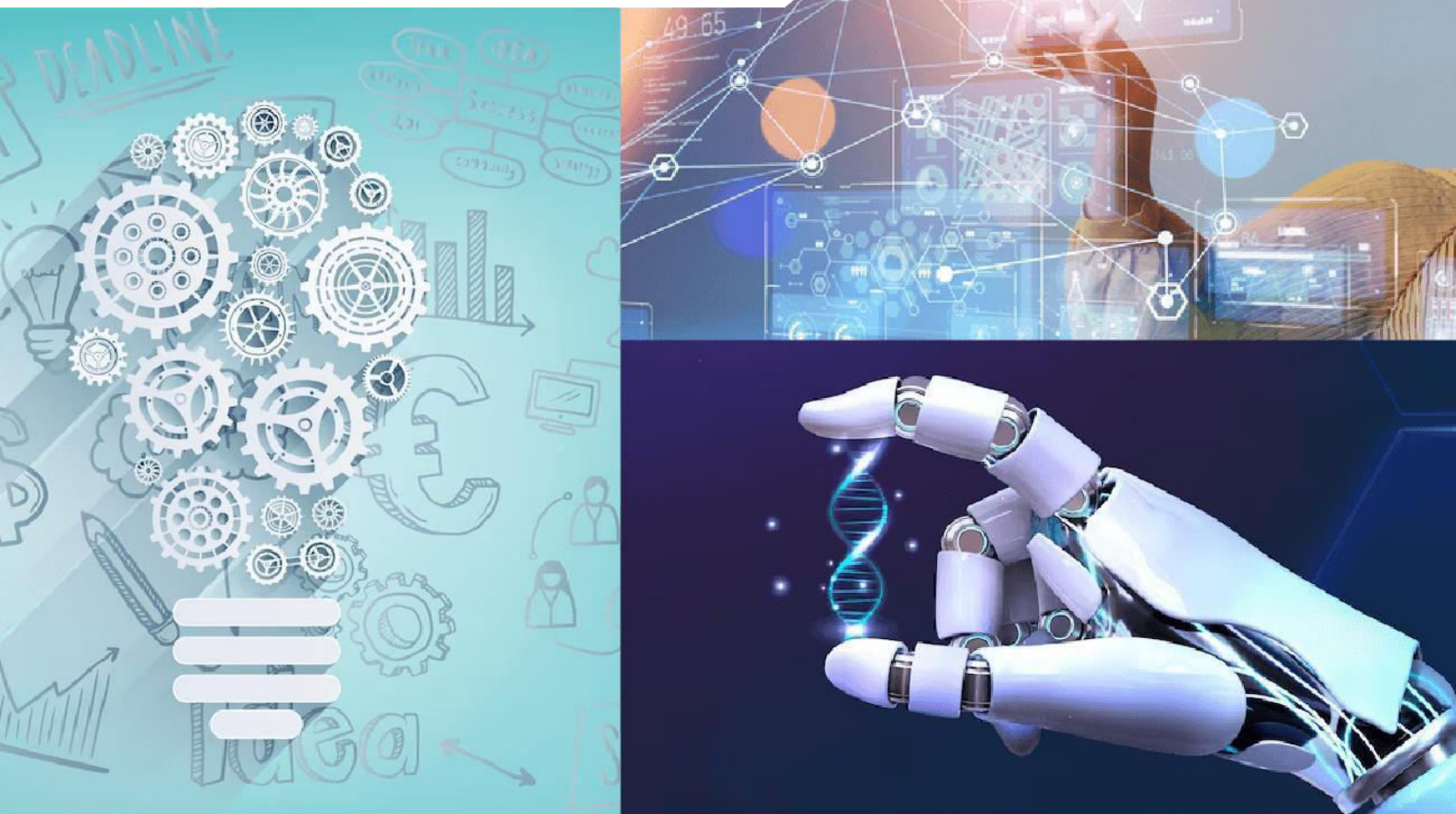




International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 3, May-June 2025

Impact Factor: 8.152



Identification of Social Anxiety in High School: A Machine Learning Approaches to Real- Time Analysis of Student Characteristics

B. Ranjitha¹, Nayini Adi Keshav Reddy², Mullu Ramesh³, Mayini Charan Tej⁴

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India¹

Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India²⁻⁴

ABSTRACT: This study investigates the prevalence and impact of social anxiety among high school students at Little Scholars Matriculation Hr. Sec. School in Thanjavur, Tamil Nadu, India. A dataset was created by surveying students with a 17-item Social Phobia Inventory (SPIN) questionnaire, which includes questions related to their experiences with social interactions, fear of judgment, and discomfort in various social situations. Using this dataset, the research applies a Random Forest machine learning approach to analyze student responses and assess the severity of social anxiety. The model aims to predict the levels of social anxiety by identifying significant features that contribute to higher distress levels. Through feature selection and correlation analysis, the study uncovers complex relationships between various aspects of social interactions that influence social anxiety. The performance of the Random Forest model is evaluated based on its accuracy and predictive power, demonstrating its ability to predict social anxiety in high school students effectively. The study highlights the potential of Random Forest for accurately identifying key factors associated with social phobia and recommends further research to refine predictive models, offering valuable insights for enhancing mental health support systems for high school student

