures [5,6]. However, deep learning (DL)-based algorithms have been extensively studied and employed in HAR [7,8,17].

Experimental research on humans suggests that behaviors can impact how our bodies adapt and cope with stress. The most common behaviors associated with self-regulatory coping behaviors (SRCB) that are harmful to health are smoking, alcohol/drug use, and overeating [9]. Body-focused repetitive behaviors (BFRBs) are another common group of conditions when individuals suffer daily from stress, boredom, and anxiety. These behaviors are spontaneously manifested without the individual's awareness of the actions. BFRBs include trichotillomania (hair-pulling), excoriation (skin picking), and nail-biting. These habits can be detrimental, causing bald spots, sores, wounds, and infections in affected areas [10]. The behaviors chosen to represent stress responses are smoking, over-eating, nail-biting and compulsive face touching. Measuring the rate of occurrence of such activities can help us to monitor the level of one's stress and boredom [10].

In this study, we make a significant contribution to the research community by introducing a novel dataset related to activities associated with stress or boredom. In today's digital age, commonly available devices such as smartphones and smartwatches have emerged as invaluable tools for collecting medical and clinical conditions-related data. Their ubiquity in daily life, coupled with their ever-evolving capabilities, presents a unique opportunity to revolutionize the field of healthcare and medical research. These devices offer continuous and non-invasive monitoring of vital signs, physical activity, sleep patterns, and other health-related metrics, providing a wealth of real-time data that was previously challenging to obtain. A wrist-worn sensors data was used to classify the complex, very fine-grained activities of daily living (ADLs) using a Deep Neural Network (DNN)-based classifier. The ease of data collection through these devices fosters patient engagement and compliance, enabling longitudinal studies and personalized healthcare interventions. Moreover, the technique proposed in this paper places a paramount emphasis on addressing and mitigating privacy concerns, which have become increasingly critical in today's data-driven research landscape. Our approach incorporates robust privacy-preserving mechanisms at its core by only collecting five motion sensors data from 40 participants.

2. Literature review

HAR models are usually evaluated based on accuracy and computational cost. Previous works have effectively recognized activities that involve unique hand movements like drinking, smoking, and writing in controlled environments using fixed window sizes [11] and features such as mean, median, minimum, and maximum [12,13] paired with ML algorithms such as K-Nearest Neighbor (KNN), Naïve Bayes NB, and Decision Tree (DT) [14,15]. Particularly in one study [14] where construction tasks, complex upper-body movements (e.g., manual tool and material handling), and whole-body actions are recognized using an accelerometer attached to the dominant wrist for adequate distinction in hand or arm movements. Optimal conditions (e.g., window sizes and features) were used with the Support Vector Machine (SVM) and multilayer perceptron to combat noise and increase classification performance. The research [15] presents a comprehensive review of different DL algorithms using HAR-specific features for the detection of simple (e.g., sitting, standing, and jogging) and complex hand movements (e.g., eating, smoking, and making coffee). Since most of these works employed methodologies that used hand-crafted features, it becomes difficult to compare and increases computational cost, thereby reducing energy efficiency. There have been attempts to pair accelerometer with ML algorithms to detect static activities and pair accelerometer with DL methodologies to detect dynamic activities [16].

A recent study [18] shows that body movements convey information relevant to one's current emotions. Therefore, it is possible to extract emotion-specific data from physical behaviors exhibited by users. Most research regarding emotion-relevant activity recognition is based on visual data (e.g., facial expressions, gait, and posture) [55] or physiological signals (e.g., changes in heart rate, skin conductivity, and dermal temperature) [56]. However, research related to inertial motion-based activity recognition for emotions is limited.

Other implementations used a combination of Artificial Neural Network (ANN) and Recurrent Neural Network (RNN) to detect activities like walking, standing, and sitting with high accuracy [19,20]. Research has also been done to recognize specific types of arm swings in badminton using multiple sensors placed at various positions on the arm with up to 82 % weighted accuracy using the Convolution Neural Network (CNN) [21]. Some have even combined inertial sensors with ambient environmental (temperature, atmospheric pressure, humidity) sensors and location context (Bluetooth reception) to detect complex activities with high accuracy using a multilayer feed-forward artificial neural network that is trained with stochastic gradient descent using back-propagation [22]. Several similar wrist movements as ours have been distinguished using N-based clustering with Hidden Markov Model (HMM) with up to 81 % accuracy [23]. In [24], smoking, eating, and face touching are detected using ANN on an Apple watch's output on the wrist.

A recent study [25] introduces a HAR system designed to categorize 18 physical activities by leveraging data from sensors embedded in smartphones and smartwatches. The system operates in two main steps: feature extraction and HAR. To extract features, a hybrid structure combining CNN and Bidirectional Gated Recurrent Unit (BiGRU) was employed. For activity recognition, the study utilized a Single-Hidden-Layer Feedforward Neural Network (SLFN) with a Regularized Extreme Learning Machine (RELM) algorithm.

In another study [26] a novel HAR system is proposed based on a wearable pedal, utilizing the differential spatio-temporal Long Short-Term Memory (DST-LSTM) method. The wearable pedal device embedded in footwear captures pedal musculoskeletal response (PMR) data of dorsum pedis. The DST-LSTM developed in this research is applied to classify five typical activity statuses: floor walking, down and up stair walking, standing, and sitting. Additionally, a study by [27] employs deep learning models, such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), for recognizing activities indicative of depression. These models were trained using two publicly available datasets, WISDM and MHEALTH. Notably, our study stands out by collecting data specifically for this research and utilizing a widely available device for this purpose. In contrast, other researchers have either utilized specialized sensors for data collection or relied on existing datasets with limited features (activities) associated with stress.

In the realm of stress detection, [28] incorporates Internet of Things (IoT) techniques and proposes an algorithm for stress level detection. The stress level prediction is based on physical activity, humidity, temperature, and step count. Another study [29] constructs a Bidirectional Long Short-Term Memory (Bi-LSTM) model to predict stress levels, utilizing the WISDM dataset for categorizing stress emotions into baseline, amusement, stress, and meditation. The focus of our study is to collect sensors data for stress associated activates from non-special sensors and develop machine learning model to recognize them using limited sensors rather than prediction of stress level from wide range of sensors data.

There are various commercial applications for stress detection, such as Spire, Zensorium's Being,[1] WellBe,[2] and Tinke.[3] These devices employ Infra-Red (IR) sensors to monitor blood flow and respiration, aiding in stress identification. Typically, these sensors are not commonly integrated into everyday smart devices, necessitating consumers to purchase these specialized gadgets. Nevertheless, this research explores other ways to harness the capabilities of readily available common devices for stress symptom detection, it could make stress assessment accessible to everyone without the need for costly specialized equipment.

3. Data collection

A publicly available dataset, the Wearable Stress and Affect Detection (WESAD) dataset collected from 15 subjects through a wrist and chest-worn wearable devices is a largely used dataset for stress detection. Apart from three-axis acceleration the training data uses features such as body temperature, electrocardiogram, electromyogram and blood volume pulse for its prediction model [30]. In contrast to that we collected data from 40 participants irrespective of their medical conditions for training data collection. We obtained ethical approval from the Capital University of Science and Technology (CUST), Pakistan, under Ethical Review Report number: CUST/ORIC/IERB/2023/02. In keeping with ethical requirements in Pakistan, we obtained verbal consent from all participants involved in our study. Verbal consent was deemed the most appropriate and feasible form of consent considering the specific context of our research, which involved populations in regions where written consent may pose logistical challenges or cultural barriers. The process of obtaining verbal consent was conducted in a thorough and respectful manner, ensuring that each participant was fully informed about the nature and purpose of the study, as well as their right to withdraw at any point without any consequences. The anonymity of participants was carefully documented, maintaining the integrity and ethical rigor of our research approach.

For the data collection, the data features only include the time series data corresponding to accelerometer, gyroscope and magnetometer which were found to be good predictor of activities associated with stress. Besides reducing the complexity of model, our approach relies on only wrist worn wearable sensors which are relatively easier to handle for collecting data.

