Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

MENTAL STRESS CLASSIFICATION USING IMPORVED RANDOM FOREST WITH THRESHOLD FEATURE EXTRACTION ON REAL TIME DATASET

Vanisha P.Vaidya
Department of Computer Science and Engineering, Dr. A. P. J. Abdul Kalam University, Indore, India, vanisha.vaidya@gmail.com

Dr Suresh.S.Asole
Department of Computer Science and Engineering, Babasaheb Naik College of Engineering, Pusad, India, suresh_asole@yahoo.com

ABSTRACT: *Mental stress is big issue today's society and everyone right from kids to adults are facing this problemthat causes a variety of chronic illnesses such as depression, malignancy, and cardiovascular disease(CVD). As a result, it is critical to handle and track a person's stress on a regular basis.We present a Random Forest (RF) method that can reliably predict mental stress levels by applying an enhanced band pass filtration technique to remove noise and exact features of R-Speak from Electrocardiogram (ECG) data in this work.When a person is stressed, a shift takes place in the ECG, which is an electrocardiogram (ECG) signal. Stress signals may be classified by analyzing ECG data and extracting particular characteristics.The suggested model obtained 98.0% accuracy based on the performance assessment criteria of the stress categorization model, confusion matrices, and precision-recall, which is a 14.7% improvement over prior research results. The suggested model was additionally validated using real-time datasets, and the model's accuracy was 98.0%.We categorised the model's output into three categories: normal, stress level 1, and stress level 2. As a result, our methodology can assist persons with stress manage their mental health.*
Keywords: Mental Stress, ECG Signal, Machine Learning, Random Forest

1. INTRODUCTION:

Stress is an emotional as well as physical response that a person may have when presented with a difficult or unexpected situation or event.Excessive stress can lead to chronic ailments like heart disease, high blood pressure, and cancer, as well as death in severe situations[1, 2]. As a result, stress monitoring is becoming increasingly vital in modern culture.

Several studies have been published that measure stress through the use of biological parameters like electromyogram (EMG), electroencephalography (EEG), oxygen saturation, & pulse waves[3–5]. However, these measuring approaches necessitate the employment of expensive and large data acquisition devices, are difficult and costly to use, and necessitate signal processing by professionals.

Stress identification and detection can be done by variety of methods, but electrocardiogram (ECG) have proved to be the most commonmethod as the collection of data of human's cardio signal is the easiest way to generate a clear waveform.Using Support Vector Machine Methodologies (SVM), 2 research obtained accuracy of 89.21% and 84.4%, respectively. [10, 11], However, obtaining right types of features will be difficult as it contains noise which may impact on signal acquisition,further leading to improver diagnosis of CVD. Two separate investigations attained 75% and 89% accuracy when taking the mean and standard deviation of the R-R period of the Heart Rhythm Variability (HRV) indication into consideration[12, 13]. Time is an important factor for prediction of CVD as man and standard deviation of R-R intervals takes a considerable amount of time. The variance in the values of the variables is minimal, accurate stress classification is difficult.

Underfitting, computing the normal deviation for the R-R period, which pertains to the duration of a long-term HRV signal, and identifying an improper R peak value are all disadvantages of the previously described stress signal classification approach based on the ECG signal. To address these issues, we offer a random forest model with pre-processed features that accurately detects mental stress. **The proposed model selects the P-Q-R-S peak points using the cut-off values, & classifies the stress-related signal using ECG analysis**.

Noise filtration with a Bandpass Filter and proper analysis of features such HRV, P, Q, R, and S Waveforms are provided as data to our random forest model to increase stress classification accuracy. We also used the clustering process to convert the unsupervised feature data into supervised machine learning, where we may apply machine learning. Confusion matrices, ROC curves, and also precision-recall (PR) charts were utilized to evaluate the performance of the stress classification model.

In this paper, we introduced an integrated Random Forest approach for classifying human mental stress using ECG data. The MIT-BIH Database data were trained, and a classification performance of over 98% was attained.The results were divided into three stress levels: normal, stress level 1, and stress level 2.The most essential part of the research endeavor is as follows:

1. Noise was filtered using a bandpass filter to provide optimum characteristics.
2. R-Speak was extracted using threshold values for identifying the appropriate signals for ECG Stress Level Classification.
3. With maximum accuracy, the newly developed Random Forest model was verified with real-time dataset.

The rest of the article is structured as follows. Section 2 offers an outline of the review of literature. Section 3 discussed methodologies such as dataset definition, preprocessing, and feature extraction, as well as numerous machine learning models.The experimental results were given in Section 4. Discussion was done in Section 5 and, Section 6 presents the conclusion.

