STRESS PREDICTION USING RANDOM FOREST LINEAR REGRESSION

Vaibhav Rajendra Mali
Department of Computer Science and Engineering
Walchand College of Engineering
Sangli, India

Prof. Anil R. Surve
Department of Computer Science and Engineering
Walchand College of Engineering
Sangli, India

Abstract—**In today's world stress has become a more familiar word because of its disastrous impact on the huge number of people worldwide. It is very important to keep stress under control every time, as it is the primitive reason for much major health issues. Some people meditate to get rid of it and others choose to use medicines to control their stress levels. Students also found with very much stressed out because of academics, projects, exams, and whatnot. There are many ways through which one can check whether you have stress or not. According to this situation, the medical diagnosis system based on human physiology becomes more requisite as compared to others. Human physiology-based study plays a important character in the detection of mental stress in persons. There have also been eventual researches which are done on the detection of stress based on facial emotions. To find out whether stressed or not we need to see a doctor and get checked, but it seems to be not practical at all times to do so. In fact, in the era of digitalism, where everyone has a smartphone there is a dearth of finding novel ways through which we can make use of technology to detect your stress levels automatically. There are wearable devices that detect stress levels based on your body activity. Many approaches aim for the detection of stress through the use of wearable devices. The approach that we are presenting in this project is predicting stress through medical data of the patients using random forest regression. Additionally, an examination between oneself fabricated convolution neural model and a portion of the pre-trained models has been finished. This is another methodology and we are getting very promising precision by utilizing sufficient research experiments on 2000 irregular trees in the model. The results achieved are the outcomes of effectively anticipated with the accuracy utilizing the model. The outcomes of this research can be useful in directing the future which explores automated stress prediction-based systems.**

Index Terms—**stress prediction, machine learning, medical data, random forest, HRV.**

I. Introduction

Understanding Stress

Stress is a commonly used word, but the meaning is difficult to settle on because it is complex and difficult to demonstrate and quantify [8]. However, since stress can not be established by scientists, how can they measure it? The Merriam-Webster dictionary describes stress as a "social, chemical, or emotional force that induces physical or mental stress and may be a factor in the causation of disease." Stress may be defined non formally as the way the body responds to any demanding or unsafe condition [9]. The brain initiates a stress reaction as feedback to sensory inputs from the ear, eye, and nose. At time of the body senses a risk or panic situation, it could be real or the virtual, mechanism for resisting such hazardous situations started by the human body. The hippocampus rapidly sends a signal of pain towards the hypothalamus. The hypothalamus is closer to the brain's command center [10].

Stress can be categorized mainly into two types such as acute stress and chronic stress. The acute is somewhat beneficiary stress because a limitation of this type of stress becomes beneficial for our daily work. For example, if we have presented work in an office, feeling a little stress makes a person busy with his work which leads to success in his work. In this way, acute stress is becoming a motivating factor for the person. Again when such a stress level increases out of limit it becomes hazardous for an individual this type of stress is considered severe type stress. Chronic stress is stress that builds up over a prolonged period when you are subjected to a hectic environment. When you enter

into this chronic mentality of stress, it will result in continuous feelings of fear, depression, or other stress signs that are very dangerous for such an individual's health. Again the question arises that how to identify that someone is under severe stress? there are some ways to identify that individual is under chronic stress such as Psychological symptoms include difficulties focusing, worrying, exhaustion, and difficulty recalling. Emotional signs such as temperate, angry, moody, or annoyed rise. Physical symptoms, such as high blood pressure, weight gain or lack, loss of immunity, and changes in women's menstrual cycle. It also leads to heart diseases like heart problems, hypertension, cancer, etc.

A. How Random Forest Regression works

Random Forest is one of the most effective machines learning models for predictive statistics, making it an industrial machine learning task. The random forest model is a type of add-on model that creates predictions by combining decisions from a sequence of basic models. where the ultimate model is the total of the basic simple models. Here, the division of each foundation is a simple decision tree. This deep method of using multiple models to obtain better guessing performance is called model integration. In random forests, all base models are built separately using a separate sample of data.

