Abstract: - Mental stress is a significant issue affecting individuals across all age groups in today's society. The impact of stress extends beyond mental well-being and can lead to various chronic illnesses, including depression, malignancy, and cardiovascular disease (CVD). Addressing mental stress and implementing strategies for stress management and prevention are crucial not only for mental well-being but also for reducing the risk of associated chronic illnesses. In this paper, we are focusing on the problem of mental stress in today's society and the significance of early prediction and management. The use of Random Forest (RF) method with enhanced band pass filtration technique on ECG data for predicting stress levels is a promising approach. Achieving a 96.73% accuracy in stress categorization, along with improvements over prior research results, highlights the effectiveness of the proposed model. The main contribution of the research work is that the results of the model are validated using real-time datasets which further strengthens its reliability and applicability in practical scenarios.

Keywords: Mental Stress, Random Forest, ECG Signals, Real time dataset

1. Introduction:

Stress typically arises when an individual's mental and physiological capacities are insufficient to meet the demands placed upon them [1]. This can occur when an individual is confronted with unfamiliar material or when a student is faced with multiple unresolved problems within a limited timeframe for a final examination. Stress is a pervasive phenomenon observed in contemporary society, exerting a significant impact on the everyday lives and overall well-being of individuals. Based on data from the American Physiological Association and American Institute of Stress, in 2014, 77% of individuals in the United States consistently had physical symptoms, while 73% experienced psychological symptoms. These symptoms were attributed to stress [2]. In addition, 33% of individuals reported experiencing severe stress, while 48% of individuals acknowledged that stress has detrimental effects on both their personal and professional spheres. In the EU, a comparable situation occurred, with over 22% of employees perceiving their health as being jeopardized due to their demanding work [3]. The ensuing expenses incurred due to stress-induced healthcare and absenteeism were incalculable.

In the United States, it amounted to a staggering 300 billion USD annually[2]. In the United Kingdom, an annual loss of 13 million working days occurred, resulting in a financial burden of 12 billion pounds [4]. Therefore, in order to alleviate stress and maintain a state of well-being, the identification and control of stress have become imperative and indispensable for individuals of average means. The autonomic nervous system (ANS) would initially exhibit the physiological response to stress[5]. The ANS is composed of two branches, namely the sympathetic and parasympathetic divisions. The occurrence of stress events has the potential to disrupt the equilibrium between the two branches. The suppression of the parasympathetic branch would result in a hyper-activation of the sympathetic branch. The transmission of this information would be directed towards the heart activity, which may be quantified by the analysis of electrocardiogram (ECG) data. Therefore, the electrocardiogram (ECG) signal is commonly utilized as a primary indication for stress assessment. Heart rate variability (HRV) refers to the fluctuation in the time interval between two successive R peaks (heartbeats) of electrocardiogram (ECG) signals [6]. HRV is considered resilient to noise and disturbances due to its reliance solely on R peaks for computation, as these peaks exhibit the largest amplitude in
electrocardiogram (ECG) signals. Therefore, heart rate variability (HRV) is frequently employed as a diagnostic instrument for stress detection. In order to get a heart rate variability (HRV) measurement, it is often necessary to connect three electrocardiogram (ECG) electrodes to the right leg (RL), right arm (RA), and left leg (LL). These electrodes are used to capture the standard Lead II ECG signal\[7\]. After the identification of all the R peaks, it is possible to extract HRV characteristics for the purpose of stress detection \[8\]. Typically, a certain time period is set, and the characteristics are computed during that period. The research on stress detection using HRV typically use a window length measured in minutes \[9\],\[10\]. Therefore, the ability to make inferences about cognitive stress is limited to a minimum delay of one minute, which is not feasible for real-time stress detection.

Numerous techniques have been suggested for the purpose of feature extraction, drawing upon physiological manifestations of stress such as heightened heart rate\[9\], alterations in the balance between the sympathetic and parasympathetic branches of the autonomic nervous system \[6\], and irregularities in heart rate rhythm \[11\], among others. Stress detection commonly involves measuring heart rate (HR) and calculating the power ratio between the low-frequency band (0.04 - 0.15 Hz) and the high-frequency band (0.15 _ 0.4 Hz), among other factors. Nevertheless, various investigations have indicated that the efficacy of a solitary feature is not uniform due to significant inter-participant variations. Consequently, the integration of multiple characteristics has been proven to yield superior performance in stress detection \[10\]. This finding suggests that individuals' reactions to stress might differ and have various expressions in heart rate variability (HRV). Subsequently, the distinguished characteristics were transmitted to the classifier for identification. Linear discriminant analysis (LDA) and support vector machine (SVM) are two often used traditional classifiers \[12\].

