Scope
Volume 13 Number 01 March 2023

Ensemble techniques for mental stress prediction

Sri. S. Naga Raju [1], Dr Nandam Gayatri[2],Sameer Mohammed[3], Sumana Sree Dandigam[4],Harsha Vardhan Adluri[5], Vijay Komma[6]

[1]AssociateProfessor,[2] Assistant Professor, [3,4,6,5]Students

[1,2,3,4,5,6] Department of Computer Science and Engineering, KITSW(Affiliated to Kakatiya University),Warangal, India-506015

Abstract: Mental stress became a social problem and may contribute to functional impairment at regular job. Chronic stress may also be a factor in a number of psychophysiological illnesses. Stress, for instance, raises the risk of heart attack, stroke, and depression. According to the most recent research in neuroscience, brain is the ultimate target of the stress since how the brain perceives a situation affects how harmful and stressful it is. In this situation, a measurement that is objective and takes into account the human brain could greatly reduce the negative impacts of the stress. A machine learning (ML) framework using the ECG and EMG of participants who are stressed is therefore suggested. The task performance and subjective feedback supported the induction of stress. The framework for suggested ML included feature selection, extraction, and classification The findings demonstrated that the suggested framework increased the accuracy of stress recognition. The suggested approach may contribute to the creation of the computer-aided diagnostic tool for stress prediction.

Keywords: 1.Machine Learning, 2.Ensemble Technique, 3.Decision Tree, 4.Gaussian NB, 5.KNN, Voting.

I. Introduction

A popular definition of stress is a condition in which a person feels overly pressured to perform and is only just able to keep up with the obligations. These requirements may be societal or psychological. Psychosocial stress is a known component of daily living that negatively impacts people's emotional behavior, ability to function at work, and mental and physical health. Numerous physiological illnesses are mostly brought on by psychosocial stress. For instance, it makes depression, stroke, heart attacks, and cardiac arrest more likely.

Stress must first be quantified into levels in order to be treated. In clinical settings, subjective techniques like questionnaires and interviews have been used to assess stress. As an alternative, the stress related physiological and the physical changes are being used as quantifiable stress predictors. Physically, for instance, stress alters the pace of blinking, the size of the pupil, and facial expressions. Contrarily, stress results in modifications to the autonomic nerve system (ANS). Therefore, the HR and heart rate variability, therespiration, and the skin conductance are physiological indicators of stress from the ANS.

The brain is the basic target of the mental stress, according to the most recent findings of neuroscience, because it is the human brain that identifies if the situation seems threatening and also stressful too. The cortex to get the best tools for measuring functional changes in brain in theresponse to the stress are not invasive methods like electroencephalography (EEG). It's significant that EEGs have demonstrated an implied relationship with other stress predictors like HR and heart rate variability generally and specifically in stress.

Several electrophysiological properties from EEG signals have been retrieved by several research in order to quantify stress from the EEG signals, and also the classification algorithms are used with highest accuracy of 96% to distinguish the two levels of the stress. These findings support the idea that EEG can be used as a stress evaluation tool. Recent research has demonstrated that stress can alter connection parameters including coherence and mutual information as well as the absolute power of the EEG. Similar findings have been made regarding the impact of HRV biofeedback during stress therapy on asymmetry in EEG alpha power. In another researchstudy, the EEG alpha asymmetry was explored, and the stressrelated issues were found in the virtual reality setting.

II. Literature Survey

There are various medical theories for detecting stress in medical field, for example doctors can find whether a person is stressed or not by asking them few questions like whether they are having symptoms like headache or lack of sleep andalso,they can use various tools like ECG (electrocardiogram). ECG is a medical tool used to record the hearts Rhythm or electronic signals in heart. Some authors predicted whether a person is stressed or not by Calculating Heart rate, Electromyography (EMG) which is a diagnostic test to check patient's muscles and nerves working, Galvanic Skin Response (GSR) test for hand and foot response data, Respiration rate and came to a conclusion that respiration is a important and most critical parameter in detecting stress in a person.

J. Li et al. (2018) published in his research in order to forecast mental stress levels, this study suggests an ensemble strategy that incorporates various classifiers.