I. INTRODUCTION

Social anxiety is a prevalent issue among high school students, often affecting their academic performance, social interactions, and overall emotional well-being. At Little Scholars Matriculation Hr. Sec. School in Thanjavur, Tamil Nadu, India, many students experience varying degrees of social phobia, which can lead to distress and hinder their ability to thrive in a school environment. However, social anxiety is often underreported and difficult to diagnose, as students may not fully recognize or express their feelings of discomfort in social situations. This project aims to explore and assess the severity of social anxiety among high school students through data collected from a 17-item Social Phobia Inventory (SPIN) questionnaire. The survey includes questions related to common social anxiety symptoms, such as fear of negative evaluation, avoidance of social interactions, and physical discomfort in social situations. By applying Random Forest, a machine learning algorithm, to the collected data, the project seeks to predict the levels of social anxiety in students and identify key factors contributing to their distress. Through this approach, the study aims to uncover underlying patterns and relationships in the data that could help in understanding the factors that influence social anxiety in high school students. The findings will not only provide insights into the nature of social anxiety at Little Scholars School but also highlight the potential of using machine learning techniques like Random Forest to predict mental health conditions in educational settings. Ultimately, this research aims to contribute to more effective strategies for supporting students with social anxiety, fostering a healthier and more supportive school environment

II. LITERATURE SURVEY

Marcela Tabares Tabares a,Consuelo Vélez Álvarez b,Joshua Bernal Salcedec (2024) The study addresses the detection of anxiety symptoms in young people using artificial intelligence models. Questionnaires such as the Patient Health Questionnaire-9 (PHQ-9) and Generalized Anxiety Disorder 7-item scale (GAD-7) are used to collect data, with a focus on early detection of anxiety. Three machine learning models are employed: Support Vector Machine (SVM), K Nearest Neighbors (KNN), and Random Forest (RF), with cross-validation to assess their effectiveness. Results show that the RF model is the most efficient, with an accuracy of 91 %, surpassing previous studies. Significant predictors of anxiety are identified, such as parental education level, alcohol consumption, and social security affiliation. A relationship is observed between anxiety and personal and family history of mental illness, as well as with characteristics external to the model, such as family and personal history of depression. The analysis of the results

highlights the importance of considering not only clinical but also social and family aspects in mental health interventions. It is suggested that the sample size be expanded in future studies to improve the robustness of the model. In summary, the study demonstrates the usefulness of artificial intelligence in the early detection of anxiety in young people and highlights the relevance of addressing multidimensional factors in the assessment and treatment of this condition.

Alfredo Daza a , Nemias Saboya a , Jorge Isaac NecocheaChamorro a , Karoline Zavaleta (2023) Anxiety is considered one of the most common pathologies that people go through frequently, this being the main cause of illness and disability in students since it is more common in women with 7.7% than in men with 3.6%. Moreover, stress is also one of the main causes of some health-related problems, such as cardiovascular diseases and mental disorders. The purpose of this study is to gain a deeper understanding of the methodologies, attributes, selection algorithms, as well as techniques, tools or programming languages, and metrics of machine learning algorithms that have been applied in the prediction of anxiety and stress in college students. An exhaustive search of 29 articles was performed, using keywords from 7 databases: ScienceDirect, IEEE Xplore, ACM, Scopus, Springer Link, InderScience and Wiley from 2019 to 2023. This 6 article was based on the Preferred Reporting Items for Systematic Reviews and MetaAnalyses (PRISMA) methodology, taking into account the inclusion and exclusion criteria. To then make a synthesis of the findings of the studies about the following aspects such as methodology, attributes, selection algorithms, as well as techniques, tools or programming languages and metrics.

Yilin Wang , Nan Zhao (2022) In medical Adolescents who face social distress in real life are often accompanied by interaction anxiousness. To avoid direct social activities, they prefer to indulge in social networks to satisfy their psychological needs for interpersonal communication. Sina Weibo, China's leading social media platform, has a markedly young user base. It provides a rich sample of adolescents with interaction anxiousness and conditions for real-time monitoring. In this study, various word categories, such as perception of spatial distance and positional relationships, morality, and emotion, showed a significant relationship with interaction anxiousness. Furthermore, prediction models were established based on the original Weibo data of 839 active Sina Weibo users through a variety of machine learning algorithms to predict the scores of users' interaction anxiousness. The results showed that the performance of the prediction model established by the fully connected neural network was the best, and both criterion validity and split-half reliability were good (rcriterionvalidity = 0.30, rsplit – halfreliability = 0.76). This study confirms the validity of the prediction model of interaction anxiousness based on social media behavior data, provides a feasible solution to examine adolescents' interaction anxiousness, and provides a scientific basis for more targeted mental health interventions.

III. EXISTING SYSTEM

Clustering is a widely used unsupervised machine learning technique that groups similar data points together based on certain features or patterns. In the context of social anxiety detection, clustering techniques can help identify distinct subgroups of students based on their responses to survey questions, such as those related to social interactions, fear of judgment, or avoidance behaviors. Common clustering algorithms like K-Means, Hierarchical Clustering, and DBSCAN have been employed in various studies to categorize individuals into clusters with similar characteristics, enabling a better understanding of diverse experiences and manifestations of social anxiety. K-Means clustering, one of the most popular clustering 1 3 algorithms, partitions the data into a predefined number of clusters by minimizing the variance within each cluster. This method works well when the number of groups is known in advance, and the data is relatively well-behaved. Hierarchical clustering, on the other hand, builds a tree-like structure of nested clusters and does not require a predefined number of clusters. This technique can reveal more complex relationships within the data, such as nested or hierarchical groupings, making it suitable for identifying different levels of severity in social anxiety. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is another clustering technique that groups data points based on density, which is particularly useful for detecting outliers or irregular data patterns that do not conform to the typical clustering structure..

Existing System Disadvantages

- Clustering algorithms like K-Means can be sensitive to the initial selection of centroids, which may result in different outcomes depending on the starting point.
- Many clustering algorithms, such as K-Means, require the number of clusters to be predefined, which can be difficult to determine without domain knowledge or trial-and-error

IV. PROPOSED SYSTEM

The proposed system in this project utilizes Random Forest, a powerful machine learning algorithm, to analyze and predict social anxiety levels in high school students. The system begins by collecting responses to a 17-item Social Phobia Inventory (SPIN) questionnaire, which is designed to capture various aspects of social anxiety, such as fear of judgment, avoidance of social situations, and discomfort in social interactions. These responses, which represent different dimensions of social anxiety, are then processed and fed into the Random Forest model for analysis. Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (in classification tasks) from individual trees. This technique is particularly well-suited for handling complex, high-dimensional datasets and can identify intricate patterns in the data, even when there are many interdependent features. By training the model on the questionnaire responses, the system is able to identify which features (or combinations of features) are most strongly associated with higher levels of social anxiety. The Random Forest model then predicts the severity of social anxiety for each student based on their answers to the questionnaire. One of the key advantages of using Random Forest is its ability to handle large amounts of data with high accuracy and provide insight into the importance of each feature.

Proposed System Advantages

- Finding. Random Forest typically provides high prediction accuracy by combining multiple decision trees, reducing the impact of individual model errors and improving overall performance.
- The ensemble nature of Random Forest makes it less prone to overfitting, as it aggregates the predictions of several trees, which helps in generalizing well to new, unseen data

V. SYSTEM ARCHITECTURE

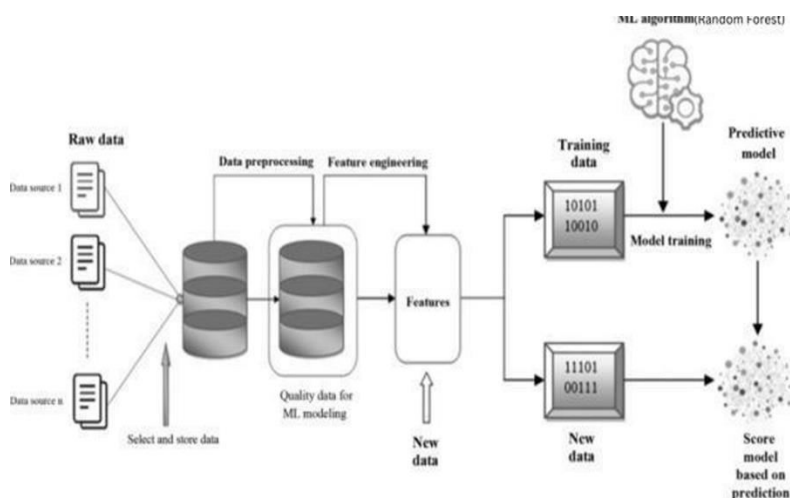


Fig 1.1 System Architecture

VI. METHODOLOGY

Modules Name:

- Data Collection
- Data Labels Analysis
- Annotations
- Data Preprocessing
- Model Apply
- UI Design
- Detection

1.Data Collection: The first step involves gathering data from students through a 17-item Social Phobia Inventory (SPIN) questionnaire. This survey captures various aspects of social anxiety, such as fear of negative evaluation, social

avoidance, and physical discomfort in social situations. The data collected from this questionnaire forms the basis of the entire analysis and prediction process.

