Stress Prediction Using Machine Learning and IoT

Vividha, Drishti Agarwal, Paras Gupta, Soham Taneja, Preeti Nagrath, and Bhawna Gupta

Abstract Stress is a mental condition that affects every aspect of life leading to sleep deprivation and various other diseases. Thus, it's necessary to analyse one's vitals to stay updated about one's mental health. This paper presents an effective method for detecting cognitive stress levels using data from a physical activity tracker device. The main goal of this system is to use sensor technology to detect stress using a machine learning approach. Individually, the impact of each stressor is assessed using ML models, followed by the construction of a NN model and assessment using ordinal logistic regression models such as logit, probit and complementary log-log. The paper uses heartbeat rate as one of the features to recognise stress and the Internet of Things (IoT) and Machine Learning (ML) to alert the situation when the person is in real danger. The patient's condition is predicted using machine learning and the patient's acute stress condition is relayed using IoT. Based on the heartbeat, a prediction of whether a person is under stress or not can be made. The paper presents a model that can predict stress levels based entirely on electrocardiogram (ECG) data, which can be measured with consumer-grade heart monitors The ECG's spectral power components, as well as time and frequency domain features of heart rate variability, are included in the model. The stress detector system takes the real-time data from the IoT device (sensor), then applies a machine learning model on the data to detect

Vividha (✉) · D. Agarwal · P. Gupta · S. Taneja · P. Nagrath
Bharati Vidyapeeth's College of Engineering, New Delhi, India
e-mail: vividha.cse1@bvp.edu.in

D. Agarwal
e-mail: drishtiagarwal.cse1@bvp.edu.in

P. Gupta
e-mail: parasgupta.cse1@bvp.edu.in

S. Taneja
e-mail: sohamtaneja.cse1@bvp.edu.in

P. Nagrath
e-mail: preeti.nagrath@bharatividyapeeth.edu

B. Gupta
Panipat Institute of Engineering and Technology, Samalkha, India

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

D. Gupta et al. (eds.), *Proceedings of Data Analytics and Management*, Lecture Notes on Data Engineering and Communications Technologies 91,
https://doi.org/10.1007/978-981-16-6285-0_49

stress levels in an individual and eventually informs/alerts the individual about their stress condition. By collecting data, this system is tested and evaluated in a real-time environment with different machine learning models. Finally, a comprehensive comparative analysis has been depicted amongst the applied models with Random Forest Classifier showing the highest accuracy. The novelty of this work lies in the fact that a stress detection framework should be as unobtrusive to the user as possible.

Keywords Electrocardiography (ECG) · Electromyography (EMG) · Galvanic skin response (GSR) · Heart rate (HR) · Internet of Things (IoT) · k-nearest neighbours (KNN) · Physionet · Random forest · Support vector machine (SVM)

1 Introduction

Stress in the twenty-first century has emerged as an integrated section of our everyday life and is a prominent notion in the healthcare sector. Nowadays, Stress has become an inherent part of everyone's life, generally in today's competitiveness of this generation. In the workplace, every person continuously faces several situations, like excessive work pressure, job insecurity, poor job satisfaction and the pressure of staying updated. The constant stress pressure can cause multiple negative health effects, which can include high blood pressure, sleep deprivation, vulnerability to infections and cardiovascular diseases. All these conditions lead to mental stress, which is the prime cause of many diseases these days. Such unfortunate effects not only affect a person's health and well-being but also result in low productivity and overall profit. Stress can be detected using data and sensors like accelerometer, keystroke dynamics, or blinking.

Generally, an ensemble of several markers is used while compromising with an increased budget and user involvement. New non-contact methods have also been discovered to quantify stress, which includes hyperspectral imaging techniques, human voice, pupil diameter, visible spectrum camera, or using stereo thermal and visible sensors.

The objective of this research are:

- To integrate the ML techniques with the IoT devices along with web deployment on the cloud servers.
- To achieve state-of-the-art performance by the project so that it can be used for stress detection accurately even in real-life scenarios.
- To analyse the results and compare the performance and accuracy of multiple ML classification algorithms.