S. K. Pal and S. Mitra's research, in order to detect mental stress, this research suggests an ensemble strategy that integrates various EEG signal processing approaches. Using a dataset of EEG signals, the strategy underwent evaluation, and it performed better than individual techniques.

Using EEG signals, the study "Ensemble of Deep Convolutional Neural Networks for Mental Stress Classification" by S. Sahu et al. (2020) suggests an ensemble strategy that incorporates various deep convolutional neural networks. Using a dataset of EEG signals, the method was tested, and it performed better than individual networks.

N. R. Mudunuri and colleagues research, this study suggests an ensemble strategy that incorporates various mental health classifiers.

III. Existing System

Machine learning has been investigated in a number of existing systems for the prediction of mental stress. These systems assess physiological signals or the EEG signals and forecast mental stress levels using a variety of the machine learning algorithms, includingthe decision tree algorithm, the support vector machines, the neural network algorithm, and convolutional neural networks. Several of the systems also employ ensemble approaches to boost prediction accuracy. The testing of these algorithms on various datasets demonstrates that machine learning is capable of accurately predicting degrees of mental stress. Depending on the method selected and the kind of signals being examined, the systems' performance varies. Overall, these algorithms demonstrate promising results in reliably forecasting mental stress levels, which can be helpful in a number of applications, including stress management and mental health monitoring.

Some existing systems include various forms of data like speech and facial expression cues for mental stress prediction in addition to physiological and EEG signals. One system, for instance, predicts mental stress levels using facial expressions as input to a deep neural network. Another approach forecasts levels of mental stress using voice inputs and the machine learning algorithms like decision tree algorithm and the support vector machines. These systems show the capability of multi-modal data processing for the prediction of mental stress.

Several applications, including stress management and performance enhancement, can benefit from this real-time monitoring.There are still significant difficulties in utilizing machine learning to predict mental stress despite the encouraging results. It might be difficult to compare the effectiveness of various systems due to the absence of uniform datasets for evaluation. The requirement for individualized models that can adjust to variations in stress reactions in each individual is another difficulty. Addressing these obstacles may result in machine learning mental stress prediction that is more precise and individualized.

In order to create responsible and reliable mental stress prediction systems, it is essential to protect the privacy and security of the data and prevent making biased predictions.

The interpretability of the models is a crucial component of mental stress prediction using machine learning. Although machine learning algorithms are quite accurate at predicting levels of mental stress, because of the complexity of the models, they are frequently referred to as "black boxes." In situations where comprehending the assumptions behind the prediction is key, interpretability is essential. Developing interpretable machine learning models for predicting mental stress levels has been the focus of some recent research. These models can shed light on the characteristics and signals that are crucial for accurately forecasting mental stress levels.

Furthermore, ethical issues are crucial in machine learning-based mental stress prediction. Data privacy, confidentiality, and potential discrimination are issues that are brought up by the use of physiological and other personal data to forecast mental stress.

The ability to reliably forecast stress levels utilizing physiological signals, EEG signals, facial expressions, and voice cues has showed promise in mental stress prediction using machine learning. These systems' effectiveness can be increased further by real-time tracking and customized models. Interpretability, data privacy and ethical issues, as well as dataset standardization, remain obstacles that must be overcome.

IV. Proposed System

In this project the main objective is to find stress in person at different levels of their life. The first step is dataset collection, Dataset which is used for implementing this project is already existing dataset which consists various values obtained from different sensors such as ECG (electrocardiogram) which measures hearts signals, Electromyography (EMG), GSR (Galvanic skin response) and other parameters. After collecting Dataset next step is pre-processing for removing redundancy so that dataset is ready to apply Machine Learning Algorithms. Next four Machine learning algorithms are applied for classification those are svm, the Decision Tree, Gaussian Naïve Bayes algorithm, the k-nearest neighbour(knn). Results obtained after applying these algorithm shows whether a person is stressed or not and also shows the accuracy. After that we apply ensemble techniques such as Bagging and boosting to find whether the accuracy of individual algorithm is higher or combined four Algorithm is higher.