2. RELATED WORK:

Monika Chauhan et al. argued for the use of many pre-preparation techniques for diagnosing stress using ECG Signals by utilizing the data of patients[1]. SVM [2]algorithm was used to categorize the stress into normal and stress state through experimental analysis.Linear discriminant analysis (LDA) and decision trees [3] models are used for analysis. Experimental analysis predicts the output with 90% accuracy.

A. Alberdi et al. developed the most exact stress detection systems, demonstrating that stress identification using electrical signals from the body is considerably more accurate than using other modalities.[4]. This is not to suggest that behavioural and context-specific data cannot be utilized to effectively detect stress; research shows that they can, but more study is needed in this area.When data from a number of different sources, such as behavioural reactions, is included, the same or greater accuracy is necessary to design a substantially less intrusive and pervasive surveillance system which is far more practicable in real life.

The rate of variation across all pulses in time is referred to as HRV. It's used to study the autonomic nervous system (ANS), it controls the body's unconscious activities including heartbeat, breathing, movements of the intestines, heart rate, blood pressure, urine, & pupil dilation/constriction. S. Ishaque et al. use signal processing and machine learning to summarize and analyze HRV research on morbidity, pain, tiredness, stress, and exercise. The significance of HRV research, as well as the limits of methodologies that might be addressed to improve study quality, have been widely investigated. Limiting the signals that were studied to ECG, EDA, PPG, and RESP yielded 25 studies studying the causes and implications of increased/decreased HRV.Increasing morbidity and stress have been linked to HRV reduction. High HRV was often associated with excellent health, notwithstanding the fact there was evidence that it might also be associated with life-threatening conditions such as weariness[5].

Luz Santamria-Granados and colleagues [6]To identify mental stress, researchers employed a Deep Neural Network (DNN) to the detection of emotion on a AMIGOS dataset.Emotion identification is accomplished by matching these behavioral indications with this dataset's intensity and valence data to determine if someone is in an effective condition.Furthermore, a suggestion for emotion analysis utilizing classic machine learning approaches to detect the properties of physiological data in the temporal, rate, and non-linear domains is proposed.CNN is used in this program to automatically extract physiological signal functions, whereas FCN predict emotion.The investigational findings of the AMIGOS dataset reveal that, when compared to the preliminary results obtained by the dataset's developers, the technique described in this research provides more precision in the categorization of emotional states.

Anusha A.S. et al., Identification of job tension based on physiological stimuli using unobtrusive sensors.[7], To determine sympathetic tension activation, EDA, ECG, and skin temperature (ST) were measured during the method. The etiology of the stress reaction was investigated using salivary cortisol levels. A fusion design was used to create classifier ensemble for multisensory stress detection techniques, combining some of the advantages of feature fusion with decision fusion. Class unbalancing techniques were utilized to settle the data unbalancing which includes under-sampling, oversampling, and SMOTE, in terms of their ability to give the best classification results. One important limitation of their research is the that dataset used was of small size.

Furthermore, the stress-testing procedures were carried out in a controlled environment with characteristics such as room temperature & activity restrictions.

Minija Mi and co. [8]suggested a system for actual time monitoring of stress on the basis of peripheral physiological parameters, with the purpose of reducing errors due to human variability and enhancing regressive stress efficiency evaluation.The proposed paradigm has been described as a transductive paradigm focusing on transductive learning, with local learning viewed as a benefit of recognizing the importance of excellent education.

To increase detection efficiency, Victoriano Montesinos et al. proposed the use of a multimodal artificial intelligence stress detection technique that incorporates numerous physiological inputs from two independent wearable sensors.The effect varies based on the subjectivity of the individual as well as the intensity and variety of stimuli[9].

In this study, Mahesh Bhargavi et al. describe the requirements that a dataset that serves as a reference for multimodal human stress identification should meet.Medical and peer-reviewed reviews extract criteria that are centered on current processes and objective facts.It was discovered that none of these compilations of data met all of the criteria for registration as a data set of comparison. Future attempts might potentially try to generate such a comparative dataset while addressing the present criteria[10].

Jie Zhang et al. identified the amount of stress using the metric Heart Rate. [8]. An ECG signal is used to compute the RR interval values.The true +Ve & -Ve rates were used to determine the RR values.The SVM classifier is used for all of the positive and negative judgments. Knowing the difference between the two sets of principles affects a person's stress level. More repetitions have been performed using the Sequential Backward Selection (SBS) method[11].

Each of the created systems gathered attributes in a unique way before building classification models with machine learning methods.Variables like heart rate, variability in blood pressure, and skin conductance have been found to be more effective in predicting a person's stress level, whereas the most powerful methods for classifying data include a SVM, a stochastic forest, and K-NN.This highlights how sensor data may be utilized to efficiently and cheaply identify levels of stress in a person by combining wearable sensors and then machine learning algorithms[12].