The paper is composed of V sections. Section I being the introduction followed by Section II, the related works in the field of stress detection. In section III, the research methodology is explained thoroughly and is further divided into subsections like Data collection and preprocessing, Feature Selection, Classification Model and IoT device

architecture. In Section IV, the results are mentioned and analysed. Finally, Section V concludes the paper and discusses future work in the given field.

2 Related Surveys

Owing to its increased importance in contemporary society, human stress has become an occasional subject in research papers (Table 1). There has been extended research work in terms of tackling stress and mental health issues in everyday life with the help of IoT and machine learning applications. [1] uses Theano in which stress is detected using the position of the eyebrow from its mean position. Many studies have been performed to determine the exact facial features associated with depression. A depressed/stressed face has the same features as a sad face that can be identified by the SVM classifier [2].

Variability of the heart rate refers to variations in the beat of the heart rate. It can be predicted, based on heart rate, whether or not a person is stressed. To evaluate the stress level of any person using his/her heart rate, an IoT system called Remote Stress Detector is used in [3]. Stress can also be detected by using sensors to measure body temperature and skin conductance, evaluating the degree of stress [4]. Signals like ECG and breathing rate are calculated with the aid of various wearable sensors [5].

Lakudzode and Rajbhoj [6] introduces a mobile phone emotion recognition device that can be used as a smart keyboard. With ML techniques, the smart keyboard senses the emotional state of an individual. [7] discusses Health Monitoring and Prognosis wearable sensor-based systems; compares different systems to address the flaws of the present biosensor networks, & offers guidance and direction for upcoming unobtrusive ways. In [8], Heart Rate Variability (HRV) trends are examined by normal healthy people during sleep. This is done for stress awareness. It is understood that HRV is an indication of the development of the autonomic nervous system. The modification of the beat-to-beat (RR) intervals represents HRV.

Minguillon et al. [9] proposes a three-level approach (stress, relaxation and neutrality) to detect stress with a few-second resolution and 86% accuracy. Both hardware, as well as the software, were installed in their laboratory, they produce a method called RABio w8 (real-time detection of biosignals, wireless, eight channels).

Three blocks (acquisition block, control block and communication block) form RABio technology. They did a study that included ten volunteer studies to test their approach. Normal relative gamma (RG), heart rate (HR), normal skin conduction (SC) and normal trapezia movement are symptoms of depression used here (TA).

Table 1 Literature survey

Paper title	Theoretical/conceptual framework	Methodology	Analysis and results
Detection of stress using image processing and machine learning techniques [1]	A real-time non-intrusive video is captured, which detects the emotional status of a person by analysing the facial expression	Employs a technique that allows to train a model and analyse differences in predicting the features. Theano is a python framework that aims at improving both the execution time and development time of the linear regression model which is used here as a deep learning algorithm	The experimental results show that the developed system is well on data with the generic model of all ages
Machine learning and IoT for prediction and detection of stress [3]	The developed prototype detects whether a person is under stress using variability in his/her heart rate	The heartbeat readings from a well-calibrated end IoT device are pushed to the server where they are filtered using a user's network id to keep track of readings for a particular individual. The data is applied with various machine learning models for detection	Stress labels are calculated and the median in our data set was stressed and map that to our detector, it is not possible to detect someone's age by their resting heart rate—while some correlations exist, the relationship is not clearly defined. The applied SVM and Logistic Regression show considerable improvement over VF—15 and Naive Bayes
Portable system for real-time detection of stress level [9]	Present and validate a portable system for real-time detection of stress level, based on the RABio w8 (real-time acquisition of biosignals, wireless, eight channels) system	The hardware was made of portable, wireless and low-cost electronics. The software was composed of an application programming interface (API) and a graphical user interface (GUI)	The classification accuracy or probability of success (pain the in the detection of stress level (stress, relax and neutral) and the 95% confidence interval were reported

3 Research Methodology

The research methodology consists of various stages including a collection of data and preprocessing, feature selection, classification using stress detection using Random Forest Classifier and the construction of IoT devices (Fig. 1).