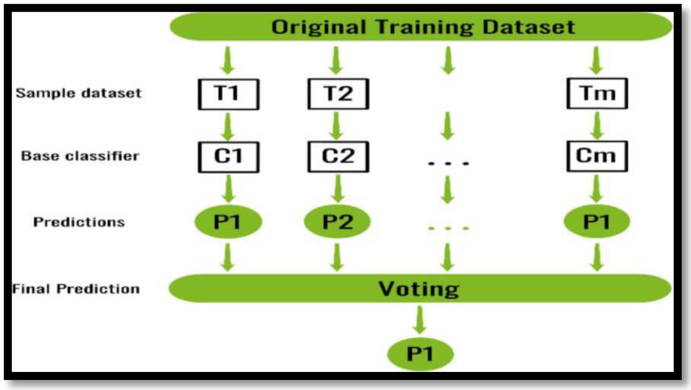
The predictions of these three classifiers would be combined in a suggested system for mental stress prediction using ensemble techniques of the decision tree, the Gaussian Naive Bayes, and K-Nearest Neighbors (KNN), which would increase the precision of mental stress prediction. Machine learning for classification tasks frequently uses decision trees, Gaussian Naive Bayes, and KNN classifiers. The decision-making process in a decision tree is based on a sequence of queries concerning the input data. A probabilistic model called Gaussian Naive Bayes makes the assumption that the characteristics are independent and have a Gaussian distribution. A non-parametric model called KNN predicts the class label based on the feature space's k nearest neighbors.

In the suggested system, the three classifiers' outputs would be integrated using an ensemble technique like majority voting or weighted voting after each classifier had been trained using a dataset of physiological signals or EEG signals. On a test dataset, the ensemble model's performance will be assessed in order to determine how well it predicts levels of mental stress.Several machine learning applications have demonstrated the benefit of ensemble approaches in increasing prediction accuracy. Several classifiers' predictions combined can lessen the chance of overfitting and increase the generalizability of the model. Ensemble models may also be more resistant to noisy or insufficient data.

Overall, a suggested method for predicting mental stress that combines decision tree, Gaussian Naive Bayes, and KNN ensemble algorithms has the potential to increase the precision of predicting mental stress from physiological or EEG inputs. However, the features of the data and the unique requirements of the application may affect the selection of the ensemble technique and the combination of classifiers.

V. Implementation

The Ensemble techniques are the methods that aim at the procedure of improving accuracy of the results in the machine learning models by combining many of the models instead of using the single machine learning model. The combined models will increase accuracy of results gradually. This has boosted popularity of the ensemble techniques in the machine learning. There are main reasons to use the ensemble technique over the single model, and they are related; they are: Performance: The ensemble can make the predictions better and achieve the better performance than single contributing machine learning model.
Robustness: The ensemble technique reduces the spread or the dispersion of predictions and the model performance.



Each machine learning model implementation starts with gathering pertinent data. In this situation, we require a dataset with both labels indicating the degree of stress and attributes that are suggestive of mental stress. Heart rate, breathing rate, body temperature, and cortisol levels are a few features that may be present in the dataset. Following preprocessing, the data has to be divided into the training and the testing sets. Testing set will be used to assess machine learning model's performance, while training set is used to train machine learning model. Typically, 80% of the data will be usedas the training and the 20% are used as testing. This guarantees that the model can generalize effectively to fresh, untested data and is not overfit to the training set of data.

The decision tree algorithm, the Gaussian Naive Bayes, and k-Nearest Neighbors classifiers should be implemented next. These classifiers are developed using a machine learning library, such as Python's scikit-learn. A common categorization approach called decision trees produces a model of decisions that looks like a tree.

Decision Tree Algorithm

The Decision tree is a algorithm which is classified as supervised learning algorithm. It is the tree type classifier, whose internal nodes stand in for the data set's characteristics, the branches for the algorithm's rules, and the leaf nodesfor the outcome. A decision tree often starts out by treating all of the data as its root. Then, depending on a certain circumstance, it begins to divide using internal nodes or branches and decides until the result is produced as a leaf. There is just one crucial fact to be aware of: when creating a tree, it adds information while reducing impurity in the characteristics to produce the desired results. Due to the algorithm's simplicity, it also has a few factors that are crucial for a data scientist to understand since they affect how well a decision tree functions when a model is finally built.