Our dataset [31] has a total of 495,446 instances for twelve columns, namely time stamp, user ID, activity label, and three-dimensional values of the accelerometer (A), gyroscope (G), and magnetometer (M) named Ax, Ay, Az, Gx, Gy, Gz, Mx, My, and Mz. The sensors' device was Samsung Galaxy S5 and the body position to wear the sensor is the wrist of the dominant arm. A custom holder was used for the wrist to ensure that each participant's orientation and placement remained the same. It was placed just above the wrist joint of the dominant hand, where smartwatches are usually worn. A similar approach was previously used to collect data related to anxiety where an Inertial Measurement Unit (IMU) was worn at the wrist position [2,32].

Data collection was done through an android application [33] with a sampling rate of 50 Hz, as studies show that the frequency of the range of human movement lies within 50 Hz [34]. The recording was started and stopped manually, and the files were stored locally for each activity. There were, in total, 40 healthy right-handed participants (20 males and 20 females) between the ages of 20–25. However, not all activities were performed by each participant. The activities were performed in a structured environment where each participant performed activities assigned to them.

The labeling was done with the help of the android app [33], which allows giving the name of the activity as the file name. The data for each activity is recorded one at a time, and later all the instances were given the activity label for which the recording was performed. The same approach is used for all the participants. Fig. 1 shows the number of instances versus the activity (a) and the user ID (b). In Fig. 1a, it can be seen that the no. of instances for each activity is almost equal, i.e., almost equal class distribution. On the other hand, Fig. 1b shows that the no. of instances recorded by each user is not the same as some recorded all five activities and some recorded one or two activities of the set. It should be noted that the data collection process was based on volunteering consent and involved no danger of emotional or physical loss. The participants were at liberty to terminate the collection process at any time. The data collection was prefaced with information sheets and a consent form. In the succeeding subsections, the recorded activities are explained in the context of this research.

3.1. Smoking

Smoking can be described as a gesture of time dependence. A session consists of at least one puff, defined as the time taken to raise the cigarette to one's lips, inhale and lower their arm. Smoking data were collected till one cigarette was smoked and burnt completely by a participant.

3.2. Eating

Eating has many different forms all over the world. The complexity of measurement of eating movements arises from (1) individual preferences in using utensils, diverse table etiquettes, and cultural dining practices; (2) unique personal habits like touching one's hair or face. This diversity makes it difficult to create a one-size-fits-all model for all users [35]. Because of these complexities, a protocol is made to avoid variations in the pattern of this activity. Eating was recorded for the movements of chocolate eating and performed until the chocolate bar was finished by a participant.

3.3. Nail-Biting

Onychophagia, or nail-biting as it is more commonly known, is a stress-coping mechanism triggered by nervousness, stress, hunger, or even boredom. It is not only unsanitary but also inadequate for the nails [36].

In this research, habitual nail biters were asked to perform the activity in sitting and standing conditions. It was observed that nail-biting

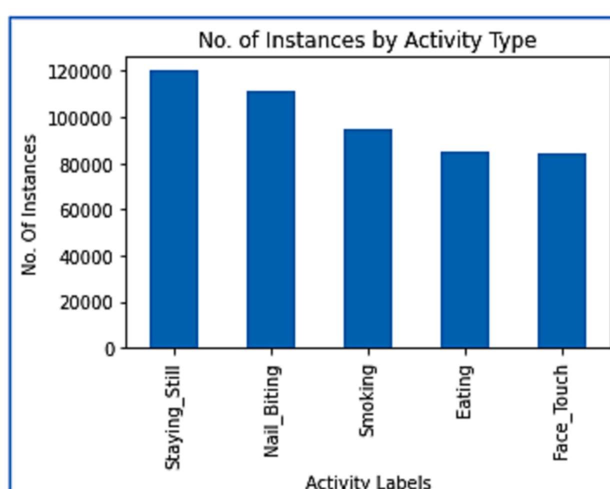

Fig. 1a. No. of instances for each activity.

[1] https://www.engadget.com/2015–01-04-zensorium-being.html.
[2] https://wellbe.me/.
[3] https://www.amazon.in/TINKE-Stress-Respiratory-Oxygen-Monitor/dp/B00DTWOSXG.

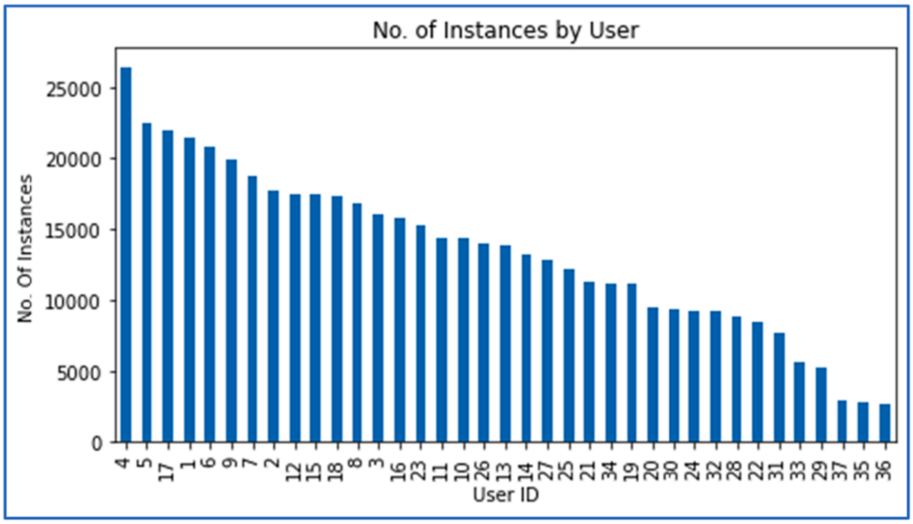

Fig. 1b. No. of instances recorded by each participant.

had more to do with the wrist movement near the face than elbow and shoulder movement, as previously seen in the activities of eating and smoking. This activity was recorded for three minutes by a participant.

3.4. Face touching

Face-touching is defined as touching any part of one's face, be it head, hair, or neck. Occasional and spontaneous touches are natural. However, negative affect and attention distraction seem correlated to the frequency of face touching, and the behavior may be involved in regulating emotion [6,37].

Here the protocol includes parts of the face based on Centers for Disease Control and Prevention (CDC) guidelines (eyes, nose, and mouth) as well as other frequently touched areas (hair, forehead, temple, ear, cheek, and chin) [38]. Participants were asked to touch their faces as they usually do in both sitting and standing conditions, and this activity was performed for three minutes by a participant.

3.5. Staying Still

Staying still was the base case (baseline); the participants were asked to stay idle and not perform other activities. The protocol allowed folding arms, grasping hips, and fumbling with fingers. Some were seen to sway or fidget while others remained still. A participant performed these in standing and sitting positions for a three-minute duration.

4. Methodology

4.1. Study design

This research comprises a prospective data analysis on the data collected from the participants as explained in detail in the last section. The sensors stream data was collected from the smartphone worn at the wrist position of the dominant arm of the participants. The sensors stream of three motion sensors, i.e., A, G, and M, were collected and recorded in a log file corresponding to each activity. After the collection of the data from all the participants, the data was integrated for further processing and analysis.

4.2. Data preprocessing

The first step was preprocessing the data, including the missing values and normalization. As more than one type of sensor had been used in this study, the chances of missing values were higher as all three sensors experience lag in their recording. Records with missing values at the beginning of the recording were excluded, as those records typically did not directly relate to the recorded activity. For other instances of missing values encountered during activity recording, imputation was performed based on the mean of the preceding or succeeding five time series values. Additionally, normalization was applied to standardize the streams to similar ranges. Normalization is one of the recommended practices for training a DNN is to normalize the data, aiming for a mean close to 0. This normalization process typically accelerates learning and facilitates faster convergence.

4.3. Data transformation

After preprocessing, the transformation of data was required. The sensor streams were in the format of time series values that needed to be transformed to apply ML algorithms. The data was transformed by selecting a window size with some overlap (or without an overlap) between consecutive window sizes. An overlap was kept between the windows to increase the number of samples, resulting in better algorithm training. In [39], the ideal overlap rests at 80 %, such that the stream of sensors is transformed into windows with an 80 % overlap, and we have kept the same setting for our data.