Lucie Maránová et al. reported the outcomes of a multifaceted inquiry into automatic heartbeat classification in their study.For the first time, non-ischemic, ischemia (two grades), in and subsequent atrial premature beats were categorized together. ECGs from rabbit isolated heart in non-ischemia and ischemic conditions were used in the investigation. This data set was utilized to assess several morphology and spectral properties, as well as categorization approaches. It has been verified that: a) morphological traits are far more useful than spectral ones; b) precision (up to 98.3% for shape and also 96.25% for spectral features) was obtained by the separation of QRS-T segment and also c) using (3 to 14 features) for model training allows achieving a high degree of accuracy (up to 98.3)[13].

A series of studies and waveforms analyses of ECG waveforms resulted in the Deep ECGNet.The author suggested the best combination of recurrent and CNN construction, based on optimization experiments and the analysis of the waveform features of ECG signals, as well as the ideal the convolutions filter height (related to what is known as the P, Q, R, S, and T wave intervals in the ECG) and combining period (related to the pulse period).Traditional techniques were also used in the trials, with HRV measures and frequency characteristics acting as a benchmark test. Deep ECGNet, a proposed model for identifying stress-related disorders, surpassed conventional approaches with a best-of-breed rate of 87.39% for Case 1 & 73.96% for Case 2, demonstrating 16.22% & 0.98% improvement over the standard HRV technique, respectively[14].

M.Vikas & Sandhu M. did the research [15]seeks to use an optimization strategy and machine learning to construct a machine learning model for classification for large-scale ECG data.The current method employs GA, ABC, & PSO to optimize three characteristics in three stages.The fitness functions of each stage were utilized to calculate the best attributes of the subsequent step.It has improved the ECG signal's characteristics. VMD aids in the reduction of ECG signal complexity in order to reduce total calculation time for improved SVM-based categorisation. Overall, it improved the suggested ECG signal prediction approach's performance above previous research. When a biggerdataset of size of 1000 ECG signals were used, the study got 98.93% accuracy, 96.83% precision, 96.83 recall, & 96.72 specificities. They also showed that the new ECG classification method beat the previous VMD and SVM-based ECG classification methods.

Utilizing a greater number of participants and a total of 1000 ECG signals, the study got 98.93% accuracy, 96.83% precision, 96.83 recall, & 96.72 specificities[16]. Nine volunteers provided ECG data while riding a 360-degree virtual roller coaster. Three assessors then manually classified the VR game in 10-second increments based on three stress levels. Then, based on spectroscopy and 1D ECG data, we offer a novel multimodal deep fusion model that can predict tension from a single 1-second frame.In trials, the suggested solution beats both typical HRV-based artificial intelligence (AI) methods (9% loss in accuracy) & basic DNN methods (2.5% gain in accuracy).The author also gives findings gleaned from the previous-generation WESAD dataset to demonstrate the model's superiority.

P.Karthikeyan et al. suggested a strategy for identifying human tension using immediate (i.e., less than 5 min) ECG & HRV data[17]. Using many pieces of data from a single sensor enhances the accuracy of detection and reliability of stress.This study employed a stress-inducing technique, data collection, preliminary processing, feature extraction, and categorization in order to identify stress.Each subject was exposed to a Stroop colour word stress-inducing task as an ECG signal was recorded simultaneously. To remove high-frequency, baseline wandering, a wavelet de-noising approach was applied. Utilizing a discrete wavelet transform (DWT), the HRV signal was obtained from ECG data after pre-processing. The ectopic spike reduction technique was used to minimize noise spikes in HRV data. The Lomb-Scargle periodogram (LSP) has been successfully utilized to detect stress caused by uneven sampling of the HRV signal.The incorporation of LSP in instantaneous HRV signals (32 s) has alleviated worries about uneven sample & power spectra data, and theory and experiment have shown the validity of such a short-term HRV signal.Theoretical study indicates that analyzing ANS activity in relation to stress necessitates at least 25 seconds of either on- or off-line ECG data.In the subject-independent mode, the overall average detection rate for immediate ECG and HRV signals is 91.66 percent, while in the subject-dependent mode, it is 94.66 percent.