- **Entropy:** It is described as the degree of the impurity in the data. When the sample achieves homogeneity, entropy is nearly zero, but it is one when it is uniformly distributed. Because it better separates the classes, the model with the lowest entropy value is superior for prediction. The following formula is used to determine entropy. The quantity of classes is denoted by n. Entropy has a maximum in the center with the value of one and the lowest at the ends with the value of zero.
- **Information Gain:**It is the measure which is used to generalize the additive entropy of a data set. The greater the access to information, the lower the entropy. A low-probability event has lower entropy and high information, while a high-probability event has the higher entropy and the low information. It is calculated in the form
  Information_Gain = Entropy_of_Parent – sum (weighted % * Entropy_of _Child)

$$Entropy = \sum_{i=1}^{C} -p_i * \log_2(p_i)$$

K-Nearest Neighbour:

The implementation modules of this system are, In theK Nearest Neighbour (KNN) is a straightforward, basic, and versatile machine learning algorithm. It's used in a variety of applications, including handwriting recognition, image recognition, and video identification. KNN is most useful when labelled data is prohibitively expensive or unavailable, and it can achieve high accuracy in a broad range of prediction-type problems. KNN is a simple algorithm that uses the local minimum of the target function to learn an unknown function with the required precision and accuracy. The algorithm also determines the range or distance from an unknown input, as well as other factors. It is founded on the concept of "information gain," in which the algorithm determines which is best suited to predict an unknown value.

The KNN algorithm is a voting system in which the majority class label decides the class label of a new data point in the feature space among its nearest 'k' (where k is an integer) neighbors. Consider a small village of a few hundred people, and you must determine which political party to vote for. You could do this by approaching your nearest neighbors and asking which political group they support. If the majority of your' nearest neighbors back party A, you will almost certainly vote for party A as well. This is similar to how the KNN algorithm works, where the majority class label among a new data point's k nearest neighbors decides the class label.

Gaussian NB:

Based on the Bayes theorem, the probabilistic ML technique which is known as the Naive Bayes is employed for numerous classification applications.Gaussian The extension of naive Bayes is nave Bayes. The Gaussian distribution is most straightforward for implementing among the various functions used to estimate data distribution, as you only have to figure out mean and the standard deviation for the trained data. The Naive Bayes techniquedevelops the assumption that predictors each have to contribute in equal proportion and independently to select output class.The name "Naive" is used because the algorithm contains functions that are independent of each other in its model. Changes in value of the one property does not directly affect t value of any other property in the algorithm. The advantage of this Naive Bayes algorithm is it is simple but most powerful algorithm. For each row (x) of X and class Y, log probability will be

log P(x, y) = log P(y) + log P(x|y),

where the log P(y) is the class prior probability
log P(x|y) is class-conditional probability.

Experimental Setup
The performance measure of our mental stress prediction is done through these performance parameters,
Sensitivity, also referred to as true positive rate. The proportion of the true positives (TP) to the true positives and the erroneous negatives is what matters. (FN). This describes the model's capacity to identify diseases accurately.

Sensitivity= $TP/TP+FN$

Specificity: True negative rate is another name for this. According to equation 2, it is the proportion of true negatives (TN) to both true negatives and false positives (FP). This defines capacity for accurate identification in the absence of illnesses.

Specificity= $TN/TN+FP$

Accuracy: This is percentage of the true positives plus the true negatives to the true positives plus the true negatives plus the false positive plus the false negative. It determines percentage of the cases that are properly classified.

$Accuracy = TP+TN/TN+FP+TP+FN$

Last but not least, the model's performance must be assessed using the assessment measures including the accuracy, the precision, the recall, and the F1 score. The Accuracy is expressed as a proportion of instances that are correctly classified. Precision is a measurement of the percentage of events that are truly positive. The percentage of true positives that were correctly classified is measured by recall. A balanced assessment of the model's performance is provided by F1 score, which is harmonic mean of the precision and the recall.

In conclusion, collecting and preprocessing data, dividing data into the training and the testing sets, implementing the classifiers and the ensemble technique, and assessing the performance of the model using evaluation metrics are required to implement mental stress prediction using ensemble techniques with the decision tree, the Gaussian Naive Bayes, and the k-Nearest Neighbors.

Using ensemble techniques with decision trees, the Gaussian Naive Bayes, and the k-Nearest Neighbors, mental stress prediction must be implemented in a number of steps, including feature selection, hyperparameter tuning, addressing class imbalance, cross-validation, and interpretability. Using these methods, a model that accurately predicts mental stress levels based on pertinent features can be produced.