The convolutional neural network (CNN) is a type of artificial neural network (ANN) that commonly includes a convolutional layer inside its hidden layers \[13\]. The technique was first utilised in computer vision applications and has since garnered attention in various fields, including biosignal classification\[14\],\[15\]. For instance, it has been employed in the classification of electromyogram (EMG) signals for gesture recognition\[16\] and the identification of electroencephalogram (EEG) patterns for assistive machine control \[14\]. CNN has been effectively utilised in several applications such as arrhythmia detection\[17\], signal component identification \[18\], and biometric recognition \[19\] in the field of ECG signal categorization. These experiments showcased the efficacy of CNN in classifying bio signals, suggesting the potential of CNN in detecting stress based on ECG data.

To the authors' knowledge, this work is the first to utilise standard machine learning with the Random Forest approach for real-time detection of mental stress using ECG data. Specifically, it aims to infer mental stress from specific segments of ECG signals. Real-time stress detection has become a crucial requirement in several practical applications, particularly in the context of acute cognitive stress detection. This necessitates the ability to make prompt judgements.

The primary objective of this work was to investigate the identification of acute cognitive stress utilising the RF technique on ECG signals within a real-time dataset. We conducted a comparative analysis of the performance of the suggested RF and standard ECG based stress estimation techniques. Our findings demonstrate that our proposed approach outperformed the old methods in terms of accurately predicting mental stress on real-time datasets. The objective of this work was to enhance the practical implementation of electrocardiogram (ECG)-based stress measurement by offering expedited, dependable, and instantaneous outcomes.

2. Methodologies
In this part, we will go over the recommended techniques. Figure 2.1 depicts the total system architecture diagram.
2.1 Dataset Description:

2.1.1 Training Dataset:
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.

2.1.2 Testing Dataset:
As we need to work on real-time dataset, we contacted various cardiologist for getting the real time ECG signal data for stress classification and prediction. We Collected the dataset from one of the famous cardiologists from Mumbai, India region. They provided us the ECG signal dataset of 11 patients for real time stress analysis for our research work. All this 11 patient’s ECG signal dataset cannot be shown here cause of paper length constraint, but 2 patients’ data was shown in following figures 2.1 and 2.2.

2.2 Preprocessing And Feature Extractions: Electrocardiography is a non-invasive 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. 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 2.2 shows Rpeak values obtained from an ECG wave. Dividing these data during and without stress allows for a more reliable examination of the ECG. 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 0.2 mV in one time frame of the signal; if the criteria value were less than 0.54 mV 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 Rpeak increases. In the unstressed condition, on the other side of the hand, the heart is generally steady, the R-R gap expands, and the Rpeak lowers. The standard deviation (SD) of Rpeak with stress was 1.47 mV in each test, and 4.25 mV when stressed.

2.3 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.

2.3.1 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. Experimental Results:
After applying the random forest model for classification of stress on 11 patients real time dataset, we found the analysis of the patient’s records based on ECG signals of 30 sec ultrashort wavelength into the 3 classification stages as Normal (Stress Level 0), Moderate (Stress Level 1) and High (Stress Level 2). The results are discussed in table 3.1.
Table 3.1: Real Time Data Stress Prediction

<table>
<thead>
<tr>
<th>Patient</th>
<th>BPM</th>
<th>P(s)</th>
<th>PR(s)</th>
<th>QRS(s)</th>
<th>Execution Time</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient1</td>
<td>90</td>
<td>0.111</td>
<td>0.188</td>
<td>0.04</td>
<td>0.17</td>
<td>Moderate</td>
</tr>
<tr>
<td>Patient2</td>
<td>91</td>
<td>0.102</td>
<td>0.186</td>
<td>0.099</td>
<td>0.21</td>
<td>Moderate</td>
</tr>
<tr>
<td>Patient3</td>
<td>86</td>
<td>0.115</td>
<td>0.185</td>
<td>0.089</td>
<td>0.19</td>
<td>Moderate</td>
</tr>
<tr>
<td>Patient4</td>
<td>64</td>
<td>0.101</td>
<td>0.159</td>
<td>0.092</td>
<td>0.14</td>
<td>Moderate</td>
</tr>
<tr>
<td>Patient5</td>
<td>48</td>
<td>0.116</td>
<td>0.18</td>
<td>0.09</td>
<td>0.28</td>
<td>High</td>
</tr>
<tr>
<td>Patient6</td>
<td>64</td>
<td>0.094</td>
<td>0.198</td>
<td>0.093</td>
<td>0.15</td>
<td>Normal</td>
</tr>
<tr>
<td>Patient7</td>
<td>86</td>
<td>0.081</td>
<td>0.152</td>
<td>0.067</td>
<td>0.21</td>
<td>Moderate</td>
</tr>
<tr>
<td>Patient8</td>
<td>106</td>
<td>0.072</td>
<td>0.151</td>
<td>-0.05</td>
<td>0.2</td>
<td>High</td>
</tr>
<tr>
<td>Patient9</td>
<td>103</td>
<td>0.101</td>
<td>0.145</td>
<td>0.103</td>
<td>0.277</td>
<td>Moderate</td>
</tr>
<tr>
<td>Patient10</td>
<td>68</td>
<td>0.094</td>
<td>0.157</td>
<td>0.084</td>
<td>0.207</td>
<td>Moderate</td>
</tr>
<tr>
<td>Patient11</td>
<td>59</td>
<td>0.107</td>
<td>0.164</td>
<td>0.144</td>
<td>0.2477</td>
<td>High</td>
</tr>
</tbody>
</table>