P.P. Singh et al. presented a strategy for assessing a person's level of anxiety in this study[18]. Continuous electrocardiograms were recorded while automobiles drove on open roads in the greater Boston region following a predefined itinerary. Driving in new settings caused a rise in stress levels.By driving in cities, highways, and rest stops.The author used the physio net dataset 'Stress Recognition in Vehicle Drivers' to generate the ECG signal. The amount of stress was determined using the ECG signal.The variables were analyzed using the discrete wavelet transform with the mother wave function 'db6' (RMS, Max, Std, the trend, min, mean, variance). Entropy connection, tilt, median, and crest factor were excluded since they do not effectively describe the transition from normal to stressed settings.Deep neural networks were used to assess whether you are stressed or not.

In their research, Xu, Hansong, and colleagues set out to develop a secure electrocardiogram (ECG) information transmission system in order to limit additional damage to patients with cardiac ailments caused by human emotional stress[19]. The author suggested a dynamic encryption solution for ECG signals and their frequency spectrums based on biometric information that assures both a high categorization rate (>90%) and device energy efficiency. Cooperative relays were utilized concurrently to improve spatial variation. Based to the simulation results, increased transmission rate & signal power capacity can help reduce the risk of a data intercept (LPI) & detection (LPD) by leveraging both timescales and geographical diversities. As a result, security of the network may be improved even more.

The goals of H.Chao et al.'s study were to give a more accurate classification strategy for detecting stress states utilizing raw electrocardiogram (ECG) data and to show how to train a DNN utilizing a smaller data set[20]. The author provided a complete system for stress detection based on raw ECG data. The architecture was built in steps, each of which included convolutional layers.To train and evaluate the model, two different types of data sets were used: a driving record set and also a mental mathematics information set that was a bit shorter than the vehicular data set.To train an example with a small quantity of data, the author employed a transfer learning approach.In terms of how it calculates receiver operating curves, the proposed model beats previous methodologies. The suggested model increases reliability from 87.39 to 90.19 percent when compared to current deep neural network algorithms that employ raw ECGs.The transfer learning strategy increases reliability by 12.01 & 10.06 percent when 10 and 60 seconds of ECG data are included in the model, respectively.

The goal of N.Keshan et al.'s work was to focus on ECG monitoring, which can now be done using safe wearable patches and sensors, in order to construct an efficient and robust system for reliably diagnosing stress[21]. The planned study was unusual in that it included a personalized personal stress analysis at three distinct tension levels: High, Medium and Low. Using machine learning methods based solely on ECG data, the authors attained an accuracy of 88.24% in recognizing the three forms of stress. They also observed that a person's stress level may be correctly identified through contrasting it to the amount of his or her hours of sleep.

3. METHODOLOGIES:

In this part, we will go over the recommended techniques. Figure 3.1 depicts the total system architecture diagram.

