Machine Learning and IoT for Prediction and Detection of Stress

Mr.Purnendu Shekhar Pandey
BML, Munjal University, Gurgaon
Haryana, India
purnendu.pandey@bml.edu.in

Abstract— **According to a cardiac surgeon, it is difficult to predict age from heart rate as it is nonlinear, but we can use a person's heart beat to predict whether that person is fit, unfit and overtrained or not, provided we have that person's age. Based on heart beat we can predict whether a person is in Stress or not. Stress is one of the main factors that are affecting millions of lives. Thus, it is important to inform the person about his unhealthy life style and even alarm him/her before any acute condition occurs. To detect the stress beforehand we have used heart beat rate as one of the parameters. Internet of Things (IoT) along with Machine Learning (ML) is used to alarm the situation when the person is in real risk. ML is used to predict the condition of the patient and IoT is used to communicate the patience about his/her acute stress condition.**

Keywords—Stress, heart beat rate, IoT, ML

I. Introduction

Remote Stress detector is an IOT device which can detect the stress level of a person using his/her heartbeat reading. When people are stressed or nervous, there is an increase in their heartbeat just like there's a spike in the heartbeat when a person is having a heart attack, scientifically known as myocardial infarction [1][6]. This device locally collects heart beat reading from a person and sends it to a server on Digitalocean. All the computation is done on the server, which then finally predicts whether the person is stressed or not. It generates a chart of the data of individual server. Node MCU is used as the development board and micro-python for programing language [2]. The micro-python code runs on an REPL (Read–Eval–Print Loop) interpreter. It is a high level language shell that takes a single input from the user, executes it and return the results back to the user. Moreover, this work uses a pulse sensor to detect the pulses from which we calculated the heartbeat rate.

The medical world has seen strong correlations between stress and heart disease, cancer and such terminal illnesses. Further stress has been shown to weaken immune systems, as well as drop performance in all metrics of success[3][5]. Stress cannot be quantified and is very difficult to detect.

978-1-5386-3893-4/17/$31.00 ©2017 IEEE

Through mapping of stress and heartbeat, we can identify a lot of things, for example whether the person is nervous or not, whether the person is in apprehension or fear, whether the person is working out, whether the person is over trained, remote monitoring of a patient with heart disease etc[4][7].
Each one of the application targets a wide user base and has significant uses in the medical industry. Moreover, there are hardly few data sets available which gives data about a user resting heart rate and elevated heart rate and the one's which are available contain data of mere 200 individuals which is not enough to implement a machine learning algorithm. Through in this paper, we can very easily gather reliable data of different individuals. Additionally, when we are training, our heartbeat rate should be elevated, especially if a person is doing any cardio-exercise and is dependent upon a person's age. Since exercising has become a part of most people's daily regime, it is crucial to identify whether he/she is exercising correctly or not.

II. Literature Review

Heart rate variability refers to the beat to beat alterations in the heart rate. According to National Institute of Health, infants generally have a heart rate of more than 100 beats per minute which settles down between 60 to 100 beats per minute (resting heart rate) for children 10 years and older, and adults. Heart rate is directly proportional to a person's fitness [8]. A person who is in good shape will have a resting heart rate in the range of 50-60 bpm compared to an average individual whose heart rate might fluctuate between 60 -80 bpm, whereas a well-trained athlete can have an heart rate as low as 40 bpm [9].
The diagnosticity of heart rate is restricted by several factors like environmental stressors and mental and physical workload. The demand control model of job strain tells that jobs in which people have low social support and no control over their job are more stressful i.e. when there is an imbalance between the job demand and the person's ability to meet those demands[10]. If the condition is long lasting, it may lead to cardiovascular diseases. There is a clear increase in a person's heart rate when he is stress or nervous, but the offset differs from individual to individual,
According to a cardiac surgeon, it is difficult to predict age from heart rate as it is nonlinear. But we can use a person's

heart beat to predict whether that person is fit, unfit and overtrained or not, provided we have that person's age. Studies have shown that if athletes have a higher resting heart rate at 7:00 am in the morning for three or more consecutive days then he/she is overtrained. Overtraining can lead to decrease in fitness and strength of an athlete. A person can have a maximum heart rate of (220 - (his/her age)). Using this as a guideline, doctors charge up the defibrillator during cardiac arrest. When a person is working out or in the gym then his heart rate should be in the range of 50% - 70% of (220 - (his/her age)). If his heart rate is less than this then he needs to exercise harder [4][6].