4.4. Raw data & statistical features

Two approaches have been implemented in this research: one with the raw sensors' data and the other with statistical features. The statistical features, namely, mean, standard deviation (std), minimum (min), and maximum (max) of each window, are calculated. With raw data mode, the raw sensors data is fed to the input layer of the neural network, so if the window size is 50 and the sensor of the accelerometer is selected, then the total number of neurons at the input layer will be 150 (50 values for each dimension x, y, and z of the accelerometer). For statistical features, the mean, std, min, and max are calculated for each dimension of selected sensors and fed into the neural network. For example, suppose the window size is 50, and the three-dimensional sensors of the accelerometer and gyroscope are used. In this case, the total number of neurons at the input layer will be 24 (four statistical features for each of three dimensions, i.e., x, y, and z of accelerometer and gyroscope). The neural network's input layer depends upon the window size, the number of sensors for raw data, and the number of sensor channels. The results of both approaches are shown in the next section of this paper.

4.5. Model development

The DNN model has been trained to recognize human activities in this research. These models are used in pattern recognition problems and require large datasets. They are made of neurons that get activated from the sensor's perceived environment, while others get activated by their weighted connections from the previously activated neurons [5]. In Equation (1), $X = x_1, x_2, \cdots x_n$ is an N-dimensional input, z is the neuron's response, ω_i is the ith weight for ith input, and ω_o is a constant bias. The relationship between the features and the labels is determined through this sequence to the input data.

$$z(X) = \sum_{i=1}^{N} (\omega_i x_i + \omega_o) \tag{1}$$

The transformed data are fed to the DNN for training the patterns associated with each activity. Seven DNN models were created with varying network architectures and parameters to find the most appropriate accuracy and recognition time model. The models' details are provided in Table 1.

$$f(x) = Max(0, x) \tag{2}$$

The overall steps are summarized in Fig. 2. The network comprises hidden layers followed by a SoftMax layer. Each hidden layer consists of specific numbers of neurons corresponding to a linear transformation and a rectified-linear (ReLU) activation function to break linearity. Since DL models run on high-performance GPUs which excel at simple operations (add, multiply), solving ReLU equation given in Equation (2) can be ten times faster than solving approximations for tanh on the GPU [40].

4.6. Evaluation & performance comparison

For evaluation, the hold-one-out method is used to split the data into training and test datasets. The data of one participant is kept to evaluate the model performance, whereas the rest of the data is used to train the model. In order to compare the performance of DNN with other ML algorithms, some traditional algorithms have also been constructed. These algorithms are Decision Tree (DT), Naïve Bayes (NB), and K-Nearest Neighbor (KNN), along with ensemble models of Random Forest (RF) and Gradient Boosting (GB). These algorithms have been employed as benchmarks in the field of HAR, with numerous research articles utilizing them to evaluate and compare the effectiveness of their proposed solutions [41–45]. Finally, the best model's performance is evaluated in real-time to assess the time it takes to recognize the activities. In the next section, the details of the experiments and results are presented.

5. Experiments and results

The experiments were conducted for both types of data, i.e., raw data and statistical features. The seven models of DNN with varying architectures were created along with some traditional ML classifiers of NB, DT, KNN, and two ensemble models of RF and GB algorithms. Sensors of A and G were used for this experiment as they are the most widely used motion sensors in the field of HAR. The window sizes (win_size) varied from 50 to 600 readings with the step size of 50, whereas the sensor's frequency was kept 50 (50 readings per second). The overlap percentage between the consecutive windows was kept at 80 % to keep sufficient samples for learning the model. The consolidated results of this experiment are given in Table 1, with statistical features in part (a) and without statistical features (raw data) in part (b). The last two columns show the maximum test accuracy for each window size and the model that achieved that accuracy. Similarly, the last two rows show the highest test accuracy by each model and the window size at which this accuracy was achieved.

It is evident from Table 2 that the test performances of statistical features-based models outperform the models trained via raw data. Moreover, statistical features-based models are less complex in input nodes, especially for large window sizes, i.e., the model's input structure remains the same for any window size, whether it is small such as 10, or large such as 600. This lesser complexity feature makes using a statistical feature set more desirable as it reduces the complexity and overall execution time of training.

As far as the models are concerned, the models of DT and NB gave good performance neither with statistical features nor with raw data, as their accuracies were pretty low for all the window sizes. Although the KNN algorithm gave comparable performance even though it started dominating the other algorithms in some cases (in Table 2a, from window size 350 to onwards), the time it takes to classify the test data was too high from all other algorithms. Due to the peculiar nature of KNN, most of the computations are performed at testing time. Hence, this technique is unsuitable for the platform where the recognition is needed in real-time. The time required to preprocess the time series values and to transform them into windows such that the ML algorithms could be applied quite long, and if the classifier itself takes a long time to classify the test cases, then the overall time requirement increases with the increase in window sizes which is not desirable aspect for real-time solutions. Hence, due to time constraints, KNN and lower performances NB and DT are not included in further experiments.

The other algorithms, i.e., all variants of DNN, RF, and GB models, gave good performances. The same models were evaluated using the data of A, G, and M sensors, as shown in Table 3. In this table, the DNN models outperformed GB and RF models for almost all the windows sizes, which shows the power of DNN over traditional ML techniques. Among DNN variants, the highest performance of 93.52 % is given by DNN3, which is closely followed by DNN1, DNN4, and DNN2 with 92.13 %, 91.9 %, and 91.67 %, respectively. Furthermore, if we compare the performances by each window size, then DNN3 performance is superior in the majority of the cases, i.e., at window sizes of 100, 250, and 450, followed by DNN2, DNN4, and DNN6, which have given the highest performances for two window sizes each. Hence, DNN3 has beaten other models in classifying the activities with the highest test accuracy. DNN has made many breakthroughs in the HAR research area. The deep architectures can represent complex functions compactly, which have outperformed most ML algorithms in many applications, such as face detection and speech recognition [46].

The graph shown in Fig. 3 is plotted between the test accuracy of model DNN3 and window sizes where the window sizes are kept from 50 to 1000 to find the most suitable window size. In this figure, it can be observed that the model's accuracy increases with the increase in window size till it reaches the point of 450. After 450, the accuracy drops and stays low till the point of 950. At 1000 window size, the curve shows a sudden spike in accuracy, which can be ignored as at point 450, an excellent comparable accuracy of 93.5 % has already been achieved. As already stated, the lower window size means quicker training and recognition in real-time systems. Hence, the most suitable window size for this problem is 450, which has been circled in Fig. 3.

A confusion matrix is presented in Fig. 4 depicts the model's performance on each activity. Face Touch and staying Still are classified with 100 % accuracy, but the activities of Eating, Nail Biting, and

Table 1
DNN models and their descriptions.