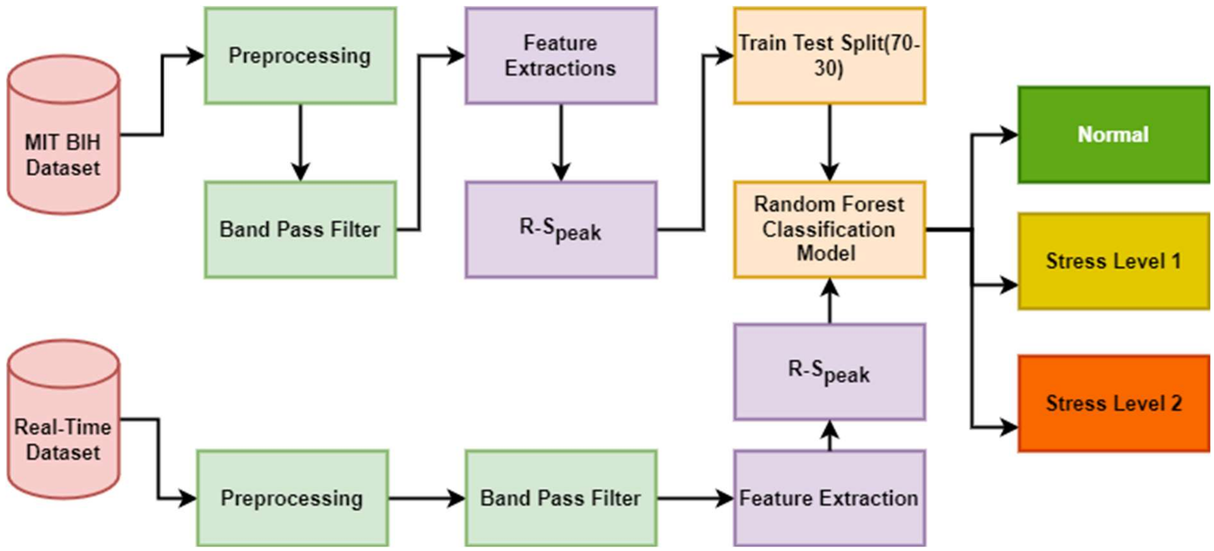

Figure 1: System Architecture

3.1 Dataset Description: The proposed study uses MIT-BIH Database for training, where ECG recording of 47 patients were taken. Twenty-three recordings were chosen at random from a set of 4000 24-hour mobile ECG recordings gathered at Boston's Beth Israel Hospital from a variety of groups of patients (approximately 60%) and outpatients (approximately 40%).Over a 10 mV range, the video recordings were digitized at 360 bits per second per lane with 11-bit resolution. Each record was annotated individually by two or more cardiologists; disagreements were resolved in order to get the computer-readable references annotations for each heartbeat (about 110,000 evaluations in total) included in the database.

For the real-time dataset, we collected ten samples of electrocardiogram (ECG) data from one of the most prestigious hospitals in the Mumbai region of India.The datasets of 5 girls and 5 males were used to validate our model. This data set's description is comparable to that of the previously reported MIT-BIH dataset, which was utilized for training purposes.

3.2 Preprocessing and Feature Extractions: Electrocardiography is a noninvasive way of assessing a person's health that examines the electrical status of the heart. Several variables cause noise during an EKG, severely lowering ECG classification accuracy[21]. We used a band-pass filter to solve this problem and discovered that a band-pass filter has a sample rate of 360 Hz & a cutoff frequency of 150 Hz removed 90.89% of the noise.

Figure 3.2 shows R Speak values obtained from an ECG wave. dividing these data during and without stress allows for a more reliable examination of the ECG[22].Rpeak & Speak were derived from an electrocardiogram (ECG) after a threshold was applied. Rpeak extracted the pole if the threshold level was greater than or equal to 0.2mV in one time frame of the signal; if the criteria value were less than 0.54mV in one time frame, Rpeak extracted the pole. The heart beats irregularly and quickly during stress, the gap between the R and R parts of the ECG signal narrows, while the R Speak increases.In the unstressed condition, on the other side of the hand, the heart is generally steady, the R-R gap expands, and the R Speak lowers[23]. The standard deviation (SD) of R Speak with stress was 1.47 mV in each test, and 4.25 mV when stressed.Figure 3 demonstrates the transition of either signal (stressed or not) into a spectrogram.

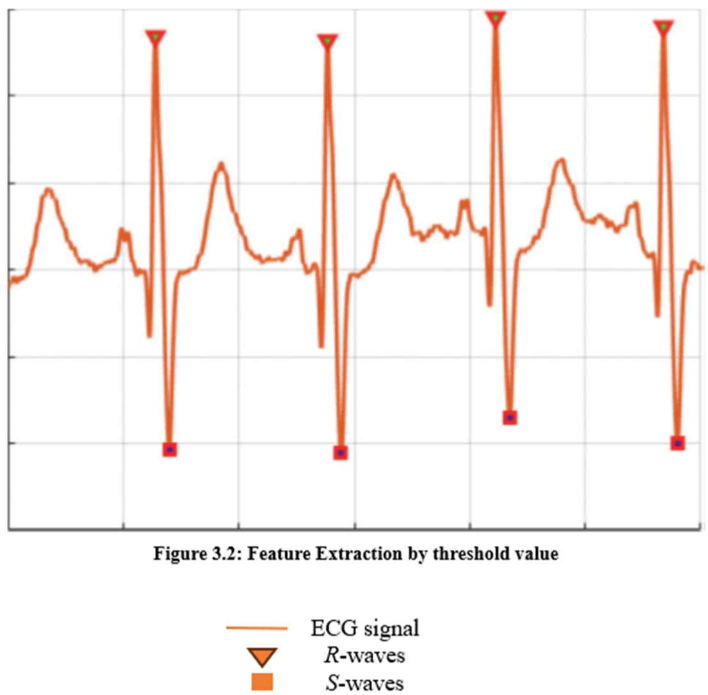

Figure 3.2: Feature Extraction by threshold value

3.3 Model Designing:

We utilized most of the machine learning classification model for categorization of stress and from the result analysis of performance of each model, we selected the best model as Random Forest based on key performance metrics such as categorization accuracy, precision, recall, and F1-Score.

3.3.1 Logistic Regression:

In binary classification jobs, the logistic Linear classification technique is used to anticipate the value of an element Y, which can have two possible values: 0 or 1. When there can be more than two possible values for Y, this strategy can be used to solve multi-classification problems.The logistic regression equation below calculates the chance that input X falls in category 1.:

$$P(X) = \frac{\exp(\beta_0+\beta_1 X)}{(1+\exp(\beta_0+\beta_1 X)} \quad (3.1)$$

Here β0 is bias and β1 is the weight that is multiplied by input X.