III. Design & Implementation

A. Methodology

The developed prototype detects whether a person is in stress using variability in his/her heart rate. It can also help in detecting pattern of changes in a person's heart rate when he/she is working out at the gym. Each device is individual specific and needs to be calibrated for it to function properly. During calibration the person should be in a relaxed mood and should be resting. This is done to set up a baseline (Fig.1), after calibration, the device uses this baseline (which is different for every individual) to determine whether that person is in stress/nervous, over trained or currently training (Fig.2).

The heartbeat readings are pushed to the server where they are filtered using a user's network id to keep track of readings for a particular individual. They are visually shown using a connected scatter plot (Fig.5).

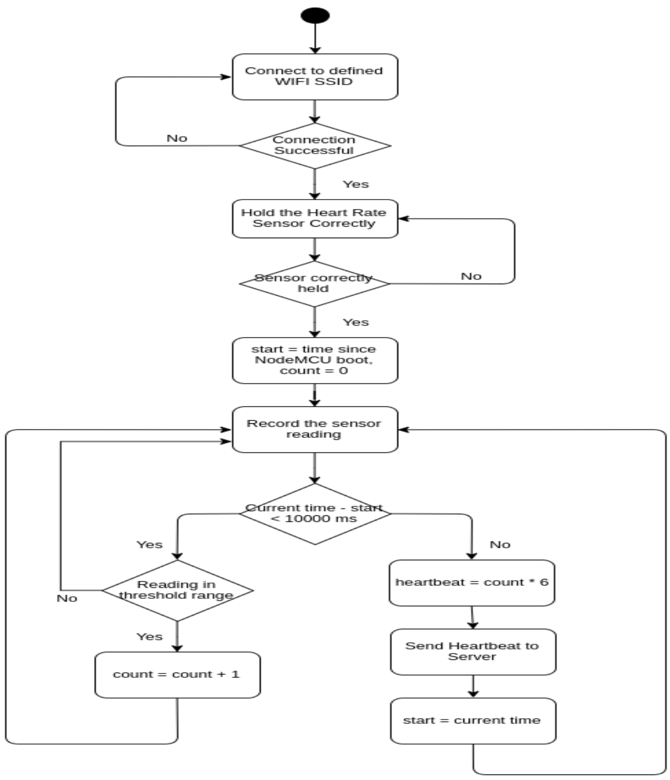

Fig.1. Data Flow Chart

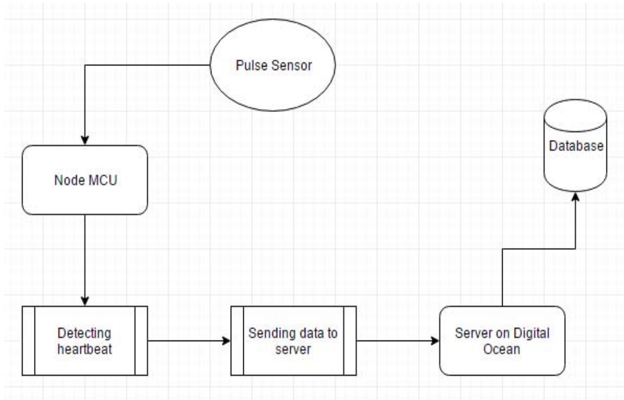

Fig.2. Block Diagram

IV. Hardware and Implementation

A. Components Used

Node MCU

It contains a wifi module ESP8266 along with a 32-bit microprocessor (Tensilica Xtensa LX106) along with 160KB RAM which is much better than 8-bit ATMega328P with 2KB RAM on Arduino UNO (Fig.3). It establishes a connection with Wifi and supports micropython which is a minimal version of python. Python allows flexibility, is quick and easy to do.

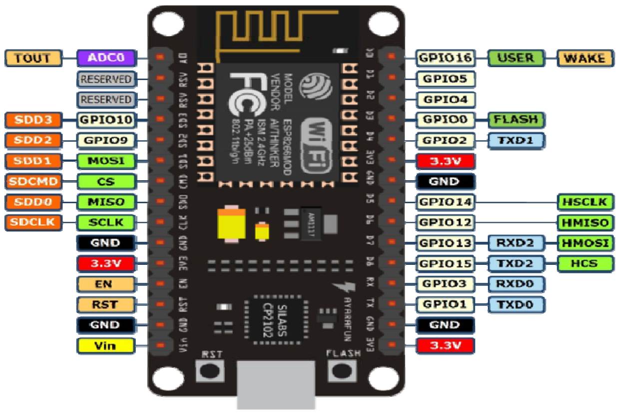

Fig.3.Node MCU

B. Pulse Sensor

It detects the pulse rate of the body which can be counted per unit time to find out the heart rate. This heart rate data can be used in multiple scenarios such as in sports, healthcare and amongst mobile developers etc. It can be included in our daily lives to monitor our anxiety and stress levels which might help in better living. It records a reading by placing the tip of a finger on it or attaching the sensor with a wrist. Pulse sensor has a threshold value ranging from 525 to 610, which needs to be calibrated (Fig.4). To count the heart rate, one needs to count the number of pulses in a minute. This is done by calculating the interbeat interval.