B. Motivation

Almost everyone has stress nowadays. When stress becomes chronic, it poses a high health risk and that's why it is a vital issue to be considered. Stress is caused by a lot of factors including deadlines, emotions, work stress, etc. It can easily be predicted by using medical data of patients using random forest regression. The Random Forest is volatile, it is convenient to use a machine learning algorithm that is built up, in the absence of hyper-parameter tuning, we get good results often. And it's one of the major widely used algorithms because of its clarity and that it can be used for regression and classification work. One of the biggest drawbacks with over-the-counter machine learning, but generally this won't happen so easily in a random forest algorithm. Hence if there are enough trees in the forest, the divider will not exceed the model.

II. LITERATURE SURVEY

As stress becomes a worldwide problem, there are so many researches are going on stress. Mr.Purnendu Shekhar Pandey in his research uses the relation between heart rate and mental stress. They use machine learning techniques like linear regression and SVM(Scaler Vector Machine) for the classification work which is useful in the prediction of stress. IoT is used here to alarm the person when they are in actual risk condition. they classify the stress level in 6 classes as Resting, warm-up, Exercising, Endurance, dangerous, and max [1]. They got an accuracy of 68% which can be improved further. Only heart rate is can't be sufficient to detect the stress of the person. We have to consider the facial expressions as well to detect the level of stress of that person. Sriramprakash.S, Prasanna Vadana. D, O. V. Ramana Murthy proposed a technique that was applied on SWELL- KW dataset on Kaggle. They Considered the physiological signals obtained from ECG (Electrocardiogram), galvanic skin response sensors, etc. Again they use the SVM and KNN (K-Nearest Neighbour) Machine Learning algorithms to classify the data. they proposed the technique for the selection of dominant features and use this with overlapping techniques to get better results. They conclude that HeartRate (HR), HeartRate Variability (HRV), Galvanic Skin Response (GSR) which are the time and frequency domain features are sufficient for the prediction of stress [2]. There proposed system gives classification accuracy by using KNN is 66.52% and by using SVM is 72.82%.

Martin Gjoreski, Hristijan Gjoreski, Mitja Lustrek, Matjaz Gams proposed a system that can be used either in the laboratory or in real life for continuous stress detection purposes. The web application was developed for collecting the data in the laboratory. In this application, they use certain equations to classify the stress levels of a person. Again in real life for the collection of data they used a combination of Ecological Momentary Assessment (EMA) as well as stress log, which was implemented on a smartphone [3]. In actual method they use Base stress-detection classifier, Activity-recognition classifier, Context-based stress detector for detection of stress. they got 92% accuracy in this work but there is the actual the challenge of collecting results from the laboratory with measuring the cortisol level in the body for a longer period.

In this era of technology and innovation where computers are used to monitor the emotions of people's, however, The problem is that there is so much technology accessible for emotional detection such as image processing but it has some demerits because it is tough to see the real feeling from artificial emotions. Because of the growth of the technological world, this field needs more awareness and research contribution. Mr. Anil R Surve and Mr. Manish N. Patil have proposed a system to make existing sensory processing extremely effective. Bio (non-invasive) sensors are used to collect sensory-related data in the human body. Using Electrodermal Activity (EDA) they measure the skin's movement of the human body to detect various processes. After the diagnosis is made using behavioral analysis like emotional patterns they can predict future illnesses and other disorders [13]. The proposed researches are done with the long-term use of wearable devices for detection and prediction of stress simply means they use the Supervised learning methods and operated on large labeled datasets. Elena Vildjiounaite, Johanna Kallio, Vesa Kyllönen, Mikko