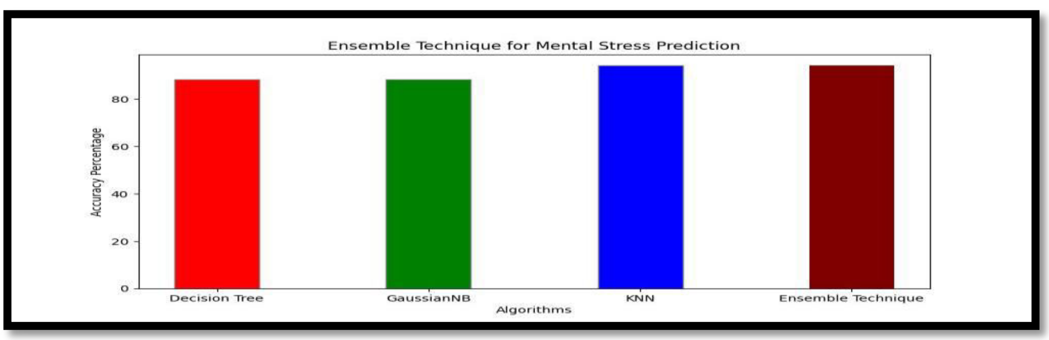
VI. Results

In this study, we used the Decision Tree, Gaussian NB, and K-Nearest Neighbour machine learning algorithms. We then computed the specificity, sensitivity, and accuracy of each approach. K-Nearest Neighbour outperformed the other three algorithms, with accuracy rates of 94.1%, 100% specificity, and 75% sensitivity. Decision Tree came in second, achieving accuracy rates of 88.2%, 88.9% specificity, and 0% sensitivity. Therefore, we may conclude that by using these three algorithms and by applying ensemble technique on these algorithms we achieve better accuracy.

According to the study, the ensemble model was more accurate than any single classifier. The study also discovered that several classifiers performed better in various situations, emphasizing the value of employing an ensemble method.

Sr. No.	Algorithm	Sensitivity (%)	Specificity (%)	Accuracy (%)
1	Decision Tree	0.0	0.889	0.882
2	Gaussian NB	1.0	0.85	0.882
3	K-Nearest Neighbours	0.75	1.0	0.941

Comparision of algorithms



From the above table and Bar graph , we can conclude that the ensembling technique gives more accuracy in predicting the mental stress compared to other individual algorithms such as Decision Tree,Gaussian Naïve bayes and KNN algorithms.

This research collectively imply that it may be possible to accurately forecast mental stress levels in a variety of circumstances, including clinical and military ones, by combining decision trees, Gaussian Naive Bayes, and k-Nearest Neighbors. These models can aid in the development of focused therapies and enhance the results related to mental health by identifying the elements that are linked to higher levels of stress. Like with any prediction model, it's crucial to take into account any data restrictions and biases and to evaluate the model across various demographics and contexts.

VII. Summary and Conclusions

After implementing this project by collecting dataset which consists of different signals from sensors and applying classification Algorithms and ensemble techniques, we can identify whether a person is stressed or not. Stress is the major problem these days regardless of age, mainly in students whose age is considered to be a carefree youth is now affected by stress. Stress in students can be of different reasons, stress may be due to exam pressure, placements, grumpyteachers, parents. Regardless of the reason identifying stress in early stage is important. By implementing this project, we can find whether a person is stressed or not by applying the four classification algorithms those are the Decision tree, Gaussian NB, K-Nearest Neighbour.We can find the target value that is either zero or one and accuracy of individual algorithms After applying ensemble techniques we will find whether accuracy of individual algorithm is higher or combining four machine learning algorithm is higher so that we can develop a model with high accuracy for predicting stress in a person.

As a future extension, block chain has wide range of

There are many potential directions for the future progress in the area of mental stress prediction using ensemble techniques with the decision tree, the Gaussian Naive Bayes, and the k-Nearest Neighbors. While ensembles can improve prediction accuracy, they can also be more difficult to interpret than individual models. Future work could explore methods for improving the interpretability of the ensemble model, such as using visualization techniques or feature selection algorithms. Mental stress levels can vary widely across individuals, and a one-size-fits-all approach may not be effective for predicting stress in all cases. Future work could explore methods for developing personalized models that take into account individual differences in stress response. Whilethe decision tree, Gaussian Naive Bayes, and the k-Nearest Neighbors are commonly used classifiers in ensembles, there may be other classifiers that could improve the accuracy of mental stress prediction.