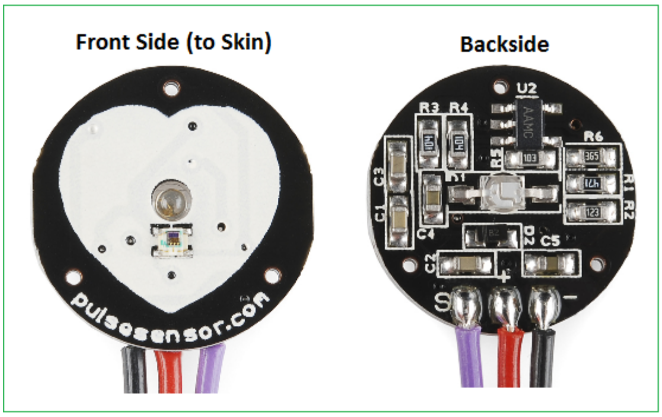

Fig.4.Pulse Sensor

C. Server and Program Flow

The NodeMCU periodically takes a heartbeat reading and sends it, along with a timestamp to our server. The server has been implemented in flask and deployed on a digitalocean droplet running Ubuntu 16.04 LTS. We are using the most basic tier that digitalocean offers, due to the lack of heavy hardware requirements for our paper.

The Server is a Flask overlay on top of a Gunicorn WSGI Server with NginX to handle asynchronous requests and perform load balancing dynamically. The server is built on the Restful API, MVC Controller model - and thus exposes various endpoints to the user, the index endpoint allows users to see the visualizations of the data posted to server, and whether they're stressed or not at the particular instant.

The /<age> endpoint exposes the activity tracker part of our paper - it takes the age provided in the request and determines whether someone is currently exercising enough or not (Fig.6 and Fig.7). The </data/add> endpoint is the location to point all the sensor data being sent via the nodemcu.
Over time, the server assembles a fingerprinted record of the user's heartbeat at different times of the day. We can then perform data analysis on the server - look at various times of the day that the user has an elevated heart rate over the baseline and thus determine his or her stress levels to some degree of accuracy. Further, the website draws visualizations of this data using the google graphs API and a few fractal drawing libraries .

V. Results and Analysis

Stress labels are calculated and median in our data set was stressed and map that to our detector. Further 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 exercise detection part of our paper does work well because that relationship is clearly defined as 50-80% of the max heart rate (220-Age)