Nieminen, Jani Mäntyjärvi, and Georgy Gimel'farb proposed a randomly controlled pressure control device based on a smartphone as a single device. They use Markov Hidden Discrete models (HMM) with Maximum posterior marginal (MPM) resolutions [4]. As they don't use any other hardware requirements they can be suitable for lifelong use. They suggested a full-day normal behavior model and distinguished it into two pieces of difficult and normal days. The main limitation for this work which uses an unsupervised technique is the compulsion to belonging a smartphone with the person all time. The various factors such as HR, HRV, ECG, EEG can be used to detect and predict stress but by using facial emotions as an additional factor we can improve the accuracy for detection and prediction of stress work to great extent.G. Giannakakis, M. Pediaditis, D. Manousos, E. Kazantzakis, F. Chiarugi, P.G. Simos, K. Marias, M. Tsiknakis proposed the study which is characterized by a paradigm for the identification and interpretation process of stress-related emotional states with the aid of video captured facial signals. The in-depth research is performed on different internal or external stressors impacting established body states such as normal, calm, and stressed. The various features like heart rate estimation through Photoplethysmography, events related to the eye, various mouth activities, and motions of the head are considered. The comparative study of relaxed state with stress or anxiety state is done [5].In classification work, they use different classifications tools like shape templates, Eigen vectors, Hidden Markov Model (HMM), Support Vector Machine(SVM), Artificial Neural Network (ANN). As a result, they validate that the eye blinking rate is increased during stress or anxiety condition. In addition to that head, movement velocity is also affected. Considering mouth activity, while watching videos that are stressful the median and temporal variability (VTI) is increased. they got an average of 86.23 % accuracy for the proposed work. Anil R Surve, Vijay R. Ghorpade, and Anil S. Patthe these authors proposed a system for continuous recognition of facial expressions. Considering the facial expressions, we can detect the emotions of that person. It is easy to detect the emotions of a person for a human being but for the computer, it is difficult to detect the emotions of a person. This proposed work gives a solution to detect one's emotion by considering facial expressions. The technique used here is Emergence-based Eigen face technique. For extracting information from image frames and to reduce the dimensionality of computations Principal Component Analysis(PCA) technique is used [6]. In classification work, they use Euclidian distance metric as well as temporal dynamics for the redundancy removing process. They present the results into four categories of emotions such as neutral, Happy, Disgust, Anger, Sad. They got an accuracy of 91.26% for the identification of emotions of the person using this technique.

III. PROPOSED SCHEME

The main purpose of this work is to develop a system that can predict stress. In this case, a machine-based approach was used. A random forest regression-based model was developed that assumes the inclusion of medical information for patients. When training is applied to the model on this data, the model predicts the level of stress of the patient whose medical information is included. Random forest Regression has been used in various fields for e.g. in banking, medical field, share market, and electronic commerce. In this field, it is also helpful for detecting fraud customers which may be responsible for scams in the bank. It is used to determine future stock performance in the finance department. In the healthcare area, it helps identify the perfect composition of components in drugs and to examine a medical history of patient which will help to detect diseases belonging to this patient. Also, efficiency metrics will apply between this work and the existing works to analyze if the proposed model based on random forest regression is efficient and feasible.

IV. METHODOLOGY

The objectives of our work were: To find how efficient and accurate the Random Forest-based models using patient's data can be in stress prediction. To carry out this work, we chose the alternative of using random forest because It can be easily predicted by using medical data of patients using random forest regression. The Random Forest is volatile, it is convenient to use a machine learning algorithm that is built up, in the absence of hyper-parameter tuning, we get good results often. Also, this technology has not been yet applied for solving stress prediction problems using patient's medical data. Thus, we planned to find a database that had medical data which plays a great role in stress. After data acquisition, the dataset was pre-processed to recover the missing values by using the mean method. After this, feature scaling was applied to the dataset so that no attribute intimidates other attributes. Finally, regression using Random Forest models was done and the results were compared.

A. Data pre-processing

The training details come from a project done at MIT by Healey as part of his Ph.D. concept, which consists of body measurements performed on young people driving in stressful environments and leisure time to create a trouble-free basic reading [12].

B. Different attributes used in Dataset:

Galvanic Skin Response (GSR) sensors are often used to measure stress by measuring skin resistance, and when a person is more stressed, sweating profusely and resisting weight loss. Heart rate variability (HRV) is a superset for the various factors found in the interval between heartbeats called RR intervals. HRV on average is measured by a clinical level electrocardiogram (ECG) [13]. The NN

interval is a different way of saying RR interval, or interval between R peaks. The letter N is used to distinguish whether these figures are intended to be found in the "normal" RR spaces, or RR intervals that represent normal heart time and are not an artifact.

Time Domain Measures are as follows:
SDNN: Standard deviation of all NN intervals.
pNN50: Percentage of differences between adjacent NN intervals that are greater than 50 ms.
AVNN: Average of all NN intervals.
RMSSD: Square root of the mean of the squares of differences between adjacent NN intervals.