Model	Description
DNN1	Input – FC (10, ReLU) – FC (10, ReLU) – Classification (SoftMax)
DNN2	Input – FC (20, ReLU) – FC (20, ReLU) – Classification (SoftMax)
DNN3	Input – FC (30, ReLU) – FC(30, ReLU) – Classification (SoftMax)
DNN4	Input – FC (40, ReLU) – FC(40, ReLU) – Classification (SoftMax)
DNN5	Input – FC (100, ReLU) – Classification (SoftMax)
DNN6	Input – FC (100, ReLU) – FC(50, ReLU) – Classification (SoftMax)
DNN7	Input – FC (100, ReLU) – FC(50, ReLU) – FC(50, ReLU) – Classification (SoftMax)

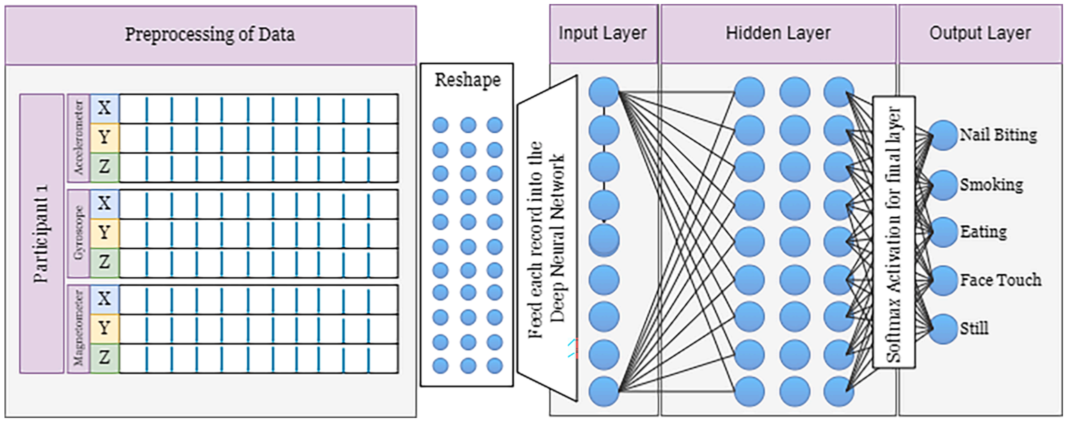

Fig. 2. Proposed model.

Table 2a
Test Accuracy (%) of models with sensors of A and G for varying window sizes with statistical features.

Win_Size	DNN1	DNN2	DNN3	DNN4	DNN5	DNN6	DNN7	DT	NB	KNN	RF	GB	Max	Model
50	83.07	83.73	82.27	82.04	82.6	81.09	81.14	68.13	58.18	80.91	82.56	79.3	**83.73**	**DNN2**
100	83.79	84.94	83.22	84.27	82.94	82.17	81.79	69.69	62.63	81.79	84.	82.65	**84.94**	**DNN2**
150	85.4	83.24	84.97	83.09	85.26	82.23	86.13	70.38	65.61	84.1	83.53	84.68	**86.13**	**DNN7**
200	84.8	85.58	85.96	87.13	81.87	85.77	81.68	72.71	66.67	85.96	83.63	85.58	**87.13**	**DNN4**
250	84.24	84.48	86.7	84.73	83.74	84.24	84.48	51.97	67.49	84.98	85.47	87.19	**87.19**	**GB**
300	87.46	88.66	85.67	87.76	86.57	86.27	83.88	53.13	70.15	86.57	87.76	86.57	**88.66**	**DNN2**
350	85.21	84.86	84.15	83.1	85.92	85.21	83.8	55.28	71.48	87.68	86.97	88.03	**88.03**	**GB**
400	85.37	86.99	85.77	86.59	85.37	85.77	79.27	57.32	73.58	89.84	86.99	87.4	**89.84**	**KNN**
450	86.11	87.04	86.57	87.04	88.89	87.96	84.72	57.41	73.15	89.35	87.96	87.96	**89.35**	**KNN**
500	85.94	86.98	86.46	86.46	86.98	86.98	84.9	57.29	73.44	92.19	88.02	86.98	**92.19**	**KNN**
550	89.53	84.3	87.79	84.88	87.79	83.14	86.05	56.4	73.84	93.02	88.37	87.79	**93.02**	**KNN**
600	86.54	87.82	88.46	86.54	89.74	85.9	81.41	57.69	73.08	91.67	86.54	88.46	**91.67**	**KNN**
Max	**89.53**	**88.66**	**88.46**	**87.76**	**89.74**	**87.96**	**86.13**	**72.71**	**73.84**	**93.02**	**88.37**	**88.46**	**93.02**	**KNN**
Win_Size	**550**	**300**	**600**	**300**	**600**	**450**	**150**	**200**	**550**	**550**	**500**	**600**	**550**	

Table 2b
Test Accuracy (%) of models with sensors of A and G for varying window sizes with raw data.

Win_Size	DNN1	DNN2	DNN3	DNN4	DNN5	DNN6	DNN7	DT	NB	KNN	RF	GB	Max	Model
50	80.34	81.9	78.6	80.01	82.13	80.95	78.5	63.74	58.98	72.8	77.23	82.74	**82.74**	**RF**
100	81.12	79.69	80.65	81.89	79.79	81.12	81.98	64.44	62.92	69.11	79.98	84.65	**84.65**	**RF**
150	80.64	84.39	75.87	80.92	81.21	80.49	81.79	64.02	65.9	69.65	80.35	83.67	**84.39**	**DNN2**
200	77.19	81.68	80.51	84.99	79.14	83.43	79.73	65.3	66.67	65.3	84.21	81.09	**84.99**	**DNN4**
250	77.59	86.45	84.24	76.6	76.6	76.35	73.89	65.02	67.24	66.01	84.24	81.77	**86.45**	**DNN2**
300	77.61	77.31	83.88	79.1	84.18	79.4	79.7	66.87	68.96	71.64	84.18	82.69	**84.18**	**GB**
350	80.28	79.58	86.97	85.56	79.58	76.41	79.58	66.2	69.37	75.7	85.56	84.51	**86.97**	**DNN3**
400	79.27	78.86	79.27	76.83	81.3	80.49	77.24	65.85	69.51	71.14	85.37	84.15	**85.37**	**GB**
450	80.56	86.11	83.8	86.11	82.41	81.94	75.93	67.13	68.98	71.3	85.65	84.72	**86.11**	**DNN4**
500	81.25	85.94	83.33	82.81	79.17	80.73	79.17	66.67	70.31	72.4	86.98	83.85	**86.98**	**GB**
550	82.56	83.72	85.47	86.05	83.14	83.72	84.3	68.02	70.35	68.02	84.3	84.3	**86.05**	**DNN4**
600	83.33	83.33	83.33	83.33	80.13	78.85	78.21	64.74	69.87	69.23	85.9	83.97	**85.9**	**GB**
Max	**83.33**	**86.45**	**86.97**	**86.11**	**84.18**	**83.72**	**84.3**	**68.02**	**70.35**	**75.7**	**86.98**	**84.72**	**86.98**	**GB**
Win_Size	**600**	**250**	**350**	**450**	**300**	**550**	**550**	**550**	**550**	**350**	**500**	**90**	**500**	

Smoking are recognized with 93 %, 92 %, and 88 %, respectively. These results make sense as these three activities are similar and involve the wrist's movement around the mouth, so the algorithm confuses these activities with each other.

5.1. Real-Time analysis

The Real-time data analysis was done using PhonePi,[4] an application that allows live data streaming from sensors present in a smartphone to a web socket supporting server. PhonePi was used to record real-time data from the A, G, and M in this research. This simple Flask-based server used WebSockets to receive sensors data and store it in a file. An open connection allows real-time data exchange to stream readings onto the server. This setup does not require any physical connections, and for testing purposes, all real-time data collection and recognition was done on localhost. The results were then rerouted to any desired port and address in file format. This way, a seamless supply of data was guaranteed.

During the real-time experiment, it was observed that readings from M were not synchronized with the readings of the A and G, and they were coming with a rolling delay of 10 s per value. Due to this delay, the

[4] PhonePi is a simple flask-based server that relies on web sockets to deliver real-time sensor data. The GitHub can be found here: https://github.com/priyankark/PhonePi_SampleServer.

Table 3
Test Accuracy (%) of short-listed models with sensors of A, G, and M for varying window sizes with statistical features.