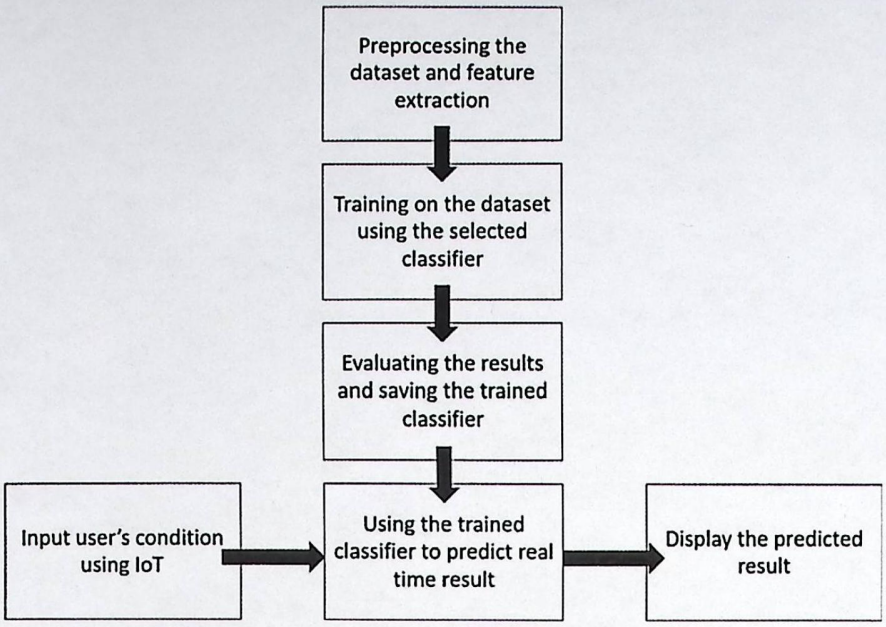

Fig. 1 Project workflow

1. **Data collection and preprocessing**

The database is in the form of a specific Physionet format which is divided into Eighteen.dat files & Eighteen.hea files having the corresponding metadata [10]. This data was made up of signals from the foot such as ECG, EMG and GSR steps from the hand, HR and Respiration. All variables are floating-point numerics, with a sampling frequency equal to 15.5 per second. To clean the database we use software created by Physionet called WFDB to format the files into comma-separated values with their column names. After that, the heart rate variation is based on the RR gaps by running a custom python programme called the WFDB system.

Heart rate variations can be determined from a 20-s instantaneous window for the closest RR interval, with 10 RR intermediate samples from the preceding immediate window and 5 RR intermediate samples fetched from the succeeding immediate window. This leads to a variable section at a time having a full size of 30 s. and prevents cutting hard using split windows.

This method can insert any events towards the end or at the beginning of windows, to determine multiple RR peaks when calculating HR variables, simultaneously resulting in a database consisting of more samples instead of the case where the windows were, e.g. 30 s without splitting. The cleaned database is stored in a.csv file.

Data enhancement ensures that binary numbers labels are 1 (stressed) or 0 (free) for ease and accuracy in categories. We erase all the lines manually and

then fill in the missing values using the inf value of inf. The heartbeat columns are further cleaned with a central filter, this is done to remove reading errors.

2. **Feature Selection**

   The galvanic skin response features are not considered here which were used to label data, as well as ECG along with EMG data were denied as these details won't be easily and simply available on wearable devices in the consumer market directly. Also, the ultra-low frequency (ULF) fraction is not considered as most values are nill, which is caused due to the short intermittent periods of the signals. Also, the low frequency band (VLF) data is rejected as a study from various researchers found that the lowest frequency band proves to be inefficient in less than 5 min of reading and also that the samples that are in the dataset are inconsistent.

   The feature set now contains the Average of all NN intervals (**AVNN), Ratio of Low to High (LFHF), Standard Deviation of NN intervals (SDNN), Percentage Difference of NN intervals (pNN50), Total spectral Power (TP), Root Mean Square of standard Deviation of difference between NN intervals (RMSSD), Low Frequency, RR intervals, Heart Rate, High Frequency**. In 4132 samples it was necessary to exclude cardiac variability factors as the RR intervals and heart rate is the obvious values from a 30-s interval.