Frequency Domain Measures are as follows:
LF Low Frequency: Total spectral power of all NN intervals between 0.04 and 0.15 Hz
HF(High Frequency): Total spectral power of all NN intervals between 0.15 and 0.4 Hz
LF/HF: Ratio of low to high-frequency power.
ULF(Ultra Low Frequency): Total spectral power of all NN intervals up to 0.003 Hz
VLF(Very Low Frequency) : Total spectral power of all NN intervals between 0.003 and 0.04 Hz

V. MISSING VALUES

Many machine learning algorithms needed numerical input numbers, and the number to be present in each row and column in the dataset. Thus, missing values can create many problems in machine learning algorithms. Hence, it is necessary to detect missing values in a dataset and change them with numerical values. There were a lot of missing values to be taken care of in the data for which techniques to replace the missing cells with mean/median were used. We have two methods for replacing missing values in the dataset such as replacement by mean and replacement by a median. Here we have used the replacement by a mean method.

VI. APPLYING RANDOM FOREST REGRESSION

We have divided the dataset into train and validation set in the ratio 90:10. The details of the parameters of the experiment are given below:

Train-validation ratio	90:10
Total categories	23
Total rows in the dataset	4129
Training rows	3716
Test rows	413

Fig. 1. Splitting ratio

VII. DESCRIPTION OF ATTRIBUTES IN TABULAR FORM

The above tables provide us with different measures (like count, mean, std, min, and max values) about the attributes, ECG, EMG, HR, and RESP used to train our machine learning model.

Measures	ECG	EMG	HR	REST
Count	4129	4023	4129	4129
Mean	0.17092	0.604475	83.136251	29.846
Std	0.137222	0.807892	17.633319	14.0245
Min	0.084500	0.108651	72.861111	28.5650
Max	0.681714	9.491700	372.00	52.089

Measures	Seconds	FootGSR	HandGSR	Interval in seconds
Count	4129	4129	4056.00	4129.000000
Mean	2278.78	6.487689	11.523244	0.788440
Std	1313.49	4.307487	6.739658	0.109768
Min	12.529	0.971111	-28.38	0.527944
Max	5005.37	22.582463	31.219800	1.040088

Measures	Marker	New time	SDNN	RMSSD
Count	3722.0	4129.0	4.007000e+03	4129.00
Mean	12.057030	35171.533570	5.781201e-02	0.027223
Std	2.612543	19946.879538	3.401004e-02	0.016964
Min	0.000000	12.529684	7.840000e-09	0.000000
Max	35.624000	70173.851200	1.645780e-01	0.080981

Measures	LF	HF	$LF_H F$	STRESS
Count	4002.00	4002.0	58.000000	4129
Mean	0.000003	0.000001	3.555695	0.516078
Std	0.000031	0.000010	2.080121	0.431116
Min	0.000000	0.000000	0.418526	0.00
Max	0.000616	0.000261	9.117240	1.000