3.3.2 Naïve Byes:

In this type of probabilistic classifier, the Bayes theorem is applied. With its categorization skills, it can, for example, categorize, filter spam emails, and diagnose illnesses.The foundation of this strategy is that the characteristics that contribute to prediction are uncorrelated with one another. It use the following formula to calculate the class's chances of success:

$$P(C|I) = \frac{P(I|C)P(C)}{P(I)} \quad (3.2)$$

The possibility of class c occurring given input I is indicated using posterior probability. In the preceding equation, P (c) represents the prior class likelihood and P (I) represents the prior feature probability.When class c is present, the probability P (I|c) expresses the potential degree to which I will occur. This method may also be used to address multi-classification problems.

3.3.3 Support Vector Machine:

Classification is accomplished by constructing a hyperplane on which all samples belonging to one class will be located on one side, and all samples belonging to another class will be located on the other side. Hyperplane optimization ensures that the distance between classes is maximized. The data points closest to the hyperplane form a support vector.

Hyperplane can be created as given in the following equation:

$$H_0: w^T + b = 0 \quad (3.3)$$

Two more hyperplanes H1 and H2 are created in parallel to the constructed hyperplane as given in the following equations:

$$H_1: w^T + b = -1 \quad (3.4)$$

$$H_2: w^T + b = 1 \quad (3.5)$$

Hyperplane should satisfy the constraints given by following equations for each input vector,

$$wI_j + b \geq +1 \; for \; I_j \; having \; class \; 1 \quad (3.6)$$

And

$$wI_j + b \geq -1 \; for \; I_j \; having \; class \; 0 \quad (3.7)$$

3.3.4 Random Forest:

To increase prediction accuracy, a bagging approach known as Random Forest is employed to blend several decision trees.Individuals are taught bagging on their own. In this technique, numerous data samples are constructed from a single dataset using replacement, and each of the decision trees is trained on a different set of data samples. The tree's characteristics are also picked at random throughout the construction process.A majority vote can be used to aggregate the forecasts of numerous trees. Increasing the random forest's accuracy through optimizing variables such as the number of estimation tools, the smallest possible size of the node, and the number of attributes used to partition nodes.

3.3.5 KNearest Neighbours(KNN)

KNN (n neighbours = 5) is a popular classification technique in machine learning.It has previously been used as a therapy for coronary artery disease. KNN is considered nonparametric since it makes no presumptions about the variance of the data.Using the machine learning (ML) method, new data is assigned to the class that is most similar to the existing classes. KNN's use helps with both regression and recognition problems. It is commonly referred to as the lazy learners algorithm due to the fact that it does not learn from a set of data that has been trained immediately. KNN uses the following equation to calculate the length of the Euclidean distance between new A (x1, y1) and the previously accessible B (x2, y2) data.

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (3.8)$$

The Euclidean formula may be used to compare the points (x2, x1) and (y2, y1) in a three-dimensional space. The new information is sorted into categories using KNN based on how near they are geometrically to one other.

3.3.6 XGBoost:

It is a decision tree solution that has been gradient-boosted for speed and performance. XGBoost was created as a networked gradient boosting library to improve efficiency, flexibility, and portability.Machine learning algorithms are created using the Gradient Boosting framework. XGBoost provides extremely fast and precise parallel tree augmentation (commonly referred to as GBDT, GBM) for a wide range of data science problems.XGBoost is a strategy for increasing group learning.In some circumstances, depending just on a single model based on machine learning may be insufficient.Through ensemble learning, the predictive power of several learners may be systematically merged. Finally, you will have an all-encompassing model that integrates the findings of several previous models.

The group, which is also referred to as a basic learner set, may be built using a variety of learning approaches.Bagging and boosting are the two most prevalent ensemble learning strategies. These two techniques can be used in various statistical models, but decision trees are by far the most commonly used.

3.3.7 Stockist Gradient Descent

Stochastic gradient descent, one of the most extensively used methods in machine learning, acts as like starting point for neural networks. If the slope of a function is equal to zero and the optimal locations cannot be found by equal it to zero, the gradient descent strategy can be utilized.This procedure starts at an indeterminate position on a curve and gradually declines until it reaches its lowest point.