Win_Size	DNN1	DNN2	DNN3	DNN4	DNN5	DNN6	DNN7	RF	GB	Max	Model
50	80.95	82.46	80.39	79.54	81.66	78.17	77.23	82.23	80.67	82.46	DNN2
100	81.03	82.65	85.99	82.36	78.17	78.84	77.5	82.75	82.75	85.99	DNN3
150	81.79	84.25	84.54	79.19	84.39	82.8	82.8	82.51	85.12	85.12	GB
200	82.07	88.11	83.04	84.02	87.33	87.13	89.08	83.82	86.74	89.08	DNN7
250	86.21	80.3	86.95	83.74	85.47	85.96	85.71	83.74	86.21	86.95	DNN3
300	84.18	85.07	85.67	85.97	84.78	88.36	78.51	85.97	87.16	88.36	DNN6
350	84.51	89.44	87.68	91.9	83.1	86.62	88.03	86.27	85.56	91.9	DNN4
400	82.11	91.46	89.43	87.4	90.65	91.46	85.37	84.55	88.21	91.46	DNN6
450	92.13	91.67	93.52	82.87	88.89	86.57	86.57	86.57	87.96	93.52	DNN3
500	91.15	91.15	90.63	91.15	83.85	88.54	84.38	85.94	87.5	91.15	DNN4
550	91.86	83.14	89.53	88.95	84.88	85.47	88.37	87.21	87.21	91.86	DNN1
600	86.54	91.03	85.9	88.46	85.9	87.18	87.82	89.1	87.82	91.03	DNN2
Max	92.13	91.67	93.52	91.9	90.65	91.46	89.08	89.1	88.21	93.52	DNN3
WS	**450**	**450**	**450**	**350**	**400**	**400**	**200**	**600**	**400**	**450**	

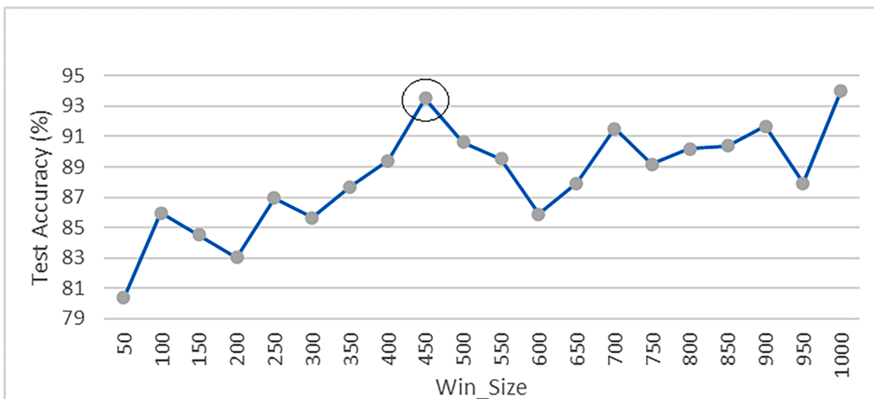

Fig. 3. DNN3 versus window size.

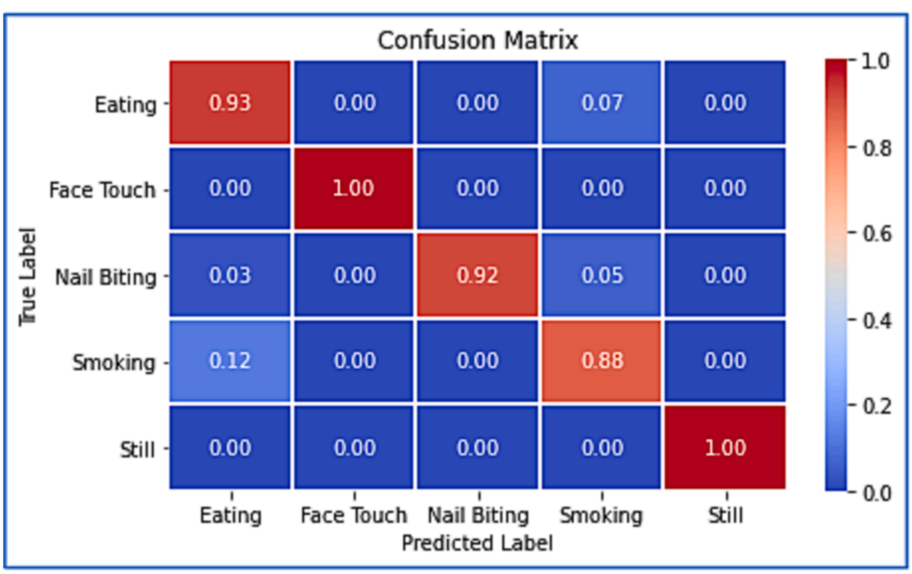

Fig. 4. Confusion matrix of the model DNN3 on test data with sensors AGM.

model could not recognize activities with the expected accuracy, as seen in the test data. For this reason, another model, DNN3, was learned with A and G sensors (the M sensor was skipped) to account for the rolling lag. This model achieved an accuracy of 87.5 % on test data. Although its performance on offline data was inferior compared to the earlier model that used a M and gave a test accuracy of 93.5 %, but due to the delay in its readings, it was not feasible to use it in the real-time system. The confusion matrix of DNN3* on test data is given in Fig. 5. In this figure, the recognition of Face Touch and staying Still is 100 %. However, the model confuses the similar activities of Eating, Nail Biting, and Smoking as the previous model (DNN3 with 93.5 % accuracy) was confusing among the same activities. For instance, the confusion matrix in Fig. 5 shows that Eating was 13 % wrongly classified as Face Touch. Similarly, Smoking was 22 % of the times erroneously classified as Eating and 7 %

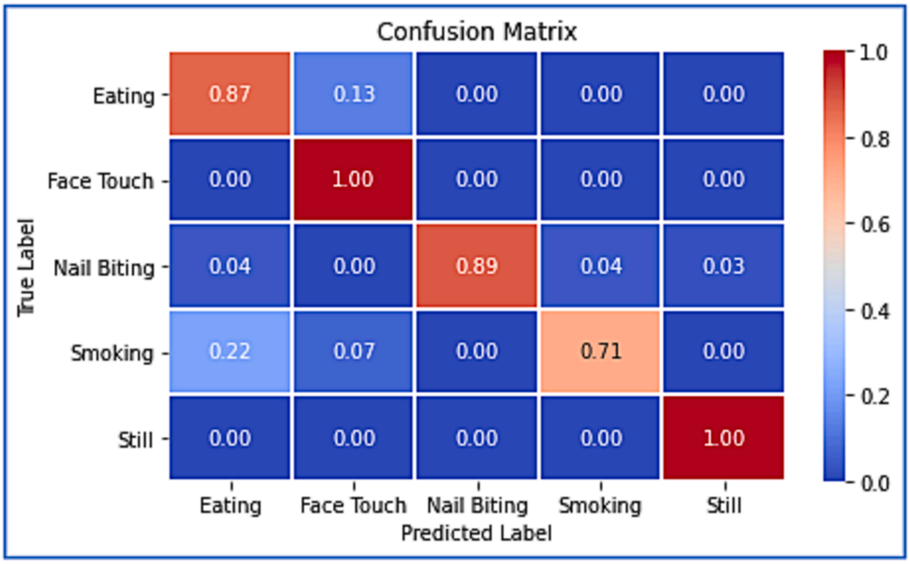

Fig. 5. Confusion matrix of the model DNN3* on test data with sensors AG.

of the times it was confused with Face Touch. Additionally, Nail Biting was 4 % wrongly classified as Face Touch, Smoking and three percent of the times it was wrongly classified as Still position.

The real-time experiments were conducted to evaluate the model in a real-world situation. A set of randomized activities were performed and displayed on a screen for the participants. They were recorded along with the predictions of the model in real-time. Initially, the real-time experiment was performed for a smaller set of activities in which every activity was performed for 40 s. The detail of this experiment is given in Table 4. The first column shows the activities performed by the subject in real-time, the second column displays all the activities recognized by the model (DNN3*), and the third column gives the time taken by the model to recognize the correct activity. The fourth column mentions the activities with which the model was confused, and the last column answers Yes or No if the model correctly recognized the activities. Out of nine activities, the model recognized seven activities correctly, giving an overall accuracy of 77.78 %.

The performance observed on real-time data does not match that achieved on test data, but this inconsistency is understandable. The test data share the same experimental settings as the training data, leading to an anticipated and often observed scenario where the performance on test data surpasses that on real-time data. Moreover, real-time scenarios introduce time constraints, impacting the model's performance evaluation based on how quickly it accurately recognizes activities.

