A SYSTEMATIC REVIEW OF MENTAL HEALTH ANALYSIS USING MACHINE LEARNING MODELS

FANISH PANDEY,

AYUSH KUMAR RAI

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING

LOVELY PROFESSIONAL UNIVERSITY

Abstract :

Machine learning (ML), deep learning (DL), and transfer learning have shown significant promise in the prediction and classification of mental health disorders, with a variety of performance metrics across different algorithms. Stacking algorithms have achieved an accuracy of 81.75%, while the Support Vector Machine (SVM) has emerged as one of the most reliable models, outperforming others like Random Forest, K-Nearest Neighbours (KNN), and Boost with an accuracy of 82%. Notably, Random Forest has reached a high accuracy of 95.45% in some studies, while deep learning models have achieved accuracy rates of up to 95% for conditions such as depression, anxiety, and suicide risk. Logis c Regression and SVM have consistently shown strong performance across metrics like accuracy, precision, and recall. Furthermore, advanced deep learning models like BERT combined with neural networks have reached nearly 99% accuracy, underscoring the potential of machine learning to provide highly reliable mental health predictions. The studies also highlight the importance of precision in the evaluation of metrics, including precision, recall, and F1-score, across complex and imbalanced datasets. Moreover, the need for more inclusive, co-produced research and be er represented in mental health studies has been emphasized to bridge existing gaps in the field and improve model robustness.

Keywords :

Machine Learning (ML), Deep Learning (DL), Mental Health Disorders, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), XGBoost, BERT (Neural Networks), Logis c Regression, AUC-ROC

Introduction :

Mental health issues among students have become a critical and rapidly growing concern across education ins towards worldwide. These challenges, including conditions like anxiety, depression, and loneliness, have been shown to significantly affect students' academic performance, social interactions, and overall well-being. Despite their prevalence, many of these mental health conditions remain undiagnosed or untreated, due to factors like s games, lack of awareness, or the absence of proper diagnosis tools. The failure to identify these conditions early can result in long-term consequences, including a decrease in academic achievement, social isolation, and in extreme cases, self-harm or suicidal tendencies. The need for early intervention and better support systems is therefore more important than ever to ensure that students can navigate their academic journey and life with better mental resilience and well-being. The **S.M.A.R.T - Student Mental Awareness and Resilience Tracker** was developed to address these challenges by leveraging the power of machine learning (ML) and data-driven techniques to predict and monitor students' mental health based on a variety of behavioral, emotional, and psychological factors. By incorporating multiple parameters such as depression (α), social anxiety (β), and loneliness (θ), the system is designed to identify students who may be at risk of developing mental health issues. Through its predictive capabilities, the S.M.A.R.T system aims to classify students as "addicted" or "non-addicted" based on these critical psychological indicators. The classification is pivotal in distinguishing students who are struggling with mental health issues and may need targeted support, counseling, or intervention from those who are mentally resilient. To ensure the effectiveness and reliability of the S.M.A.R.T system, various performance metrics are integrated into its structure to evaluate the predictive power and accuracy of the models being used. These metrics, such as accuracy, precision, recall, F1-score, and AUC-ROC (Area Under the Receiver Opera ng Characteristic Curve), are vital in assessing how well the system can correctly classify and predict mental health conditions. Accuracy

measures how well the model predicts the correct outcomes, while precision and recall assess how effectively the model identifies true positives and avoids false positives and negatives. The F1 score provides a balanced evaluation of precision and recall, which is crucial when the focus is on a high number of true positives with minimal false results. AUC-ROC, on the other hand, serves as a visual representation of a model's ability to distinguish between different classes, ensuring that the predictions made by the system are both robust and trustworthy. The research surrounding the S.M.A.R.T system compares different machine learning models to identify which algorithms provide the best performance in mental health prediction and classification. Among these models are Random Forest, Support Vector Machine (SVM), and Logis c Regression, all of which have demonstrated significant accuracy in predicting mental health outcomes. SVM has emerged as one of the most reliable algorithms in this context, known for its capability to handle both linear and nonlinear data effectively, leading to higher prediction accuracy in mental health classification. Through the integration of cu ng-edge ML models and comprehensive performance metrics, the S.M.A.R.T system represents a significant leap forward in the management of student mental health. By enabling institutions to make data-driven decisions regarding mental health, it fosters a more supportive and understanding academic environment, where mental well-being is given the same priority as academic achievement. Ultimately, the system aims to empower students, educators, and mental health professionals to work together in identifying and addressing mental health issues at an early stage, creating a healthier and more resilient student population. By contributing to a culture of mental health awareness and intervention, the S.M.A.R.T system hopes to inspire
positive change and promote students' overall well-being, helping them academically and personally.

Review of Literature:

We read different research papers and each paper was based on different parameters like age group, work pressure, marital status, studying in a different country, screen time, spending time with family, and friends, having someone with whom things can be shared, income, etc. Each algorithm performed on different parameters and in some cases one algorithm was not performing well but in other cases, the same algorithm was performing well and giving positive results accordingly.

M.Azizur Rahman et al, in analyzing the mental health of students Multilayer Perceptron, and SVM, achieve up to 95.45% accuracy for Random Forest which is good enough to analyze the mental health of different factors. The most considerable factors affecting mental health were discovered to be financial problems, academic tension, homesickness, loneliness, and culture shock.[1]

[Mohit Mittal et al] In determining Student’s Mental Health in playing Online Mobile Games Support Vector Machine (SVM) algorithm gives the best accuracy at 91.68% [2]

[Sriteja Kataru et al] In determining the Mental Health Issues in Children who are below the 5th standard the KNN gives the best accuracy of 86.78%.[3]

[Shivani Goel et al] In Predicting Stress at Workplace Decision Tree has the best accuracy of 82%.[5]

[Solomon O. Akinola et al] In analyzing mental health conditions AdaBoost has the best accuracy of 81.75% .[6]

[R Rajkumar et al] In analyzing Student Stress due to academic challenges the SVM has the highest accuracy of 90%.[11]

[Shakir Khan1 et al] Hybrid models have also been used in detecting mental health and the combination of (TF-IDF + Logistic Regression) has the highest Accuracy of 99.4%.[19]

[Sarthak Dash et al] In determining the mental health condition based on the social media content of the user the BERT algorithm achieves the highest accuracy of 91.5%, followed by CNN.[23]

[K Praveen Tumuluru et al] In determining public mental health Random Forest archives the top accuracy of 90% followed by Logistic Regression at 85%. [29]

[Ayesha Khaliq et al] In determining public mental health using sentiment analysis, AdaBoost performs with the highest accuracy which is 96%, more than other algorithms.[34]

[B Ashwath Rao et al] In determining mental health illness, the Stacking algorithm gives the best result with an accuracy of 81.75% followed by Random Forest with an accuracy of 81.22%.[35]

[Abdullah Al Foysa et al] In determining how stress affects students' mental health Random Forest gives the best result with an accuracy of 91%.

Henrik Larsson et al and number While predicting mental health in adolescence with help of different machine learning, Random Forest algorithms gave the best accuracy, which was 95%,[36]

[Na Guo] while predicting the mental health of college students using integrated machine learning algorithms, in which the Decision tree algorithm achieved the highest accuracy of 97% compared to c1 and Random forest algorithms.[38]

[Hikaru Ooba] While predicting the mental health of pregnant women using machine learning, where EPDS algorithms gave the highest accuracy, which was 97% with 84% precision.[39]

[Qingzhong Liu] In determining the depression caused by the use of social media the algorithms RoBERTa and DeBERTa both have an accuracy of 98.0%. [40]

[Farah Jemili] Neural networks gave the best accuracy, more than 97%, in the prediction of early detection of mental health in students using multiple machine learning models.[42]

Research Questions:
The research accompanying this study focuses majorly on understanding:

What are the most productive techniques and algorithms currently used to predict mental stress levels?

Table 1:

Algorithm	Accuracy (%)	Best for
[1]SVM	95.45	International students
[2]SVM	91.68	Online Mobile Games
[3]KNN	86.78	Children
[5]DT	82	Stress at Workplace
[6]Adaboost	81.75	All Age Groups
[11]SVM	90	Student Academic Challenge
[12]TF-IDF + LR	99.4	All Age Groups
[14]Random Forest	90	Public Mental Health
[23]BERT	91.5	social media content
[24]AdaBoost	96	public mental health using sentiment analysis
[26]Stacking	81.75	mental health illness
[37]RF	91	stress affecting students
[36]RF	95	adolescence
[38]DT	97	college students
[39]EPDS	97	pregnant women
[41]RoBERTa, DeBERTa	98	depression by the use of social media
[42]Neural Network	97	early detection of mental health in students

What are the measured advantages and disadvantages of implementing a predictive analysis system to assess and address mental stress levels among college students?

Table 2:

Aspect	Pros	Cons
Predictive Accuracy	High accuracy rates are achieved through algorithms such as SVM, Random Forest, and deep learning models like BERT, with accuracy rates up to 99%.	Variability in accuracy across different models, with some models like Logistic Regression showing lower performance in certain cases.
Comprehensive Evaluation Metrics	Uses precision, recall, F1-score, and AUC-ROC, providing a well-rounded evaluation and ensuring robustness in classification accuracy.	Performance metrics can vary significantly across imbalanced datasets, requiring careful handling to avoid misleading results.
Early Intervention Potential	Enables identification of students at risk for mental health issues, allowing for timely intervention and support.	Potential over-reliance on model predictions may overshadow personalized evaluations by mental health professionals.
Algorithmic Versatility	Utilizes multiple algorithms including SVM, KNN, Random Forest, and deep learning models, allowing for comparative performance analysis.	Some algorithms, such as deep learning models, may require extensive computational resources, limiting their accessibility in low-resource settings.
Inclusive Parameters	Includes a range of psychological indicators such as depression, social anxiety, and loneliness, increasing the system's relevance and adaptability.	Expanding parameters could add complexity, making the model prone to overfitting if not managed effectively.
Potential for Stigma Reduction	Encourages mental health awareness by offering an unbiased assessment tool, potentially reducing stigma in mental health diagnostics.	sensitive categories (e.g., "addicted" vs. "non-addicted"), which could lead to stigma if privacy isn't managed properly.
Data-driven Decision-Making	Supports institutions in making data-informed decisions regarding mental health interventions, fostering proactive student support strategies.	Heavy reliance on data may raise privacy and ethical concerns if student data is not handled with strict confidentiality protocols.
Scalability for Diverse Populations	Can be adapted to diverse student populations due to its algorithmic flexibility, offering insights across various demographic groups.	Generalization to broader populations may reduce accuracy if the model is not adequately trained on representative datasets.
Encouragement of Collaboration	Fosters cooperation between students, educators, and mental health professionals, promoting a more supportive environment.	Implementation requires extensive training and cooperation, which may not be feasible in institutions with limited mental health support infrastructure.
Limitations in Diagnostic Accuracy	Some models, like Logistic Regression, perform well but may lack the complex diagnostic accuracy of specialized medical assessments	While useful as a support tool, the S.M.A.R.T system may not replace the need for thorough mental health evaluations by qualified professionals.

User-Friendly Visualization (AUC-ROC)	AUC-ROC provides clear visual insights into classification accuracy, helping users understand model performance intuitively.	Without ongoing updates and the inclusion of recent research, models may become outdated, impacting the system's effectiveness over time.

What datasets are used in this project to analyse and prediction of students' mental stress levels, and how do these datasets contribute to identifying key stress factors as well as its remedies?

DATASET USED:

Data Collection and Preprocessing

The first and foremost step involved in our review paper is the collection of data. It can be any type of categorical data that has been approved by the specific institution. It can be of any type, either primary or secondary datasets but it should be related to mental analysis only or the factor which helps us to get to know about the mental health of people specifically for students.

So, for this, we have gone through n number of research papers and all the types of datasets that are approved by the specific institution.

For this, we have studied and utilized datasets that were collected or researched from 2010-2020.

The dataset, referenced on the particular dataset- sites like Kaggle. com(https://www.kaggle.com/search?q=mental+health+analysis)

Data.gov(https://catalog.data.gov/dataset?q=Mental+health+analysis)

These are the basic sources of collection of secondary data sources.

For the primary dataset, we have collected all data from the research papers.

Names of the research papers through the dataset have been chosen:

1). Early Detection and Intervention for Mental Health Issues in Children using Machine Learning models.[3]

Description: The dataset was taken from HKF's (My Healthfirst) program in Santa Clara County, and put on air about children's social emotional, and behavioural health through licensed professionals. All this Data is taken from 2600 elementary school students using a modified Kaiser Permanente Paediatric Quality of Life (POQ2) survey, to access children's health-related quality of life across 16 features, such as aggression loneliness, and academic stress.

2). Prediction of Mental Health Treatment Adherence using Machine Learning Algorithms.[4]

Description: The dataset in the study contains individual demographic, behavioral, and clinical factors to showcase the risk that mental health disorders may require medical intervention. Age, gender, education level, sleep cycle, family stress, and income level.

3). Predicting Mental health Stress of people at the Workplace using the Machine Learning method [5]

Description: Open Sourcing Mental Health is a kind of non-profit agency that is enhanced to provide information about raising awareness and contributing resources to support mental

wellness in a large open-source and tech culture. OSMI did a Mental Health Survey in 2021 and the data they collected that data used in the paper The dataset has 130+ different responses from working professionals or people. It further had 110+ components consisting of questions from both their private and professional life. Available: https://www.kaggle.com/datasets/osmihelp/osmh-2021-mental-health-in-tech-survey-results

4). Machine Learning method for predicting depression in College Students Based on Non-Clinical Information.[10]

Description: The dataset used in this research was obtained from a different research project and the project was "The Situation and Educational Productivity System of Undergraduate Level" in ***Mae Fah Luang University*** in 2017 which proved to be very helpful for us to get to know more about youngsters.

5). Machine Learning in Safety and Health Research: A Scient metric Analysis[16]

Description: Web of Science (WoS) and Scopus Databases: The collected data is publications from January 1, 1996, to December 31, 2021. WoS developed more publications than Scopus after duplicate removal, with the final dataset comprising 1148 unique documents: 741 from WoS and 407 from Scopus.

S.No.	Dataset	Category
1.	Mental Health Diagnosis and Treatment Monitoring [14] (Dataset Link)	Age, Gender, Sleep Hours, Income, Therapy type.
2.	Personality Classification Type Dataset Dataset Link	Neutral, Aggressive, Vivid, Confused.
3.	Student Mental health Dataset link	Student's CGPA, Placement
4.	Student Stress Factors: A Comprehensive Analysis Dataset Link	Sleep Quality, Breathing Problems, Blood Pressure, Noise Level.
5.	Mental Health Dataset Dataset Link	Habits, Treatment, family support, Indoors, employed(finance)

So, after a careful review of all datasets, the dataset which is utilized for this project is a unique collaboration of multiple datasets created with a deep understanding of psychological techniques. By taking datasets from various sources on student mental health analysis with the custom approach of GENAI(GPT) a new elaborative dataset was developed.

The dataset includes major contributors to student mental health and how are they affecting or ruining it are academic stress, financial stress, family support, and various other parameters like CGPA, Parental income, placement status, sleep hours, and student behavior pattern.

Each category was taken into account to reflect the challenges faced by youngsters and also offers actionable suggestions to support mental well-being.

The target of this approach is to provide a powerful base for predictive analysis. ***The link for the dataset***

Student Mental Health Analysis

Conclusion.

This review looked at more than 60 research that used multi-modal biometric signals and machine learning (ML) to diagnose mental health conditions. Machine learning techniques have improved diagnostic accuracy Compared to conventional methods, indicating considerable promise for advancing mental health care. The model's performance has been further enhanced by incorporating multi-modal data, including behavioral and physiological information. There are still issues, though, such as requiring bigger, labeled datasets and dealing with moral dilemmas like algorithmic prejudice and data privacy.

Future Scope

Future work should concentrate on creating attention-based models that better capture the intricate relationships in multi-modal data; incorporating explainable AI (XAI) techniques will be critical to establishing clinicians' confidence in ML models; standardizing datasets and evaluation metrics will streamline model comparison and enhance reproducibility; investigating optimized feature extraction techniques will assist in identifying the most pertinent signals for various mental health conditions; and, lastly, cooperation between data scientists and mental health professionals is critical to improving these models and guaranteeing their practical applicability.

REFERENCES

1. Muhammad Azizur Rahman, Tripti Kohli. Mental health analysis of international Students using machine learning techniques
2. Shaifali Chauhan, Mohit Mittal, Hardeep Singh . Predictive Analysis on Student's Mental Health Towards Online Mobile Games Using Machine Learning
3. Sriteja Kataru, Kathleen King, Lachin Fernando. Machine Learning-Based Early Detection and Intervention for Mental Health Issues in Children
4. Madhavan Bajaj, Priyanshu Rawat. Prediction of Mental Health Treatment Adherence using Machine Learning Algorithms
5. Anika Kapoor, Shivani Goel.Predicting Stress at the Workplace using Machine Learning Techniques
6. Mental Health Predictive Analysis using Multiple Machine Learning Techniques
7. Md Zia Uddin • Kim Kristoffer Dysthe • Asbjørn Følstad • Petter Bae Brandtzaeg.Deep learning for prediction of depressive symptoms in a large textual dataset
8. Puneeth N. Thotad; Shanta Kallur; Laxmi Mundaragi; Swati H. Kadam. Mental Health Tracker Using Machine Learning Approaches
9. Elizabeth Oluyemisi Ogunseye. Predictive Analysis of Mental Health Conditions Using AdaBoost Algorithm
10. Pattaramon Vuttipittayamongkol, Pornrat Nison. A Machine Learning Approach for Depression Screening in College Students Based on Non-Clinical Information
11. Madhurana Sen, R Rajkumar. Student Stress Patterns: A Data-Driven Exploration Using Machine Learning Amidst Academic Challenges and Health Impacts.
12. Shumaila Aleem ,Noor ul Huda. Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions
13. H.arrif Hussain, Hizbullah Hassan shahen Akhtar. Mental health monitoring with multimodal sensing and machine learning: SCIENCE DIRECT
14. Chloe R. Green Nicholas Hill Kate Johnston-AtaataRenata. A critical review of research into mental health consumers' perspectives on their physical health: Is there an absence of consumers in the design, conduct, analysis, and reporting of this research?: Frontiers
15. Md. Monirul Islam a f, Shahriar Hassan b,Sharmin Akter c, Ferdaus Anam Jibon d, Md. Sahidullah e. A comprehensive review of predictive analytics models for mental illness using machine learning algorithms

16. Davi Sofyan Khairul Hafezad Abdullah . Machine Learning in Safety and Health Research: A Scientometric Analysis

17. Anoushka Thakkar1, Ankita Gupta1, and Avinash De Sousa2. Artificial Intelligence in Positive mental health: A Narrative review

18. Shaurya Bhatnagar a, Jyoti Agarwal b, Ojasvi Rajeev Sharma a b. Detection and classification of anxiety in university students through the application of machine learning

19. Shakir Khan1 · Salihah Alqahtani1. Hybrid machine learning models to detect signs of depression

20. , Haziq Ahmed, Ahmad Saad, hamza Asaad, Muhammad Sulaiman, mubashar Mushtaq. Improving Mental Health with Machine Learning: Classification of Mental Disorders Using Local Data

21. Ahmad Radwan, Mohannad Amarneh, Hussam Alawneh, Huthaifa I. Ashqar.Predictive Analytics in Mental Health Leveraging LLM Embeddings and Machine Learning Models for Social Media Analysis