VIII. References

1. Jung, Yuchae, and Yong Ik Yoon. "Multi-level assessment model for wellness service based on human mental stress level." Multimedia Tools and Applications 76.9 (2017): 11305-11317.

2. Norizam, Sulaiman. Determination and classification of human stress index using nonparametric analysis of EEG signals. Diss. UniversitiTeknologi MARA, 2015.

3. Subhani, Ahmad Rauf, Wajid Mumtaz, Mohamed Naufal Bin Mohamed Saad, Nidal Kamel, and Aamir Saeed Malik. "Machine learning framework for the detection of mental stress at multiple levels." IEEE Access 5 (2017): 13545-13556.

4. Ghaderi, A., Frounchi, J., Farnam, A.. Machine learning-based signal processing using physiological signals for stress detection. In: 2015 22nd Iranian Conference on Biomedical Engineering (ICBME). 2015, p. 93–98.
5. Gjoreski, M., Gjoreski, H., Lustrek, M., Gams, M.. Continuous stress detection using a wrist device: in the laboratory and real life. In: ˇProceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. ACM; 2016, p. 1185–1193.
6. Villarejo, M.V., Zapirain, B.G., Zorrilla, A.M. A stress sensor based on galvanic skin response (GSR) controlled by ZigBee. Sensors 2012; 12(5):6075–6101.
7. Selvaraj, N. Psychological acute stress measurement using a wireless adhesive biosensor. In: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2015, p. 3137–3140.
8. Xu, Q., Nwe, T.L., Guan, C.. Cluster-based analysis for personalized stress evaluation using physiological signals. IEEE Journal of biomedical and health informatics 2015; 19(1):275–281.
9. Khosrowabadi, Reza, Chai Quek, Kai Keng Ang, Sau Wai Tung, and Michel Heijnen. "A Brain-Computer Interface

for classifying EEG correlates of chronic mental stress." In IJCNN, pp. 757-762. 2011.

10. Adnan, Nadia, et al. "University students stress level and brainwave balancing index: Comparison between early and end of study semester." Research and Development (SCOReD), 2012 IEEE Student Conference on. IEEE, 2012.
11. G. C. Dieleman et al., ''Alterations in HPA-axis and autonomic nervous system functioning in childhood anxiety disorders point to a chronicstress hypothesis,'' Psychoneuroendocrinology, vol. 51, pp. 135–150,jan. 2015.
12. A. R. Subahni, L. Xia, and A. S. Malik, ''Association of mental stress withvideo games,'' in Proc. 4th Int. Conf. Intell. Adv. Syst. (ICIAS), Jun. 2012,pp. 82–85.
13. J. Wielgosz, B. S. Schuyler, A. Lutz, and R. J. Davidson, ''Long-termmindfulness training is associated with reliable differences in resting respiration rate,'' Sci. Rep., vol. 6, pp. 1–6, 2016.
14. A. Liapis, C. Katsanos, D. Sotiropoulos, M. Xenos, and N. Karousos,''Recognizing emotions in human computer interaction: Studying stressusing skin conductance,'' in Human-Computer Interaction (Lecture Notesin Computer Science), vol. 9296, J. Abascal, S. Barbosa, M. Fetter,T. Gross, P. Palanque, and M. Winckler, Eds. Cham, Switzerland: Springer,2015, pp. 255–262.
15. K. Dedovic, R. Renwick, N. K. Mahani, V. Engert, S. J. Lupien, andJ. C. Pruessner, ''The montreal imaging stress task: Using functionalimaging to investigate the effects of perceiving and processing psychosocial stress in the human brain,'' J. Psychiatry Neurosci., vol. 30, no. 5,pp. 319–325, 2005.
16. A. R. Subhani, L. Xia, and A. S. Malik, ''Association of electroencephalogram scalp potential with autonomic nervous system,'' Adv. Sci. Lett., vol. 19, no. 5, pp. 1340–1344, 2013.