4. Experimental Results:

Table 4.1 displays the accuracy, accuracy, specificity, sensitivity, and precision predicted values derived using formulae 4.1 to 4.4 [26-28] to evaluate each classification model's performance.Formula (4.1) specifies accuracy and gives the likelihood of correctly identifying all under pressure and without stress.In the formula, the initials TP, TN, FP, and FN stand for true beneficial, true negative, false beneficial, & false negative, respectively.

$$Accurcay = \frac{TP+TN}{TP+TN+FP+F} \quad (4.1)$$

The fraction of data accurately identified as stress-free to all stress-free information (actual observed data) is referred to as sensitivity.

$$Sensitivity = \frac{TP}{FN+TP} (4.2)$$

The fraction of data accurately categorized as under pressure among all under pressure data (actual seen data) is referred to as specificity.

$$Specificity = \frac{TN}{TN+F} (4.3)$$

Precision is defined as the ratio of data accurately categorized as stress-free by the stress categorization method to the total value of data classified as stress-free.

$$Precision = \frac{TP}{TP+} (4.4)$$

Model Name	Precision	Recall	F1-Score	Accuracy
Logistic Regression	95.23%	70.23%	74.87%	90%
Support Vector Classification (SVC)	93.93%	73.80%	79.99%	86%
Decision Tree(DT)	63.33%	66.66%	64.91%	93%
Random Forest (RF)	98.24%	83.33%	87.98%	96%
XGBoost	63.33%	66.66%	64.91%	93%
Stockiest Gradient Descent (SGD)	98.24%	83.33%	87.98%	6%
k-nearest neighbors(KNN)	63.33%	66.66%	64.91%	93%
Naïve Bayes	86.18%	93.38%	88.48%	91%

According to table 4.1, Random Forest performed the best for our MIT BIH Dataset with 98.00% accuracy when compared to other machine learning models, so we concluded that model for testing the real-time dataset obtained from one of the well-known cardiologists in the Vidarbha Region of Maharashtra, India.We used the random forest model to test our model on the real-time dataset provided by the cardiologist, and we achieved 98.00% testing accuracy. We were able to quickly determine the stress levels of the patients who were classified as anxiety level 1, tension level 2, and normal patients.

5.DISCUSSIONS: An improved combined random forest model (RF model) was tested on real-time datasets in this work to improve the effectiveness of stress categorization and reduce overfitting.To improve stress classification accuracy, the properties of normal, stress level 1, as well as stress level 2 are taken into account utilizing the average value of the R Component of the ECG signal.We discovered that the random forest model yields 98.00% accuracy while training the MIT-BIH Datasets and evaluating the real-time dataset.

6. CONCLUSION: In this work, we suggested an enhanced Random Forest (RF) model by implanting a bandpass filter to remove noise and drawing accurate R-Speak features to increase accuracy. The suggested RF model exhibited 98.3% accuracy in stress categorization. In the future, we plan to improve the conditioning method by removing subtle noise compared to biological signals and further identifying ECG signals into time domain and frequency domain respectively, which will improve the clearness of ECG signals and allow us to accurately classify the stress level separately.In the future, we are also going to expand the size of the testing dataset, which will aid in making the model more general for everyday people.The stress classifier offered by us is expected to be beneficial in mental health treatment since it can quickly and efficiently define the stress which exists in modern persons, as we have previously tested on real-time datasets.Through periodic stress management, it is also intended to aid in the prevention of many ailments such as anxiety, high blood pressure, depression, and diabetes.

Acknowledgment: Dr. Kapil Rathi (Consultant Physician, Cardiac Specialist) from My Health Clinic in Mumbai, India, provided the real-time dataset of ECG signals for this study.

REFERNCES:

[1] M. Chauhan, S. V. Vora, and D. Dabhi, "Effective stress detection using physiological parameters," *Proc. 2017 Int. Conf. Innov. Information, Embed. Commun. Syst. ICIIECS 2017*, vol. 2018-Janua, pp. 1–6, 2017, doi: 10.1109/ICIIECS.2017.8275853.

[2] R. Ribeiro and J. P. Cunha, "A regression approach based on separability maximization for modeling a continuous-valued stress index from electrocardiogram data," *Biomed. Signal Process. Control*, vol. 46, pp. 33–45, Sep. 2018, doi: 10.1016/j.bspc.2018.06.006.

[3] P. M. Mohan, V. Nagarajan, and S. R. Das, "Stress measurement from wearable photoplethysmographic sensor using heart rate variability data," *2016 International Conference on Communication and Signal Processing (ICCSP)*. IEEE, 2016. doi: 10.1109/iccsp.2016.7754331.

[4] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *J. Biomed.*

Inform., vol. 59, pp. 49–75, 2016, doi: https://doi.org/10.1016/j.jbi.2015.11.007.

[5] S. Ishaque, N. Khan, and S. Krishnan, "Trends in Heart-Rate Variability Signal Analysis," *Front. Digit. Heal.*, vol. 3, p. 13, 2021, doi: 10.3389/fdgth.2021.639444.