VIII. GRAPHS REPRESENTING THE RELATIONSHIP BETWEEN STRESS AND OTHER ATTRIBUTES
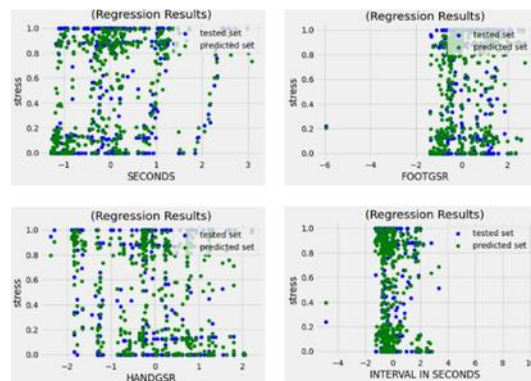
Fig. 1. Stress vs Attributes
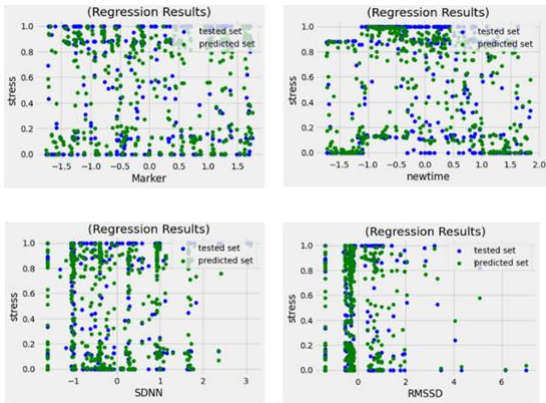
Fig. 2. Stress vs Attributes
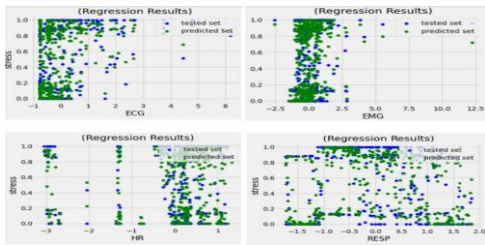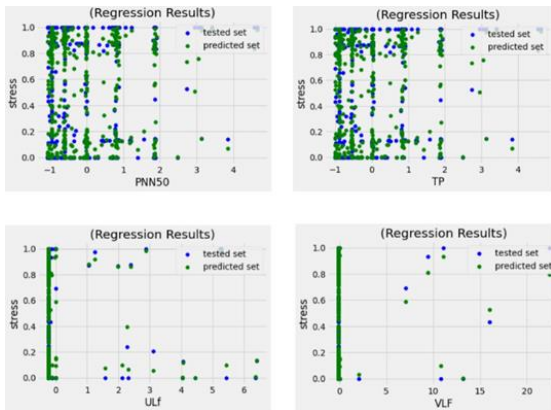

Fig. 3. Stress vs Attributes

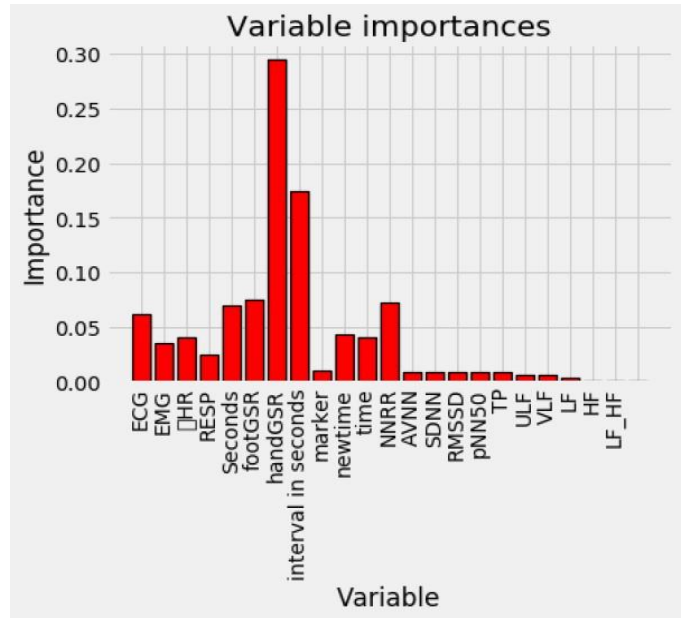Fig. 4. Stress vs Attributes

A. *Variable Importances*

Fig 5. Variable Importances

According to the above graph, handGSR is an attribute on which prediction of stress highly depends. After that interval in seconds is the second most important attribute in stress prediction. While footGSR and NNRR have the same importance in stress prediction. The graph also represents that stress is also related to ECG, EMG, and HR. RESP is a factor next to them in stress prediction. Which seems complete we were there as a factor that can be used to determine the stress or are relatable to stress. This graph also shows that the remaining attributes have very little relation with stress.

IX.   NUMERICAL MEASURES FOR EVALUATING THE MODEL

To find how reliable our model is we have calculated several measures. One in them is explained the difference in values. This value leads to individual estimates points for corresponding diversity Target variation. This setting determines the value of the whole world that has captured the unbroken diversity. Multi-output value = 'variance weighted' is the default value of the r2 score for backward compatibility. Thus our explained variance score is 0.921048 and the r2 score is 0.920961 which shows our model is very much reliable as the higher the value of these variables closer to the one the higher will be the reliability which shows our model is very good in its prediction.