A. Snapshots

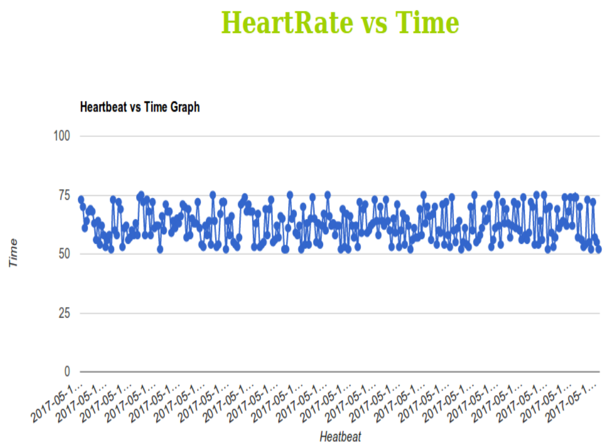

Fig.5. Heart rate vs time graph

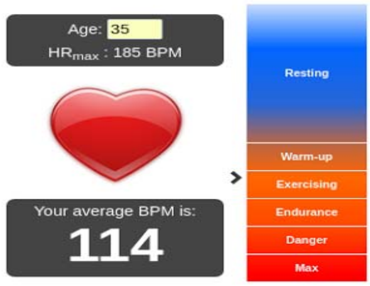

Fig.6. Determining whether the person is exercising

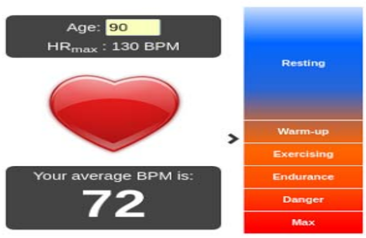

Fig.7. Determining whether the person is at rest or relaxed state

VI. Classification

A. Logistic Regression

Supervised learning comprises in taking in the connection between two datasets: the watched information X and an outer variable y that we are attempting to anticipate, as a rule called "target" or "names". Frequently, y is a 1D cluster of length n_samples.
Every supervised estimator in scikit-learn actualize a fit(X, y) method to fit the model and a predict(X) technique that, given unlabeled perceptions X, gives back the anticipated marks y.
In the event that the expectation assignment is to order the perceptions in an arrangement of limited marks, at the end of

the day to "name" the items watched, the undertaking is said to be a grouping errand. Then again, if the objective is to foresee a ceaseless target variable, it is said to be a regression task.
Logistic regression is the fitting relapse examination to lead when the needy variable is dichotomous (binary). Like all relapse examinations, the calculated relapse is a prescient investigation. Strategic relapse is utilized to depict information and to clarify the connection between one ward double factor and at least one ostensible, ordinal, interim or proportion level autonomous factors.

Procedure :

- Training Size = 318
- Number of Features = 270
- Using One vs. Rest Classifier
- Creating Confusion Matrix
- Calculating Training and Test Accuracy

B. Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a discriminative classifier formally characterized by an isolating hyperplane. At the end of the day, given named preparing information (managed taking in), the calculation yields an ideal hyperplane which classifies new cases.

Procedure :

- Training Size = 318
- Number of Features = 270
- Creating Confusion Matrix
- Calculating Training and Test Accuracy

VII. Comparing Results

In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix)as shown in fig.8 and fig.9 respectively and its test accuracy is shown in fig.10 and fig.11. Each column of the matrix represents the instances in a predicted class while each row represents the instances in an actual class (or vice versa). The name stems from the fact that it makes it easy to see if the system is confusing two classes (i.e. commonly mislabelling one as another).
In the confusion matrices below predicted label is in the y-axis and the true label is the x-axis.

A. Trained Model Confusion Matrix

Confusion Matrix of Logistic Regression

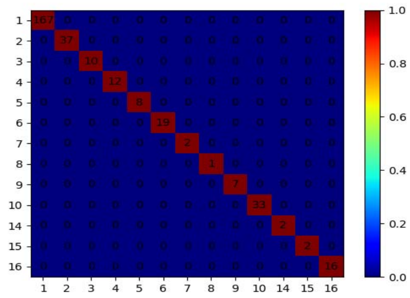

Fig.8. Accuracy = 100 %

SVM Trained Model Confusion Matrix

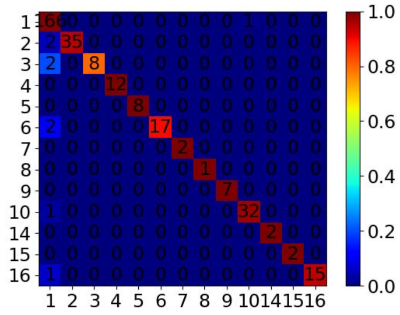

Fig.9. Accuracy = 97 %

B. Test Accuracy

Logistic Regression

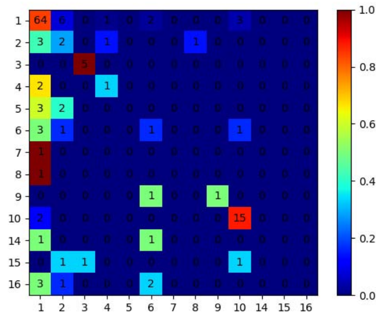

Fig.10. Accuracy = 66 %

SVM

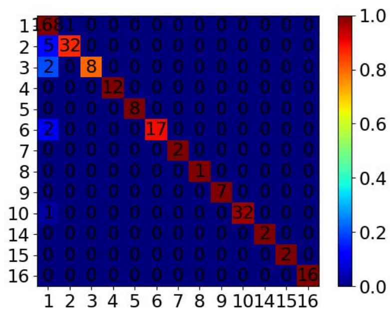

Fig.11. Accuracy = 68 %

VIII. CONCLUSION

For calculating accuracy we find the missed predictions in the confusing matrix which gives us the error rate, subtracting it from 1 gives us the accuracy of the classifier.