3. **Classification Model**

   Machine Learning is applied for classification based on the features selected. Several ML classifiers were implemented to get results. Out of all the classifiers, Random Forest Classifier gave the best accuracy. It is applied to detect whether a person is under stress using physical attributes such as skin conductance, EMG and ECG.

Random Forest Classifier

Random Forest Classifier (RF) is a guided machine algorithm used for partitioning, retreating and other operations using decision trees. Random forest planning creates a set of decision trees from a randomly selected set of training sets. A collection of deciduous trees (DT) from a randomly selected training set and Collect votes on various decision trees to determine the final prediction.

The Random Forest Classifier, as its name states, has hundreds of tropical trees that serve as a team. Every tree in a random forest event explains the forecast of the category and the label having the majority of the votes becomes the output of this RF model. The basic concept of a random forest is simple yet powerful—the wisdom of the crowds. In terms of detail, the reason why the informal forest model works distinctively is that the large number of unrelated species (trees) that function as a jury will outweigh any other species. Low integration between models is key. As small-scale investments (such as stocks and bonds) come together to create a portfolio greater than its total assets, unconventional models can produce an increased number of accurate forecasts in contrast to any other prediction. This positive effect is observed because the trees prevent one another from their flaws (as long as they do not deviate in the same way).

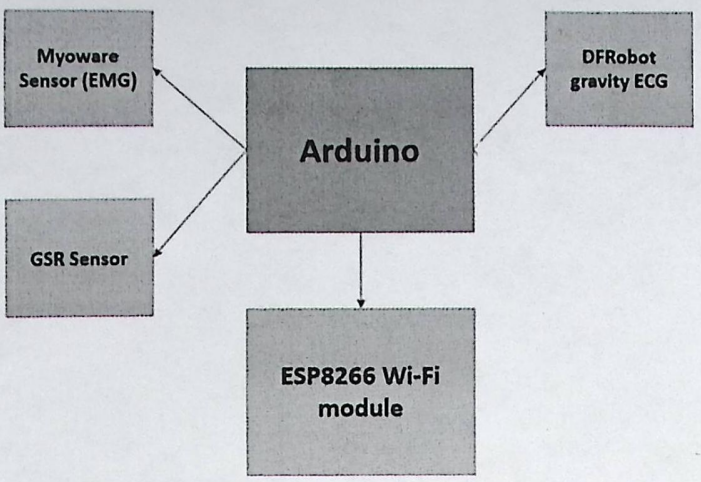

Fig. 2 Device architecture (block diagram)

While some trees may not be good, many other trees will be good, so as a group the trees can move in the right direction. Therefore, the requirements for a random forest to function properly are required to have a certain trait in our features so that models built using those features perform better than random guesses (so mistakes) made by individual trees need to be a fusion of each other (Fig. 2).

3.1 IoT Device Architecture

The device is based on the basic principles of IoT, involving sufficient use of sensors that will measure the required vitals of our users [11] (Fig. 3).

The backbone of any IoT based device is the microcontroller which connects and implements all these sensors together and makes them simultaneously. Arduino is the most effective hardware-software platform to work upon devices involving sensors. Along with this, the data will be sent to a cloud database with the help of the ESP8266 Wi-Fi module which will provide internet access to our Arduino board for data transfer.

4 Results

The Random Forest classifier achieved an accuracy of about 0.735. Figure 4 shows the test results where the algorithm successfully detected the stress, based on the heart rate feature over a particular interval of time which eventually derives the required Heart Rate Variability. As the Heart Rate shows peaks values the algorithm