Out of 11 patient’s ECG prediction for stress, we found only 1 patient as normal, 3 patients with high stress and rest 7 patients having moderate level of stress. The Random Forest model has produced 96.73% accuracy for the stress classification and prediction when tested on real time dataset which is higher comparative with other studies with real time classification of the stress using ECG Signals.

Figure 3.1 show the stress prediction pie chart for real time dataset of 11 patients. We found that 64% patients are under moderate stress, 27% are under high stress and 9% are under normal stress.

When we done the comparative analysis with our research work with most recent research work then we found that our model has performed better compared to all the recent research work. We are comparing on the basis of their model’s performance on ECG dataset for stress classification. Following table 3.2 summarizes the comparative analysis with most recent work.
Table 3.1: Comparative Analysis of the proposed work

<table>
<thead>
<tr>
<th>Reference</th>
<th>Accuracy</th>
<th>Model</th>
<th>Type of Signals</th>
<th>Window Size</th>
<th>Stress Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>[20]</td>
<td>90.19</td>
<td>CNN</td>
<td>ECG</td>
<td>10 s</td>
<td>2</td>
</tr>
<tr>
<td>[21]</td>
<td>89.8</td>
<td>CNN</td>
<td>ECG</td>
<td>60 s</td>
<td>2</td>
</tr>
<tr>
<td>[22]</td>
<td>87.39</td>
<td>CNN-RNN</td>
<td>ECG</td>
<td>10 s</td>
<td>2</td>
</tr>
<tr>
<td>[23]</td>
<td>83.9</td>
<td>CNN</td>
<td>ECG and RSP</td>
<td>50 s</td>
<td>2</td>
</tr>
<tr>
<td>[24]</td>
<td>82.7</td>
<td>CNN</td>
<td>ECG</td>
<td>10 s</td>
<td>2</td>
</tr>
<tr>
<td>[25]</td>
<td>92.8</td>
<td>CNN</td>
<td>ECG</td>
<td>25 s</td>
<td>3</td>
</tr>
<tr>
<td>[26]</td>
<td>86.5</td>
<td>CNN-BiLSTM</td>
<td>ECG</td>
<td>10 s</td>
<td>3</td>
</tr>
<tr>
<td>[27]</td>
<td>85.45</td>
<td>CNN</td>
<td>ECG</td>
<td>30 s</td>
<td>3</td>
</tr>
<tr>
<td>[28]</td>
<td>80.00</td>
<td>SVC</td>
<td>ECG</td>
<td>60 s</td>
<td>2</td>
</tr>
<tr>
<td>[29]</td>
<td>98.5</td>
<td>KNN</td>
<td>ECG</td>
<td>60s</td>
<td>2</td>
</tr>
<tr>
<td>Proposed Work</td>
<td>96.73</td>
<td>Random Forest</td>
<td>ECG</td>
<td>30s</td>
<td>3</td>
</tr>
</tbody>
</table>

Figure 3.2: Comparative Analysis

4. Conclusion: By applying the random forest which was best performing model during the training our database, on the real time database procured by the cardiologist from Mumbai, the RF model produced 96.73% accuracy. When tested with ECG signal database of 11 patients, we found that 1 patient got into the normal stress classification, 7 into Moderate stress and 3 into high stress classification. This is one of the unique studies as per our knowledge who applied Stress classification with Random Forest classification algorithm with 96.73% accuracy on real time database with early prediction of average prediction time as 0.17 seconds.

Acknowledgement: We sincerely thanks to Dr. Kapil Rathi (Consultant Physician, Cardiac and Chest Specialist) from My Health Clinic, Andheri(E), Mumbai, India for providing the real time dataset of ECG Signals.

REFERENCES:


