Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The Prediction of Mental Stress by Utilizing Supervised and Unsupervised Machine Learning Techniques

Amanpreet Kaur [1] Dr. Kamal Malik [2]

[1] Research Scholar, Computer Applications, CT University, Ludhiana, Punjab 142024
amanp9675@gmail.com

[2]Professor, CSE, CT University, Ludhiana, Punjab,142024

kamal17203@ctuniversity.in

ARTICLE INFO

Received: 25 Dec 2024

Revised: 15 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Mental Stress is the most prominent and pertinent issue nowadays in people all across the globe and can be considered a disease that needs to be addressed on time to avoid suicides. Artificial Intelligence plays a major role in resolving this societal issue by predicting the stress level. In this paper, a hybrid model using various Machine Learning algorithms is constructed to predict the stress based on some crucial parameters. The constructed hybrid model is hyper-tuned with Adaboost and Bagging techniques and the model's accuracy was raised to 96.4%. The stacked model accurately predicts the stress level and can provide concrete insights into the reasons that are highly responsible for stress among people of all ages.

Keywords: Mental Stress Detection, Supervised Learning,

Introduction

Mental stress has become a pervasive and critical challenge in contemporary society, significantly impacting individual well-being, productivity, and overall quality of life (Barua *et al.*, 2024). As technological advancements continue to transform healthcare and psychological research, machine learning techniques offer unprecedented opportunities to understand, predict, and potentially mitigate mental stress more effectively than traditional diagnostic approaches (Barua *et al.*, 2024). The rapid proliferation of digital technologies, wearable devices, and continuous monitoring systems has generated vast amounts of physiological, behavioral, and psychological data. These rich datasets provide researchers with unique insights into the complex mechanisms underlying mental stress. The Predictive models which usually classify the stress levels based on multiple parameters are constructed by utilizing the various Supervised learning algorithms. The Psychological assessments that are usually considered in the predictive models are patterns of sleeping, heart rate, cortisol levels, etc., Conversely, unsupervised learning techniques can help uncover hidden patterns and clusters within stress-related data, potentially revealing novel insights into stress manifestation and progression that might not be apparent through traditional analytical methods. This research aims to bridge the gap between advanced computational techniques and mental health assessment by demonstrating the potential of machine learning in stress prediction (Gupta *et al.*, 2023). By integrating diverse data sources and employing sophisticated algorithmic approaches, the study seeks to develop a comprehensive framework for early stress detection, personalized intervention strategies, and enhanced mental health monitoring. The significance of this research extends beyond academic exploration, offering potential transformative implications for healthcare professionals, mental health practitioners, employers, and

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

individuals seeking proactive stress management solutions. Machine learning techniques represent a promising frontier in understanding and addressing the complex phenomenon of mental stress.

Literature Review

(Tuarob *et al.*, 2017), Mental health challenges like stress and anxiety can cause significant personal suffering and potentially lead to severe consequences. Accurately predicting an individual's mental state at present and in the future could be invaluable for healthcare professionals. Traditional mental health assessments rely on psychological experts conducting evaluations, but these methods have significant limitations. Current psychological assessment techniques are resource-intensive and time-consuming, which restricts their widespread application. Moreover, individuals may lack self-awareness about their mental state or feel uncomfortable during evaluations, potentially resulting in prolonged undetected mental health issues. This research aims to explore how advanced machine learning techniques can develop mathematical models capable of predicting an individual's current and future mental states.

(Verma and Singh, 2023) Stress has emerged as a critical mental health challenge with far-reaching health implications, potentially triggering numerous medical conditions including heart disease, migraines, infertility, obesity, and insomnia. Early stress detection is therefore paramount for preventive healthcare. Traditional stress monitoring methods using complex electrode and wire-based sensor systems were often intrusive and uncomfortable for participants. In contrast, modern smartwatches offer a more convenient approach, integrating seamlessly into daily life while providing continuous physiological data tracking through built-in sensors that monitor heart rate and activity levels.

(Ilias, Mouzakitis and Askounis, 2024), Social media platforms have become powerful tools for understanding mental health, as users frequently express their inner thoughts and emotions through online posts. Researchers have been exploring these platforms as potential avenues for early detection of mental health conditions like stress and depression. Previous research approaches have primarily focused on two methodological strategies: traditional feature extraction techniques combined with shallow machine learning classifiers. While transformer models have demonstrated notable performance improvements, they often struggle to capture comprehensive contextual knowledge. To address model reliability, the researchers implemented label smoothing techniques to improve both model performance and prediction confidence. By testing their approach across three publicly available datasets, they demonstrated significant performance enhancements when linguistic features were systematically integrated into transformer models.

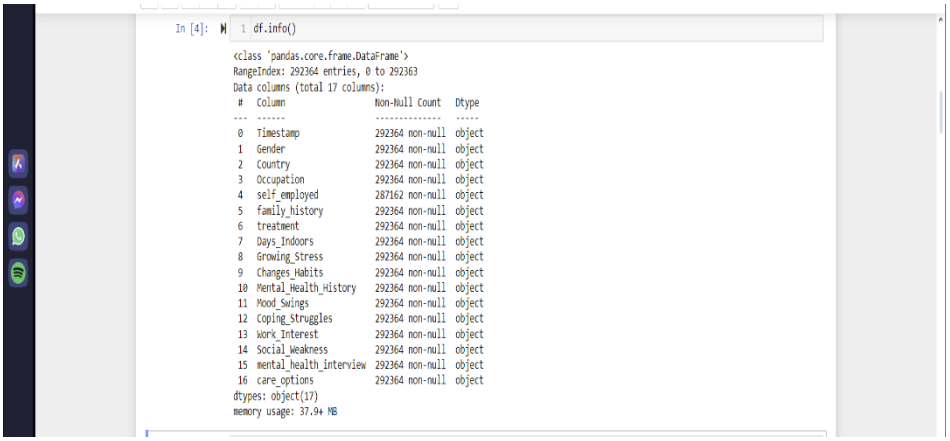
1. Research Methodology

The study addresses a very prominent and sensitive societal issue, i.e., mental health prediction. Mental Health is the root cause of various allied diseases and problems like depression, and anxiety. If they are not treated and cured on time, they can become fatal and lead to suicidal tendencies which is a great threat not only to youngsters but to the entire society. In this research, the notion is to detect the mental health of a person beforehand using supervised learning techniques so that they can be prevented and cured on time. So, the dataset of 4,923 K instances is taken and has undergone various phases like pre-processing, feature extraction, and classification. The prediction of mental health conducted earlier through AI techniques in various state-of-the-art models has resulted in a very low accuracy and all the parent models are inefficient in providing a strong insight to the datasets. The second, major drawback was that all the state-of-art models are using lower-dimensional data sets with few instances. The proposed model works on both aspects where the higher dimensional datasets are used and the higher level of accuracy is also achieved.

Dataset Input and Processing: A secondary data set with 4923K instances has been taken from Kaggle with various attributes. The Snapshot of the dataset is as shown in fig below.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

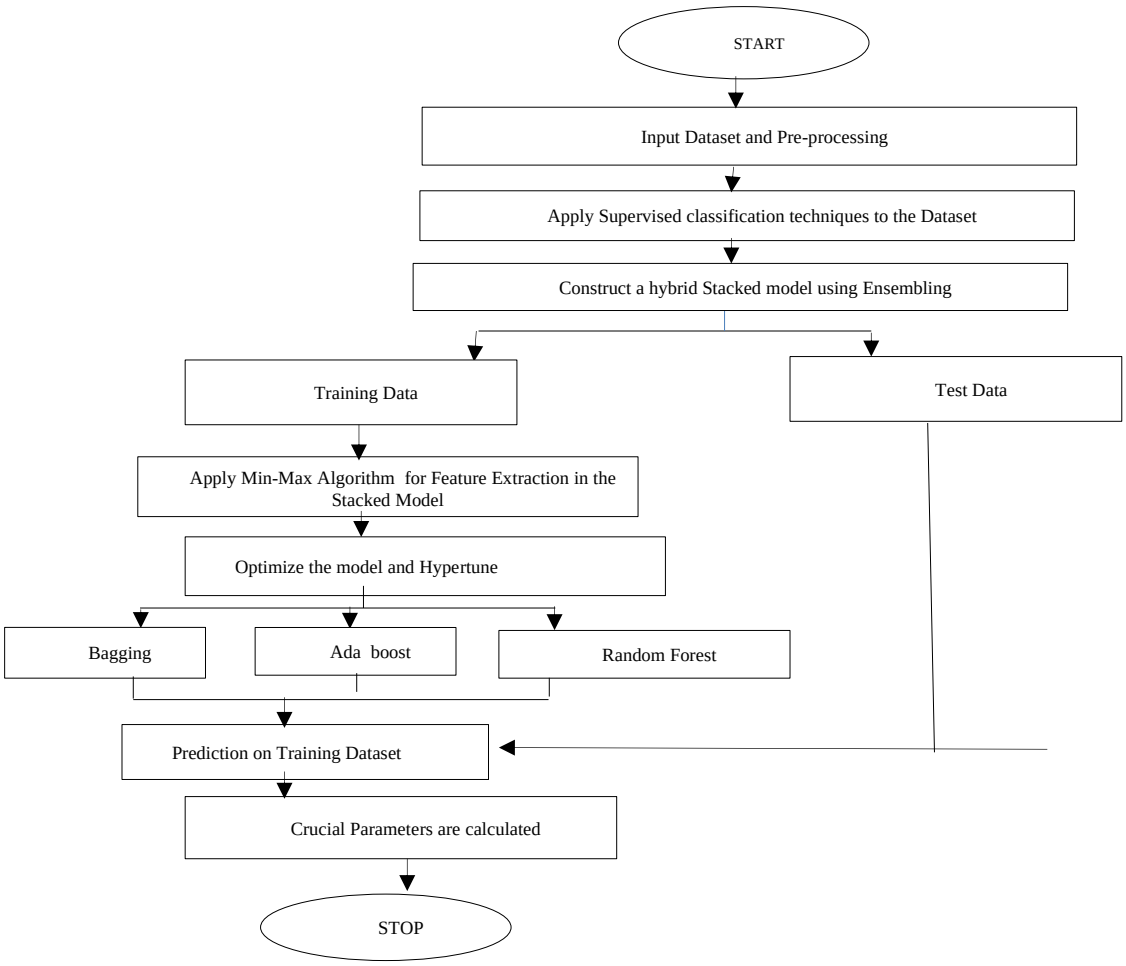


```
In [4]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 292364 entries, 0 to 292363
Data columns (total 17 columns):
 #   Column                  Non-Null Count   Dtype
---  ------                  --------------   -----
 0   Timestamp               292364 non-null  object
 1   Gender                  292364 non-null  object
 2   Country                 292364 non-null  object
 3   Occupation              292364 non-null  object
 4   self_employed           287162 non-null  object
 5   family_history          292364 non-null  object
 6   treatment               292364 non-null  object
 7   Days_Indoors            292364 non-null  object
 8   Growing_Stress          292364 non-null  object
 9   Changes_Habits          292364 non-null  object
 10  Mental_Health_History   292364 non-null  object
 11  Mood_Swings             292364 non-null  object
 12  Coping_Struggles        292364 non-null  object
 13  Work_Interest           292364 non-null  object
 14  Social_Weakness         292364 non-null  object
 15  mental_health_interview 292364 non-null  object
 16  care_options            292364 non-null  object
dtypes: object(17)
memory usage: 37.9+ MB
```

Fig 1. Description of Dataset

3.1.Proposed Algorithm



Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

3.1.1. Description of Algorithm

i. Data Preprocessing and Feature Extraction

First, the mental health dataset was collected from Kaggle, and data preprocessing was conducted, where all the duplicate and null entries were removed using the appropriate imputation techniques. Then, outliers were identified and treated using Statistical methods, including Interquartile range and Z-Score methods. All the crucial features were identified and extracted through Autoencoding and Min-Max Scaling to ensure consistent scale across different variables and mitigate potential bias in the classification model.

ii. Classification Models

The Supervised Machine Learning Models were applied, The Dataset was trained on Logistic Regression, K-Nearest Neighbor, Support Vector Machine, Decision Tree and Finally Random Forest was used and the individual accuracies of all these models were calculated, after ensembling of these models, a stacked model is created and its accuracy is compared with the other models. The stacked model was tested against the various statistical parameters like Accuracy, R-Square, F1-Value, and Recall through the Confusion matrix The stacked model was way better than the individual models and it has a high accuracy of 97.76 percent which is quite better than the already existing state of the art models.

iii. Splitting of Dataset

The Dataset of mental health is divided into two categories i.e., Training and Testing datasets where 70% of the dataset contributes to the training one, and 30% of the dataset is considered for testing purposes to achieve accurate prediction and results.

iv. Calculation of Various Parameters against Stacked Model

The different parameters like Accuracy, F1 Call, Recall, R-squared value, and Precision-score, are calculated for the various models like Logistic Regression, KNN, Decision, Random Forest, and Bagging. Optimization of all the models is conducted with Ada boosting and Bagging and then a hybrid stacked model is constructed and compared.

4. Result and Discussion

4.1. Performance Analysis Parameters

The following are the various performance analysis parameters: -

Accuracy: Accuracy is one of the prominent metrics that is utilized to calculate the performance of the model and evidence domain recovery. This metric depicts the ratio of the data points that are successfully classified to the total no. of data points and is usually represented by the equation:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision: This metric is also used for evaluating the performance of the training data set and it is indicated as the ratio of correctly indicated positive data points to the total no. of indicated positive data points

$$Precision = \frac{TP}{TP + FP}$$

Recall: Recall refers to the ratio of the corrected instances retrieved to the total no. of retrieved instances and it can be formalized as below:

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$Recall = \frac{TP}{TP + FN}$$

F1 Score: The F1 Score serves as a fundamental metric for evaluating classification model performance, offering a well-rounded assessment of how well a model makes predictions. By combining precision and recall into a single value, creates a balanced measure that helps data scientists understand their model's overall effectiveness. The mathematical formula, 2 * (Precision * Recall) / (Precision + Recall), takes the harmonic mean of these two crucial metrics.

2 * (Precision * Recall) / (Precision + Recall)

In this section, the results of the proposed model are presented along with a comparison with the existing schemes for various parameters.

Table 1 Comparison of various ML Techniques against the performance Metrics

	Training Accuracy	**Testing Accuracy**	**Precision**	**Recall**	**F1 Score**
Logistic Regression	75.97	76.24	81.88	84.28	83.06
K-NN	69.45	69.12	69.20	99.68	81.69
Decision Tree	76.02	76.05	81.19	85.05	83.07
Bagging	76.41	73.04	78.03	84.89	81.31
Random Forest	76.02	75.96	81.55	84.28	82.89
Gradient Boosting	76.03	76.02	81.15	85.07	83.06
Stacked Model (Proposed)	93.24	96.02	92.86	92.07	93.08

Table 1 presents the performance metrics for several machine learning models on a certain task or dataset. The metrics included are precision, F1 score, recall, Training accuracy, and Testing Accuracy.

Looking at the specific model performance, Logistic Regression has the highest training accuracy at 75.97%, but a lower testing accuracy at 76.24%. K-NN has fairly consistent training and testing accuracy around 69%. Decision Tree has a training and testing accuracy of around 76%. Bagging, Random Forest, and Gradient Boosting have training accuracy of around 76% and testing accuracy of around 75-76%. The "Stacked Model (Proposed)" has the highest testing accuracy at 96.02%.

As depicted in Table 1, the dataset was trained on various classification models individually, Logistic Regression was used in Binary Classification. In K-NN 9 neighbors were taken to enhance the accuracy but the training and testing accuracy reached up to 69% approximately.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

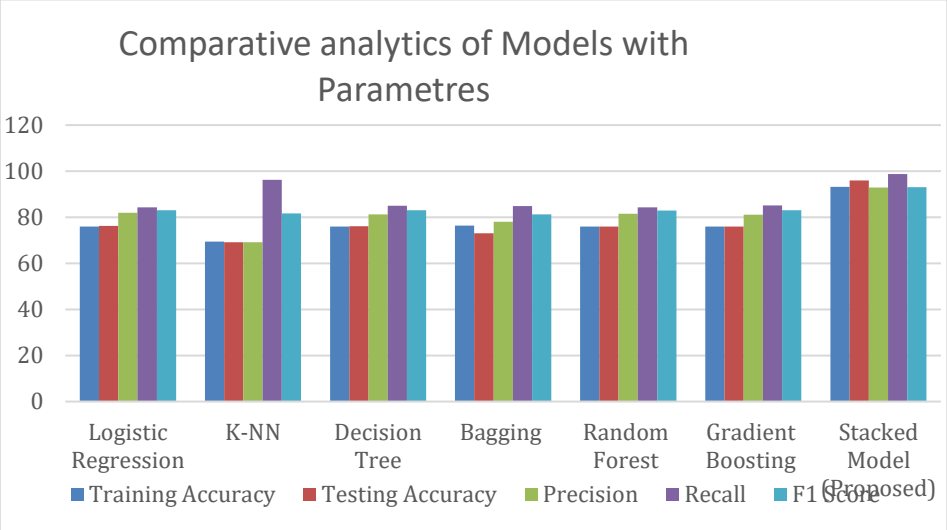


Figure 2. Comparative Analytics of Supervised Machine Learning with Parameters

5. Conclusion

In this paper, it is assumed that the persons who are willing to take treatment are considered as the patients suffering from depression. To make the appropriate predictions of patients before actually going on the acute terms of depression, the various Supervised, Supervised Machine Learning techniques like K-Nearest Neighbours with 9 neighborhood points, Logistic Regression, Decision Tress, Bagging, Random Forest, Gradient Boosting, and finally the stacked model. The classification done through the stacked model shows an accuracy of 93.08% which overpowers all other Supervised models when utilized individually. It also indicates that growing Stress, change in Occupation, and family history are the most prominent factors responsible for hampering the mental health of a person and finally converge to Depression.

6. References

[1] Barua, P.D. *et al.* (2024) "Artificial intelligence assisted tools for the detection of anxiety and depression leading to suicidal ideation in adolescents: a review," *Cognitive Neurodynamics*. Springer Science and Business Media B.V., pp. 1–22. Available at: https://doi.org/10.1007/s11571-022-09904-0.

[2] Delgadillo, J. and Atzil-Slonim, D. (2023) "Artificial intelligence, machine learning and mental health," in *Encyclopedia of Mental Health, Third Edition: Volume 1-3*. Elsevier, pp. V1-132-V1-142. Available at: https://doi.org/10.1016/B978-0-323-91497-0.00177-6.

[3] Gupta, D. *et al.* (2023) "Humans' Emotional and Mental Well-Being under the Influence of Artificial Intelligence," *Journal for ReAttach Therapy and Developmental Diversities*, 6(6).

[4] Ilias, L., Mouzakitis, S. and Askounis, D. (2024) "Calibration of Transformer-Based Models for Identifying Stress and Depression in Social Media," *IEEE Transactions on Computational Social Systems*, 11(2). Available at: https://doi.org/10.1109/TCSS.2023.3283009.

[5] Tuarob, S. *et al.* (2017) "How are you feeling?: A personalized methodology for predicting mental states from temporally observable physical and behavioral information," *Journal of Biomedical Informatics*, 68. Available at: https://doi.org/10.1016/j.jbi.2017.02.010.

[6] Verma, P. and Singh, R. (2023) "Mental Stress Prediction Using Wrist Wearable Through Machine Learning Approaches," in *2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology, ICSEIET 2023*. Available at: https://doi.org/10.1109/ICSEIET58677.2023.10303348.