[6] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, and N. Arunkumar, "Using Deep Convolutional Neural Network for Emotion Detection on a Physiological Signals Dataset (AMIGOS)," *IEEE Access*, vol. 7, no. c, pp. 57–67, 2019, doi: 10.1109/ACCESS.2018.2883213.

[7] A. S. Anusha, J. Jose, S. P. Preejith, J. Jayaraj, and S. Mohanasankar, "Physiological signal based work stress detection using unobtrusive sensors," *Biomed. Phys. Eng. Express*, vol. 4, no. 6, p. 65001, 2018, doi: 10.1088/2057-1976/aadbd4.

[8] M. Li, L. Xie, and Z. Wang, "A transductive model-based stress recognition method using peripheral physiological signals," *Sensors (Switzerland)*, vol. 19, no. 2, 2019, doi: 10.3390/s19020429.

[9] V. Montesinos, F. Dell'Agnola, A. Arza, A. Aminifar, and D. Atienza, "Multi-Modal Acute Stress Recognition Using Off-the-Shelf Wearable Devices," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 2196–2201, 2019, doi: 10.1109/EMBC.2019.8857130.

[10] B. Mahesh, E. Prassler, T. Hassan, and J. U. Garbas, "Requirements for a Reference Dataset for Multimodal Human Stress Detection," *2019 IEEE Int. Conf. Pervasive Comput. Commun. Work. PerCom Work. 2019*, pp. 492–498, 2019, doi: 10.1109/PERCOMW.2019.8730884.

[11] J. Zhang, W. Wen, F. Huang, and G. Liu, "Recognition of Real-Scene Stress in Examination with Heart Rate Features," *2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*. IEEE, 2017. doi: 10.1109/ihmsc.2017.13.

[12] B. Ramani, "An Energy-Efficient Stress Monitoring System Using Physiological Data," vol. 9, no. 8, pp. 132–137, 2021.

[13] L. Maršánová *et al.*, "ECG features and methods for automatic classification of ventricular premature and ischemic heartbeats: A comprehensive experimental study," *Sci. Rep.*, vol. 7, Dec. 2017, doi: 10.1038/s41598-017-10942-6.

[14] H. B, Y. J, V. T, M.-G. I, P. C, and Z. BT, "Deep ECGNet: An Optimal Deep Learning Framework for Monitoring Mental Stress Using Ultra Short-Term ECG Signals," *Telemed. J. E. Health.*, vol. 24, no. 10, pp. 753–772, Oct. 2018, doi: 10.1089/TMJ.2017.0250.

[15] V. Malhotra and M. Sandhu, "Improved ECG based Stress Prediction using Optimization and Machine Learning Techniques," *ICST Trans. Scalable Inf. Syst.*, p. 169175, Jul. 2018, doi: 10.4108/EAI.6-4-2021.169175.

[16] Z. Ahmad, S. Rabbani, M. R. Zafar, S. Ishaque, S. Krishnan, and N. Khan, "Multi-level Stress Assessment from ECG in a Virtual Reality Environment using Multimodal Fusion," Jul. 2021, Accessed: Sep. 22, 2021. [Online]. Available: https://arxiv.org/abs/2107.04566v1

[17] P. Karthikeyan, M. Murugappan, and S. Yaacob, "Detection of human stress using short-term ECG and HRV signals," *J. Mech. Med. Biol.*, vol. 13, no. 2, pp. 1381–1384, 2013, doi: 10.1142/S0219519413500383.

[18] P. Singh, A. Wadhwani, and S. Wadhwani, "Stress estimation of ECG signal by using DWT," *Int. J. Bio-Science Bio-Technology*, vol. 7, pp. 277–282, Oct. 2015, doi: 10.14257/ijbsbt.2015.7.5.26.

[19] H. Xu and K. Hua, "Secured ECG signal transmission for human emotional stress classification in wireless body area networks," *EURASIP J. Inf. Secur. 2016 20161*, vol. 2016, no. 1, pp. 1–12, Jan. 2016, doi: 10.1186/S13635-015-0024-X.

[20] H.-M. Cho, H. Park, S.-Y. Dong, and I. Youn, "Ambulatory and Laboratory Stress Detection Based on Raw Electrocardiogram Signals Using a Convolutional Neural Network," *Sensors* , vol. 19, no. 20. 2019. doi: 10.3390/s19204408.

[21] N. Keshan, P. V Parimi, and I. Bichindaritz, "Machine learning for stress detection from ECG signals in automobile drivers," in *2015 IEEE International Conference on Big Data (Big Data)*, 2015, pp. 2661–2669. doi: 10.1109/BigData.2015.7364066.