X. CONCLUSIONS

The Random forest algorithm is one of the proven algorithms which were used in various researches. It is the most promising option as compared to other different algorithms, because of its compatibility with both classification and regression. The research attempt is configured in such a way that it predicts the stress level of patients whose data was made available. The Random forest regression-based model takes the medical data of patients as input and it gives results into the stress level of that patient as output.

The proposed stress-predictive system using randomized learning equipment based on a random forest rescue model has achieved 95.6 percent accuracy which is the optimum finding of the research.

XI. REFERENCES

[1]P. S. Pandey, "Machine Learning and IoT for prediction and detection of stress," Proc. 2017 17th Int. Conf. Comput. Sci. Its Appl. ICCSA 2017, 2017, doi: 10.1109/ICCSA.2017.8000018.

[2]S. Sriramprakash, V. D. Prasanna, and O. V. R. Murthy, "Stress Detection in Working People," Procedia Comput. Sci., vol. 115, pp. 359–366, 2017, doi: 10.1016/j.procs.2017.09.090.

[3]M. Gjoreski, H. Gjoreski, M. Lu s̆trek, and M. Gams, "Continuous stress detection using a wrist device - in laboratory and real life," UbiComp 2016 Adjun. - Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput., pp. 1185–1193, 2016, doi: 10.1145/2968219.2968306.
.

[4]E. Vildjiounaite et al., "Unobtrusive stress detection on the basis of smartphone usage data," Pers. Ubiquitous Comput., vol. 22, no. 4, pp. 671–688, 2018, doi: 10.1007/s00779-017-1108-z.

[5]G. Giannakakis et al., "Stress and anxiety detection using facial cues from videos," Biomed. Signal Process. Control, vol. 31, pp. 89–101, 2017, doi: 10.1016/j.bspc.2016.06.020.

[6]A. R. Surve, V. R. Ghorpade, and A. S. Patthe, Continuous Facial Emotion Recognition System Using PCA for Ambient Living, vol. 839. Springer Singapore, 2019.

[7]https://www.stress.org/what-is-stress/ .

[8]https://www.helpguide.org/articles/stress/stress-symptoms-causes-and-effects.htm ¿ .

[9]http://www.health.harvard.edu/stayinghealthy/understanding-the-stress-response .

[10]K. Plarre, A. Raij, S.M. Hossain, A.A. Ali, M. Nakajima, M. Al'absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, L.E. Wittmers, Continuous inference of psychological stress from sensory measurements collected in the natural environment, Proceedings of the 10th ACM/IEEE International Conference on Information Processing in Sensor Networks, 2011, pp. 97–108

[11]Moors A. Theories of emotion causation: a review. Cognit Emot 2009;23(4):625–62.

[12]https://dataespresso.com/en/2019/01/30/stress-detection-with-wearable- devices-and-machine-learning/ .

[13]M. N. Patil and A. R. Surve, "IOT in HealthCare: Smart Emotion Detector Utilizing Wearable Bio Sensors," 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information and Communi- cation Technology (RTEICT), 2018, pp. 1755-1760, doi: 10.1109/RTE- ICT42901.2018.9012232.

[14]S. M. Rajbhoj and S. F. Lakudzode, "Human stress monitoring system using electrocardiogram," 2016 International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1-5, doi: 10.1109/IN-VENTIVE.2016.7823216.

[15]S. Gedam and S. Paul, "A Review on Mental Stress Detection Using Wearable Sensors and Machine Learning Techniques," in IEEE Access, vol. 9, pp. 84045-84066, 2021, doi: 10.1109/ACCESS.2021.3085502.

[16]Stanley Leonard Tito, A Simplified Outpatient Health Monitoring Sys- tem in Resource Constrained IoT Infrastructure , International Journal of Engineering Research and Advanced Technology (ijerat): Vol. 5 No. 12: December-2019

[17]J. Wijsman, B. Grundlehner, H. Liu, H. Hermens and J. Penders, "Towards mental stress detection using wearable physiological sen- sors," 2011 Annual International Conference of the IEEE Engineer- ing in Medicine and Biology Society, 2011, pp. 1798-1801, doi: 10.1109/IEMBS.2011.60905