22. Kannan Nova. Machine Learning Approaches for Automated Mental Disorder Classification Based on Social Media Textual Data

23. Vaishali M Deshmukh, Dr. Rajalakshmi B, Sarthak Dash. Analysis and Characterization of Mental Health Conditions Based on User Content on Social Media

24. Mrs.G.Parimala,Dr.R.Kayalvizhi,Dr.S. Nithiya . Mental Health: Detection & Diagnosis

25. Jiahui Gao, Guopu Shi. Mental_Health_Evaluation_of_College_Students_Based_on_Similar_Trajectory_Clustering_ Algorithm

26. Saloni Jage, Shubham Chaudhari, Manthan Jatte. Predicting Mental Health Illness Using Machine Learning

27. Machine Learning Techniques for Prediction of Mental Health

28. Ranjana R,T Subha, Sreenidhi K L, Akshaya K B. Mental Health Prediction using Natural Language Processing

29. CH.M.H.Saibaba, Alekhya K V K, Yeshwanth, KPraveen Tumuluru. Prediction of Public Mental Health by Using Machine Learning Algorithms

30. ROHIZAH ABD RAHMAN 1 , ,KHAIRUDDIN OMAR 1 ,, SHAHRUL AZMAN MOHD NOAH1 , ,(Member, IEEE), MOHD SHAHRUL ,NIZAM MOHD DANURI 2 , MOHAMMED ALI AL-GARADI3. Application of Machine Learning Methods in Mental Health Detection: A Systematic Review

31. Ngumimi Karen Iyortsuun 1 ,Soo-Hyung Kim 1, Min Jhon 2, Hyung-Jeong Yang 1 and

Sudarshan Pant 1. A Review of Machine Learning and Deep Learning Approaches on Mental Health Diagnosis

32. Gyeongcheol Cho1 , Jinyeong Yim2 ,, Younyoung Choi3, Jungmin Ko4 , Seoung-Hwan Lee.Review of Machine Learning Algorithms for Diagnosing Mental Illness

33. Yutao Sun,Hui Li. Machine Learning-based Mental Health Analysis and Early Warning for College Student

34. Saad Awadh Alanazi, Ayesha Khaliq, Fahad Ahmad, Nasser Alshammari, Iftikhar Hussain, Muhammad Azam Zia, Madallah Alruwaili, Alanazi Rayan, Ahmed Alsayat and Salman Afsar. Public's Mental Health Monitoring via Sentimental Analysis of Financial Text Using Machine Learning Techniques

35. Konda Vaishnavi, U Nikhitha Kamath, B Ashwath Rao, and N V Subba Reddy. Predicting Mental Health Illness Using Machine Learning Algorithms

36. Ashley E. Tate ,Ryan C. McCabe,Henrik Larsson,Sebastian Lundström,Paul Lichtenstein,Ralf Kuja-Halkola. Predicting mental health problems in adolescence using machine learning techniques

37. Biodoumoye Bokolo. Comprehensive analysis of stress factors affecting students: a machine learning approach

38. Na Guo. Mental Health State Prediction Method of College Students Based on IntegratedAlgorithm

39. Hikaru Ooba Jota Maki, Hisashi Masuyama. Voice Analysis and Deep Learning for Detecting Mental Disorders in Pregnant Women

40. Qingzhong Liu. Deep Learning-Based Depression Detection from Social Media: Comparative Evaluation of ML and Transformer Techniques

41. Maleeha Illahi, Isma Farah Siddiqui, Qasim Ali, Fizza Abbas Alvi. Ensemble Machine Learning Approach for Stress Detection in Social MediaTexts

42. Farah Jemili. Early Detection of Mental Health Issues through Machine Learning: A Comparative Analysis of Predictive Models

43. Pradeep Kumar Tiwari , Muskan Sharma , Payal Garg , Tarun Jain , Vivek Kumar Verma and Afzal Hussain. Sentiment Analysis of Mental Illness Using Machine Learning Techniques

44. S. A. Edalatpanah,Ramin Safa*,Ali Sorourkhah. Predicting mental health using social media: A roadmap for future development.