Given that the problem involves time series analysis, the correct classification relies on previous sensor values, and as time passes, better recognition is expected due to additional data availability. Real-time applications always present a challenge in balancing recognition time and accuracy. Achieving higher accuracy requires more time for data accumulation, impacting recognition speed. Conversely, quicker detection may lead to inaccuracies due to limited data for model feeding.

This trade-off should be carefully considered by field experts when determining the time constraints based on the nature of the problem. For instance, in violence detection, minimal recognition time is crucial for prompt actions, while for stress and boredom detection, a more relaxed timeframe is acceptable, as immediate actions are typically not required.

6. Discussion

The current study's exploration into using non-intrusive, widely available smart devices for detecting stress and boredom represents a significant advancement in human emotion detection. Traditionally, emotion detection has heavily relied on physiological signals (e.g., EEG, GSR) captured in controlled environments [47] or on methods that may be considered intrusive, such as video recordings for facial expression analysis [48]. By employing motion sensors in smartphones to recognize activities associated with stress and boredom, this research not only circumvents the need for controlled settings but also addresses privacy concerns associated with video and audio data collection.

A critical aspect of this study is its methodological innovation, particularly the comprehensive exploration of model construction dimensions, which have been somewhat overlooked in the literature [49–51]. The investigation into sensor placement, window sizes, and the comparative analysis of using raw versus processed data offers valuable insights into optimizing model performance. This nuanced approach to model development challenges the existing reliance on adding new sensors or employing computationally expensive algorithms, proposing instead a more resource-efficient pathway to improve detection

Table 4
Summarized detail of the real-time experiment.

Original Activity	Activities recognized by model DNN3*	Time taken (in sec) to recognize correct activity	Model confused with Activities	Correctly recognized? (Y/N)
Still (standing)	Still	0	—	Yes
Smoking	Still, Smoking, Eating	17	Eating	No
Still (sitting)	Eating, Still	5	—	Yes
Eating	Still, Eating	9	Still	Yes
Still (standing)	Eating, Still	4	—	Yes
Face Touch	Still, Face Touch	15	Still	Yes
Still (sitting)	Face Touch, Still	7	—	Yes
Nail Biting	Still, Face Touch	11	Face Touch	No
Still (standing)	Still	5	—	Yes
Accuracy				**77.78 %**

accuracy.

The provision of a benchmark dataset is another cornerstone of this study, addressing a critical gap in the field by facilitating the development of machine learning models for mental health issues beyond stress and boredom [31]. This contribution is pivotal for the research community, providing a foundation for future studies to build upon and extend the application of smart device data in mental health monitoring.

The potential for real-time monitoring and intervention, as demonstrated by the study's findings, underscores the practical implications of this research in mental health care. The development of smart wearable devices capable of generating emergency alerts presents a promising avenue for timely intervention in stress-related crises. However, the challenge of ensuring privacy and data security in such applications remains a critical concern. The balance between effective monitoring and preserving individuals' privacy rights calls for innovative solutions that prioritize ethical considerations in the design and implementation of real-time mental health interventions.

Despite the promising results and contributions of this study, it is crucial to acknowledge its limitations. The reliance on smart devices as the primary data source introduces constraints related to the diversity and accuracy of the detected activities. Future research should aim to integrate additional data sources, such as physiological signals collected via wearable technology, to enrich the dataset and enhance the model's predictive accuracy [5,52]. Moreover, the generalizability of the findings may be limited by the study's sample size and demographic characteristics. Expanding the research to include a broader and more diverse population would help to validate the findings across different contexts and enhance the robustness of the model [53].

The study marks a significant step forward in the application of smart devices for mental health monitoring, offering a detailed approach to detecting stress and boredom. By critically integrating and expanding upon existing literature, the discussion highlights the study's contributions to methodological innovation, real-time monitoring potential, and the challenges of ensuring privacy. As the field progresses, continued exploration and refinement of these methodologies will be essential for harnessing the full potential of technology in mental health care, ultimately leading to more accessible, effective, and personalized interventions.

7. Conclusion

This paper introduces a novel dataset designed to capture stress and boredom-related activities using smartphones, thereby reducing the reliance on expensive gadgets. This dataset offers advantages in terms of privacy and user-friendliness. Additionally, we propose an approach for wrist-worn devices, employing DNN to accurately detect intricate in-home activities associated with stress and boredom. Our method achieves high classification accuracy, particularly for complex activities that are challenging to classify using other algorithms. Furthermore, it performs well in real-time testing. Findings of this have significant implications for applications in smart healthcare, specifically in emotion recognition through HAR. Successful stress identification can lead to earlier diagnosis of stress-related disorders and improve overall mental health and well-being.

The performance disparity between real-time and test data in this study is acknowledged, attributed to the shared experimental settings between test and training data. Real-time scenarios, due to time constraints, affect the model's performance evaluation, with a trade-off between recognition time and accuracy. Balancing this trade-off is crucial, especially considering the nature of the problem; for instance, violence detection requires minimal recognition time, while stress and boredom detection allow for a more relaxed timeframe.

The study demonstrates the practicality of using smart devices and machine learning for recognizing behavioral markers associated with stress and boredom. The promising results suggest the extension of this approach to identify markers for other mental health issues, addressing the challenge of insufficient recognition and consideration in mental health diagnoses. In situations where societal attitudes and stigma hinder open discussions about mental health, smart devices offer an accessible means for self-assessment, eliminating the need for immediate visits to psychologists. Additionally, by enhancing mobile applications to provide appropriate responses and facilitating real-time connections between individuals and psychologists, timely interventions and support can be offered, especially for severe cases.

8. Limitations and future directions

The exploration of human activity recognition through wearables in the context of mental health is a burgeoning field that is still in its early stages. One notable constraint is the scarcity of available datasets within this specific domain, posing a considerable challenge to accessibility. Consequently, we undertook the task of compiling our own dataset, encountering resource limitations in the process. Our dataset primarily encompasses a restricted range of activities related to stress and boredom.

Faced with challenges in data collection, we opted to use smartphones positioned at the wrist due to the absence of specialized smart devices. Additionally, the study's scope is currently constrained in terms of both the number of activities and participants. Our intention is to broaden the study's scope by incorporating a more diverse range of activities, participants from various backgrounds, and a wider array of devices. This expansion is designed to provide a comprehensive understanding of the genuine applicability and effectiveness of our methodology in the realm of mental health research.

Furthermore, we recognize the importance of broadening participant inclusion to encompass diverse backgrounds, demographics, and ethnicities. This diversification is crucial for enhancing the generalizability of our study and mitigating biases inherent in a more limited participant pool.

As part of our future endeavors, we plan to record activities in a more natural environment. Participants will be encouraged to carry out their daily routines, with their activities recorded to capture the nuances of behavior in a genuine setting. This approach aims to enrich the dataset with more naturalistic data, enabling the model to be trained on a more representative sample.

Moving forward, we intend to explore advanced deep learning models, including but not limited to Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and combinations such as CNN-LSTM and ConvLSTM. By subjecting these models to our expanded dataset, we seek to assess their performance and identify the most suitable architecture for our research objectives. We also intend to explore the integration of physiological sensors like heart rate and skin conductance, along with geolocation data, to create an emotion map using our model on mobile devices. Context awareness plays a pivotal role in robust, pervasive, and wearable computing. Furthermore, given our devices are battery-powered, we aimed to investigate low computational complexity and power-efficient detection methods.

9. Summary Table

9.1. What is already known on the topic

- The use of smart devices for Human Activity Recognition (HAR) shows great potential in applications such as sports, fitness, and remote health monitoring.
- This process requires the analysis of time series data, for which deep learning algorithms are highly effective in identifying various activities.
- It's crucial to carefully design the study and adjust parameters like window size and the choice of sensors, taking into consideration the specific data and activities being monitored.

10. What this study added to our knowledge

- This study investigates the potential applications of HAR and smart devices in the realm of mental health and introduces a distinctive dataset designed for identifying stress and boredom indicators.
- The CNN-LSTM model has shown significant effectiveness in identifying these indicators through the use of standard inertial sensors available in smart devices.
- The capability for real-time activity detection demonstrates the method's relevance where immediate feedback is essential. Such outcomes indicate broad possibilities for combining wearable technology with the diagnosis and ongoing observation of mental health.