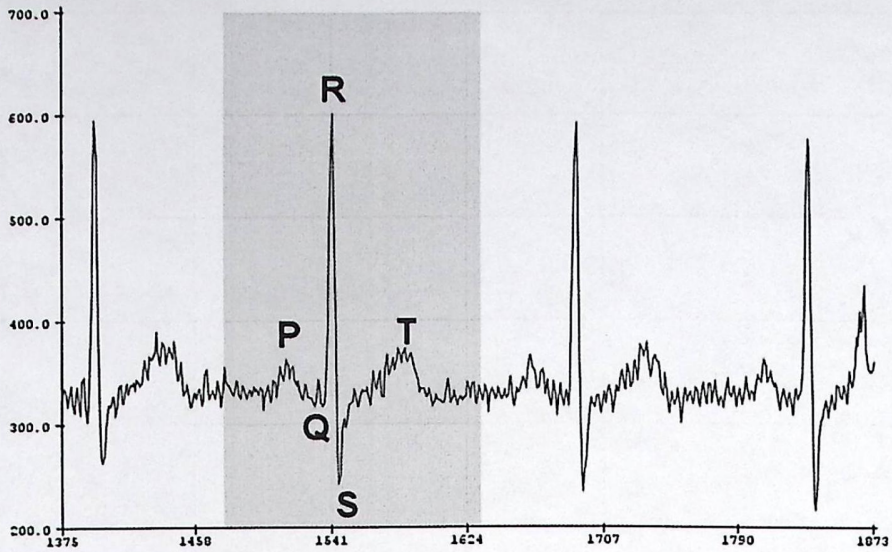

Fig. 3 ECG Generated waveform

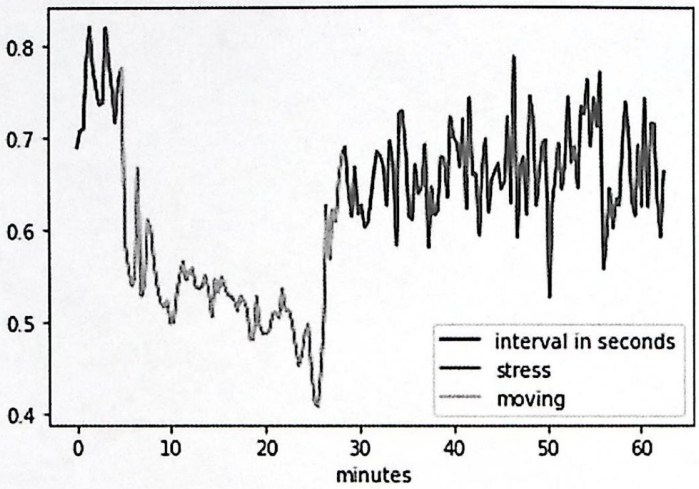

Fig. 4 Graph depicting the relation between HRV and stress of random forest classifier

detects a stressed condition (represented as 1). Thus, establishing a relation between HRV and Stress. The study compared the accuracy of different classifiers which consisted of SVM, KNN and Random Forest Classifier. On comparing the accuracy of Random Forest Classifier was found to be the best with an accuracy of 0.735 (approx.). Figures 5 and 6 show the various Graphs obtained on plotting the HRV and time interval using different classifiers.