Classifier	Train Accuracy	Test Accuracy
Logistic Regression	100 %	66 %
SVM	97 %	68 %

Two algorithms for classification are being used VF - 15 algorithm, which is a feature interval based classifier, which creates classification intervals during training and use it to test the classifier gives an accuracy of 62 % and Naive Bayes approach which is a Bayesian classification algorithm gives 50 % of accuracy while testing.

Classifier	Test Accuracy
VF - 15	62 %
Naive Bayes	50 %
VF - 15 with weights to features	68 %

We applied SVM and Logistic Regression which show considerable improvement over VF - 15 and Naive Bayes without any external weights provided. In the weighted VF - 15 the accuracy was comparable to SVM, but in SVM no external weights were provided as they were in the weighted VF - 15 using genetic algorithms.

Using Logistic Regression and SVM we get an accuracy of 66 % and 68 % respectively, which shows an improvement in accuracy after using SVM.

IX. FUTURE SCOPE

This paper provide an insight into the applications of heart rate monitoring and serve as a stepping stone for any new research work in this field. Moreover, the paper faces the challenge of inadequate data, as any machine learning algorithm can only give correct readings/predictions if it is applied on reliable data. Hence, the next step in this work would be to gather heart rate readings of different individuals.

We can integrate this work with any health monitoring device and safety device. To get better results, we are using his heart rate and also his daily activity pattern to determine at what time of the day he is exercising and what time he is stressed, as any device can give a false alert when the person is not in danger. One of the most promising future aspects of this work can be to use a person's profile and his daily heart rate measurements along with his galvanic skin response to determine the mood of a person. Crucial analysis like Arrhythmia and stress we need high accuracy, for which more data is needed which can be preferably acquired from sources like health bands, smartphones, smart watches etc. Also for more accurate modelling we can use Neural Networks, which work quite well on datasets with high number of features.

REFERENCES

[1] Effects of physical and mental tasks on heart rate variability. http://etd.lsu.edu/docs/available/etd-08222009-204359/unrestricted/thesis.pdf by By Satya A. Paritala

[2] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, and Gerhard Troster, Member, IEEE, and Ulrike Ehlert, "Discriminating Stress From Cognitive Load Using a Wearable EDA Device", IEEE transactions on Information Technology in Biomedicine, Vol. 14, No. 2, March 2010

[3] http://www.sweetwaterhrv.com/documentation/stressandhrv.pdf

[4] Hock Beng Lim, "A Soldier Health Monitoring System for Military Applications," *International Conference on Body Sensor Networks*.

[5] Y. Kawamoto, H. Nishiyama, N. Kato, Y. Shimizu, A. Takahara, and T. Jiang, ``Effectively collecting data for the location-based authentication in Internet of Things," *IEEE Syst. J.*, vol. PP, no. 99, pp. 1_9, Sep. 2015, doi: 10.1109/JSYST.2015.2456878.

[6] P. Gope and T. Hwang, ``BSN-care: A secure IoT-based modern healthcare system using body sensor network," *IEEE Sensors J.*, vol. 16, no. 5, pp. 1368_1376, Mar. 2016.

[7] H. Viswanathan, E. K. Lee, and D. Pompili, ``Mobile grid computing for data- and patient-centric ubiquitous healthcare," in *Proc. 1st IEEE Workshop Enabling Technol. Smartphone Internet Things (ETSIoT)*, Jun. 2012, pp. 36_41

[8] N. Yang, X. Zhao, and H. Zhang, ``A non-contact health monitoring model based on the Internet of Things," in *Proc. 8th Int. Conf. Natural Comput. (ICNC)*, May 2012, pp. 506_510.

[9] G. Yang *et al.*, ``A health-IoT platform based on the integration of intelligent packaging, unobtrusive bio-sensor, and intelligent medicine box," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2180_2191, Nov. 2014.

[10] C. Doukas and I. Maglogiannis, ``Bringing IoT and cloud computing towards pervasive healthcare," in *Proc. Int. Conf. Innov. Mobile Internet Services Ubiquitous Comput. (IMIS)*, Jul. 2012, pp. 922_926.

Author

Mr. Purnendu Shekhar Pandey is currently working as Assistant Professor at BML, Munjal University, Hero Group, Gurgaon, Haryana. PhD from Indian Institute of Information Technology, Allahabad, India. His areas of specialization include Mobile IPv6, Wireless Network, Cellular networks, Machine Learning and IoT. He has authored many research papers in National and International Journals (SCI) and Conferences. He had reviewed various IEEE conferences and SCI Journals. He is a professional member of various Technical bodies such as ACM and IEEE.

E-mail: purnendu.pandey@bml.edu.in