Funding

This research received no specific grant from any funding agency in the commercial, or not-for-profit sectors. Dr. Gulnaz Anjum's participation in this project was supported by the Department of Psychology (PSI), University of Oslo (UiO). The funding was provided by Tildeling av kvalifiseringsstipend 2022, sub-project number: 102603241.

CRediT authorship contribution statement

Nida Saddaf Khan: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Saleeta Qadir:** Writing – original draft, Visualization, Validation, Software, Resources, Data curation. **Gulnaz Anjum:** Writing – review & editing, Supervision, Project administration. **Nasir Uddin:** Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

Dr Muhammad Sayeed Ghani offered outstanding support and guidance throughout a research related to this research paper. This project greatly benefited from his previous mentorship and involvement.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2024.105401.

References

[1] K. Chen, D. Zhang, L. Yao, B. Guo, Z. Yu, and Y. Liu, "Deep Learning for Sensor-based Human Activity Recognition: Overview, Challenges, and Opportunities," *ACM Comput. Surv.*, vol. 54, no. 4, p. 77:1-77:40, May 2021, doi: 10.1145/3447744.

[2] N.S. Khan, M.S. Ghani, G. Anjum, ADAM-sense: anxiety-displaying activities recognition by motion sensors, Pervasive Mob. Comput. (Nov. 2021), https://doi.org/10.1016/j.pmcj.2021.101485.

[3] X. Li, Y. Zhang, I. Marsic, A. Sarcevic, R.S. Burd, "Deep learning for RFID-based activity recognition", in proceedings of the 14th ACM conference on embedded network sensor systems CD-ROM, in SenSys '16, Association for Computing Machinery, New York, NY, USA, Nov. 2016, pp. 164–175, https://doi.org/10.1145/2994551.2994569.

[4] R. Grzeszick, J. M. Lenk, F. M. Rueda, G. A. Fink, S. Feldhorst, and M. ten Hompel, "Deep Neural Network based Human Activity Recognition for the Order Picking Process," in Proceedings of the 4th international Workshop on Sensor-based Activity Recognition and Interaction, Rostock Germany: ACM, Sep. 2017, 1–6. 10.1145/3134230.3134231.

[5] L. Rachakonda, S.P. Mohanty, E. Kougianos, P. Sundaravadivel, Stress-lysis: a DNN-integrated edge device for stress level detection in the IoMT, IEEE Trans. Consum. Electron. 65 (4) (Nov. 2019) 474–483, https://doi.org/10.1109/TCE.2019.2940472.

[6] Y. Shibasaki, K. Funakoshi, and K. Shinoda, "Boredom Recognition Based on Users' Spontaneous Behaviors in Multiparty Human-Robot Interactions," in MultiMedia Modeling, L. Amsaleg, G. Þ. Guðmundsson, C. Gurrin, B. Þ. Jónsson, and S. Satoh, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017, pp. 677–689. doi: 10.1007/978-3-319-51811-4_55.

[7] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan, "Deep Activity Recognition Models with Triaxial Accelerometers," ArXiv151104664 Cs, Oct. 2016, Accessed: Jun. 24, 2021. [Online]. Available: http://arxiv.org/abs/1511.04664.

[8] F.J. Ordóñez, D. Roggen, Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition, Sensors 16, no. 1, Art. no. 1 (2016), https://doi.org/10.3390/s16010115.

[9] B. Mezuk, et al., Stress, self-regulation, and context: evidence from the health and retirement survey, SSM - Popul. Health 3 (Dec. 2017) 455–463, https://doi.org/10.1016/j.ssmph.2017.05.004.

[10] A.S. Mathew, T.P. Davine, I. Snorrason, D.C. Houghton, D.W. Woods, H.-J. Lee, Body-focused repetitive behaviors and non-suicidal self-injury: a comparison of clinical characteristics and symptom features, J. Psychiatr. Res. 124 (May 2020) 115–122, https://doi.org/10.1016/j.jpsychires.2020.02.020.

[11] M. Zeng *et al.*, "Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors," in Proceedings of the 6th International Conference on Mobile Computing, Applications and Services, Austin, United States: ICST, 2014. doi: 10.4108/icst.mobicase.2014.257786.

[12] H. Cho and S. M. Yoon, "Divide and Conquer-Based 1D CNN Human Activity Recognition Using Test Data Sharpening," Sensors, vol. 18, no. 4, Art. no. 4, Apr. 2018, doi: 10.3390/s18041055.

[13] M.M. Hassan, M.Z. Uddin, A. Mohamed, A. Almogren, A robust human activity recognition system using smartphone sensors and deep learning, Future Gener. Comput. Syst. 81 (Apr. 2018) 307–313, https://doi.org/10.1016/j.future.2017.11.029.

[14] J. Ryu, J. Seo, H. Jebelli, S. Lee, Automated action recognition using an accelerometer-embedded wristband-type activity tracker, J. Constr. Eng. Manag. 145 (1) (Jan. 2019) 04018114, https://doi.org/10.1061/(ASCE)CO.1943-7862.0001579.

[15] M. Shoaib, S. Bosch, O.D. Incel, H. Scholten, P.J.M. Havinga, Complex human activity recognition using smartphone and wrist-worn motion sensors, Sensors 16, no. 4, Art. no. 4 (Apr. 2016), https://doi.org/10.3390/s16040426.

[16] G. Bhat, Y. Tuncel, S. An, H. G. Lee, and U. Y. Ogras, "An Ultra-Low Energy Human Activity Recognition Accelerator for Wearable Health Applications," ACM Trans. Embed. Comput. Syst., vol. 18, no. 5s, p. 49:1-49:22, Oct. 2019, doi: 10.1145/3358175.

[17] N.S. Khan, M.S. Ghani, A survey of deep learning based models for human activity recognition, Wirel. Pers. Commun. 120 (2) (Sep. 2021) 1593–1635, https://doi.org/10.1007/s11277-021-08525-w.

[18] Q. Li, R. Gravina, and G. Fortino, "Posture and Gesture Analysis Supporting Emotional Activity Recognition," in *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Oct. 2018, pp. 2742–2747. doi: 10.1109/SMC.2018.00468.

[19] N.Y. Hammerla, S. Halloran, T. Plotz, "Deep, convolutional, and recurrent models for human activity recognition using wearables", in twenty-fifth international joint conference on artificial intelligence (IJCAI-16), New York, AAAI Press, USA, 2016.

[20] H. Amroun, M. Temkit, and M. Ammi, "DNN-Based Approach for Recognition of Human Activity Raw Data in Non-Controlled Environment," in 2017 IEEE International Conference on AI Mobile Services (AIMS), Jun. 2017, pp. 121–124. doi: 10.1109/AIMS.2017.26.

[21] T. Steels, B. Van Herbruggen, J. Fontaine, T. De Pessemier, D. Plets, E. De Poorter, Badminton activity recognition using accelerometer data, Sensors 20, no. 17, Art. no. 17 (Jan. 2020), https://doi.org/10.3390/s20174685.

[22] P. Vepakomma, D. De, S. K. Das, and S. Bhansali, "A-Wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities," in 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN), Jun. 2015, pp. 1–6. doi: 10.1109/BSN.2015.7299406.

[23] M. Abreu, M. Barandas, R. Leonardo, and H. Gamboa, "Detailed Human Activity Recognition based on Multiple HMM," Jan. 2019, pp. 171–178. doi: 10.5220/0007386901710178.

[24] C. Cole, B. Janos, D. Anshari, J. Thrasher, S. Strayer, H. Valafar, "Recognition of smoking gesture using smart watch technology", Int. Conf. Health Inform. Med. Syst., (2016) [Online]. Available: https://scholarcommons.sc.edu/csce_facpub/208.

[25] T.-H. Tan, J.-Y. Shih, S.-H. Liu, M. Alkhaleefah, Y.-L. Chang, M. Gochoo, Using a hybrid neural network and a regularized extreme learning machine for human activity recognition with smartphone and smartwatch, Sensors 23 (6) (2023) 3354.