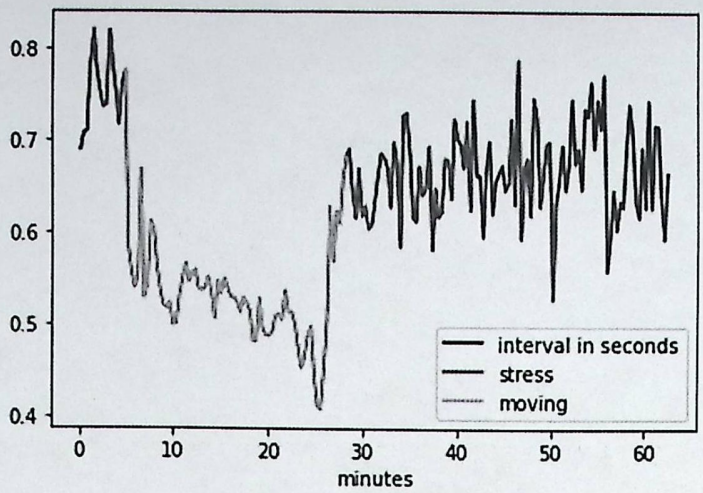

Fig. 5 Graph depicting the relation between HRV and stress of SVM classifier

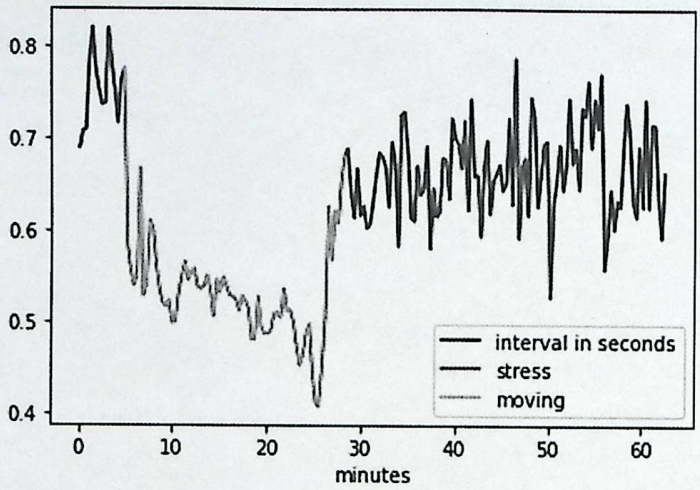

Fig. 6 Graph depicting the relation between HRV and stress of KNN classifier

It can be seen that (Table 2) the RF Classifier outperforms the other ML models with an accuracy of 73.5% as compared to the SVM classifier with an accuracy of 62.2%(approx.) and the KNN Classifier with an accuracy of 73%. Hence, it can be inferred that tree-based ensembling algorithms are suitable for the task of stress prediction.

Table 2 Accuracies of different classifiers used

Models	Accuracy
RF	0.7348668280871671
SVM	0.6222760290556901
KNN	0.7300242130750605

5 Conclusion and Future Scope

In this research study, using IOT device architecture and Machine learning classification algorithms, a device to detect stress was studied and implemented. For the stress detection model, a combined framework using ML and IoT was proposed. The classification of stress using the user's physical attributes such as ECG, heart rate, skin conductance was achieved using a Random Forest classifier. For the physical attributes, IoT device architecture was established with different end-points for different features of the user to detect stress.

For future scope, the research would extend to stress in specific audiences such as maternal stress or stress in students due to different cognitive approaches. Along with this, to ensure greater accuracy in real-world cases, a two-level stress detection approach would be implemented which would take into account the facial and visual aspects of the user as a feature using deep learning-based image classification techniques such as Convolutional Neural Networks (CNNs).

References

1. Raichur N, Lonakadi N, Mural P (2017) Detection of stress using image processing and machine learning techniques. Int J Eng Technol 9(3)
2. Venkataraman D, Parameswaran NS (2018) Extraction of facial features for depression detection among students. Int J Pure Appl Math
3. Pandey PS (2017) Machine learning and IoT for prediction and detection of stress, IEEE
4. Kumar T, Kumar RS et al (2019) Health monitoring and stress detection system, IRJET, March 2019
5. Hesham, Elkhorazaty Y, Aloul F (2016) Emotion recognition using mobile phones, IEEE
6. Lakudzode SF, Prof. Rajbhoj SM (2016) Review on human stress monitoring system using wearable sensors
7. Pantelopoulos A, Bourbakis NG (2009) A survey on wearable sensor-based systems for health monitoring and prognosis. In: IEEE transactions on systems, man, and cybernetics, part c (applications and reviews); vol 40, issue 1
8. Amir Muaremi; Bert Arnrich; Gerhard Troster; Towards Measuring Stress with Smartphones and Wearable Devices During Workday, Sleep. Zurich, Switzer- land. 2013
9. Minguillon J, Perez E, Lopez-Gordo MA, Pelayo F, Sanchez-Carrion MJ (2018) Portable system for real-time detection of stress level. Sensors 18(8):2504
10. Healey J (2005) Wearable and automotive systems for affect recognition from physiology
11. Healey J, Picard W (2000) Wearable and automotive systems for affect recognition from physiology