[26] H. Wu, et al., A novel pedal musculoskeletal response based on differential spatio-temporal LSTM for human activity recognition, Knowl.-Based Syst. 261 (2023) 110187.

[27] F. Anjum, S. Alam, E.H. Bahadur, A.K.M. Masum, M.Z. Rahman, Deep learning for depression symptomatic activity recognition, in: Presented at the 2022 International Conference on Innovations in Science, Engineering and Technology (ICISET), IEEE, 2022, pp. 510–515.

[28] F.M. Talaat, R.M. El-Balka, Stress monitoring using wearable sensors: IoT techniques in medical field, Neural Comput. Appl. (2023) 1–14.

[29] P. Sa-Nguannarm, E. Elbasani, J.-D. Kim, Human activity recognition for analyzing stress behavior based on Bi-LSTM, Technol. Health Care no. Preprint (2023) 1–11.

[30] P. Schmidt, A. Reiss, R. Dürichen, K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection", in ICMI 2018,

Boulder, F. Wolling, USA, 2018 [Online]. Available: https://ubicomp.eti.uni-siegen.de/home/datasets/icmi18/.

[31] S. Qadir, N.S. Khan, G. Anjum, N. Uddin, "StresSense: Dataset for Stress and Boredom-Displaying Activities" vol. 1 (2023) Dec, https://doi.org/10.17632/2dn3hpbm5m.1.

[32] "ADAM-Sense: Anxiety-Displaying Activities Recognition by Motion Sensors." Mendeley Data, Mar. 11, 2021. [Online]. Available: https://data.mendeley.com/datasets/6g6pxwjj48/draft?a=cd0c1648-3ee8-4d69-8210-c51ae57aced9.

[33] M. Shoaib, H. Scholten, P.J.M. Havinga, Towards physical activity recognition using smartphone sensors, in: In 2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing, 2013, pp. 80–87, https://doi.org/10.1109/UIC-ATC.2013.43.

[34] B. Bhandari, null JianChao Lu, null Xi Zheng, S. Rajasegarar, and C. Karmakar, "Non-invasive sensor based automated smoking activity detection," Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf., vol. 2017, pp. 845–848, Jul. 2017, doi: 10.1109/EMBC.2017.8036956.

[35] S. Zhang, W. Stogin, N. Alshurafa, I sense overeating: motif-based machine learning framework to detect overeating using wrist-worn sensing, Inf. Fusion 41 (May 2018) 37–47, https://doi.org/10.1016/j.inffus.2017.08.003.

[36] A. Sachan and T. P. Chaturvedi, "Onychophagia (Nail biting), anxiety, and malocclusion," Indian J. Dent. Res. Off. Publ. Indian Soc. Dent. Res., vol. 23, no. 5, pp. 680–682, Oct. 2012, doi: 10.4103/0970-9290.107399.

[37] M. Grunwald, T. Weiss, S. Mueller, L. Rall, EEG changes caused by spontaneous facial self-touch may represent emotion regulating processes and working memory maintenance, Brain Res. 1557 (Apr. 2014) 111–126, https://doi.org/10.1016/j.brainres.2014.02.002.

[38] X. "Anthony" Chen, "FaceOff: Detecting Face Touching with a Wrist-Worn Accelerometer," ArXiv200801769 Cs, Aug. 2020, Accessed: Jun. 25, 2021. [Online]. Available: http://arxiv.org/abs/2008.01769.

[39] M. Gjoreski, H. Gjoreski, M. Luštrek, and M. Gams, "Recognizing atomic activities with wrist-worn accelerometer using machine learning," p. 4.

[40] H. Gjoreski, J. Bizjak, M. Gjoreski, and M. Gams, "Comparing Deep and Classical Machine Learning Methods for Human Activity Recognition using Wrist Accelerometer," p. 7.

[41] G. Bhat, V. V. Chaurasia, H. Shill, and U. Y. Ogras, "Online Human Activity Recognition using Low-Power Wearable Devices," in IEEE/ACM International Conference on Computer-Aided Design (ICCAD), San Diego, CA: IEEE, 2018, pp. 1–8. Accessed: Mar. 30, 2020. [Online]. Available: https://www.researchgate.net/publication/327260776_Online_Human_Activity_Recognition_using_Low-Power_Wearable_Devices.

[42] S. A. Rokni, M. Nourollahi, and H. Ghasemzadeh, "Personalized Human Activity Recognition Using Convolutional Neural Networks," in arXiv:1801.08252 [cs], Jan. 2018. Accessed: Apr. 03, 2020. [Online]. Available: http://arxiv.org/abs/1801.08252.

[43] T. R. Mauldin, M. E. Canby, V. Metsis, A. H. H. Ngu, and C. C. Rivera, "SmartFall: A Smartwatch-Based Fall Detection System Using Deep Learning," Sensors, vol. 18, no. 10, Art. no. 10, Oct. 2018, doi: 10.3390/s18103363.

[44] Y. Zhang, Y. Zhang, Z. Zhang, and Y. Song, "Human activity recognition based on time series analysis using U-Net," p. 21.

[45] J. O'Halloran and E. W. J. Curry, "A Comparison of Deep Learning Models in Human Activity Recognition and Behavioural Prediction on the MHEALTH Dataset," presented at the Irish Conference on Artificial Intelligence and Cognitive Science, 2019. Accessed: Oct. 02, 2023. [Online]. Available: https://www.semanticscholar.org/paper/A-Comparison-of-Deep-Learning-Models-in-Human-and-O%E2%80%99Halloran-Curry/04353e288c5be638aa30a1218ac180ae6ce770ce.

[46] Y. Bengio, "Learning Deep Architectures for AI," Found. Trends® Mach. Learn., vol. 2, no. 1, pp. 1–127, Jan. 2009, doi: 10.1561/2200000006.

[47] O. AlShorman, et al., Frontal lobe real-time EEG analysis using machine learning techniques for mental stress detection, J. Integr. Neurosci. 21 (1) (2022) 20.

[48] J.R. Saadon, et al., Real-time emotion detection by quantitative facial motion analysis, Plos One 18 (3) (2023) e0282730.

[49] F. Wahle, T. Kowatsch, E. Fleisch, M. Rufer, S. Weidt, Mobile sensing and support for people with depression: a pilot trial in the wild, JMIR MHealth UHealth 4 (3) (Sep. 2016) e111.

[50] R.S. McGinnis, et al., Rapid detection of internalizing diagnosis in young children enabled by wearable sensors and machine learning, PLOS ONE 14 (1) (Jan. 2019) e0210267.

[51] J.T. O'Brien, et al., A study of wrist-worn activity measurement as a potential real-world biomarker for late-life depression, Psychol. Med. 47 (1) (Jan. 2017) 93–102, https://doi.org/10.1017/S0033291716002166.

[52] F. Demrozi, G. Pravadelli, A. Bihorac, P. Rashidi, Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey, IEEE Access 8 (2020) 210816–210836, https://doi.org/10.1109/ACCESS.2020.3037715.

[53] J. Klenk, et al., The FARSEEING real-world fall repository: a large-scale collaborative database to collect and share sensor signals from real-world falls, Eur. Rev. Aging Phys. Act. 13 (1) (Oct. 2016) 8, https://doi.org/10.1186/s11556-016-0168-9.

[54] N. S. Khan, S. Ghani, and S. Haider, "Real-Time Analysis of a Sensor's Data for Automated Decision Making in an IoT-Based Smart Home," *Sensors*, vol. 18, no. 6, Art. no. 6, Jun. 2018, doi: 10.3390/s18061711.

[55] S. Xu, et al., Emotion Recognition From Gait Analyses: Current Research and Future Directions, IEEE Trans. Comput. Soc. Syst. 11 (1) (Feb. 2024) 363–377, https://doi.org/10.1109/TCSS.2022.3223251.

[56] L. Shu, et al., A Review of Emotion Recognition Using Physiological Signals, Sensors 18 (7) (Jun. 2018) 2074, https://doi.org/10.3390/s18072074.