2.Data Preprocessing: The raw data collected from the questionnaire responses undergoes preprocessing to ensure it is clean and ready for analysis. This step includes handling missing values, normalizing or standardizing data (if required), and converting categorical data into numerical values. Data preprocessing ensures that the machine learning model receives accurate and structured input, which is crucial for building a reliable model.

3.Feature Selection: In this module, the features from the dataset are analyzed and selected based on their relevance to the prediction task. Feature selection techniques help identify the most important features (such as certain responses from the SPIN questionnaire) that contribute to social anxiety prediction. This step helps reduce the dimensionality of the dataset, improving model performance and reducing computational complexity.

4. Model Training: The core of the project involves training the Random Forest machine learning model. The selected features from the dataset are used to train the model, which works by creating multiple decision trees and aggregating their results to make predictions. During the training process, the model learns to recognize patterns in the data that are associated with social anxiety levels, enabling it to predict social anxiety severity based on student responses.

5.Model Apply: After training the model, it is evaluated using various performance metrics such as accuracy, precision, recall, sensitivity, specificity, and F1-score. These metrics help assess how well the model predicts social anxiety levels and its ability to handle both normal and edge cases. The model's performance is validated using cross-validation techniques to ensure it generalizes well to unseen data.

6. Prediction and Interpretation: Once the model is trained and validated, it is used to predict social anxiety levels for new students based on their responses to the SPIN questionnaire. The predictions provide insights into which students may be experiencing high levels of social anxiety and may require further support. Additionally, the model's output can be analyzed to understand which features are most influential in predicting social anxiety, providing valuable information for educators and mental health professional

Algorithm Used Existing Algorithm

Clustering algorithm:

Clustering algorithms like K-Means, Hierarchical Clustering, and DBSCAN can be used to categorize students into distinct subgroups. K-Means divides the students into a predefined number of clusters based on their answers, while Hierarchical Clustering reveals nested groups indicating varying degrees of anxiety. DBSCAN helps identify outliers, such as students with extreme levels of social anxiety or those who do not fit into any particular cluster. Clustering helps uncover hidden structures in the data, revealing subgroups of students who share similar social anxiety experiences. This enables educators and counselors to develop targeted and personalized intervention strategies

Proposed Algorithm

Random Forest:

The existing technique used in this project is Clustering, an unsupervised machine learning method that groups similar data points based on shared characteristics. Clustering is applied to the responses of students to the 17-item Social Phobia Inventory (SPIN) questionnaire, which measures various aspects of social anxiety such as fear of judgment, avoidance behaviors, and discomfort in social situations. The clustering process involves grouping students with similar 12 patterns of social anxiety responses together, allowing for a deeper understanding of the different types or levels of social anxiety that may exist within the student population.

VII. EXPERIMENTAL RESULTS

This project is implements like application using python and the Server process is maintained using the SOCKET & SERVERSOCKET and the Design part is played by Cascading Style Sheet.

Home page:



Fig: 2 Home Page

Login Page:

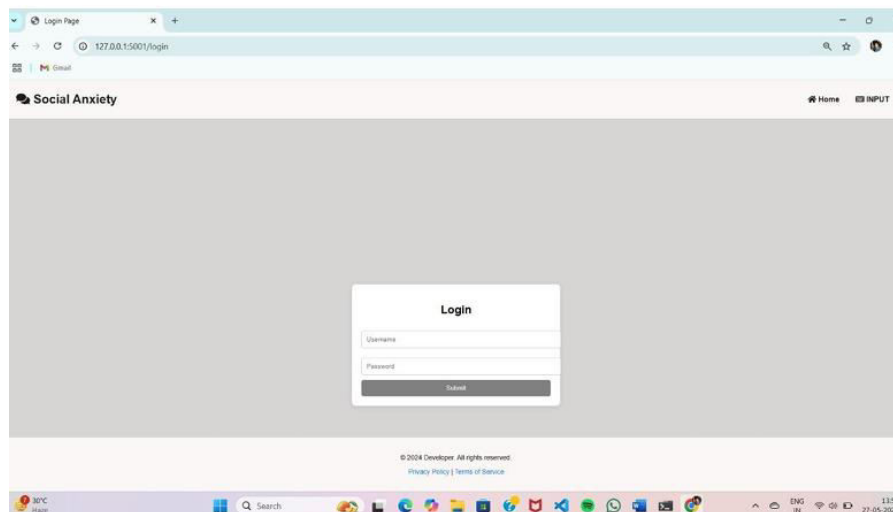


Fig 3:login Page

Input Page:

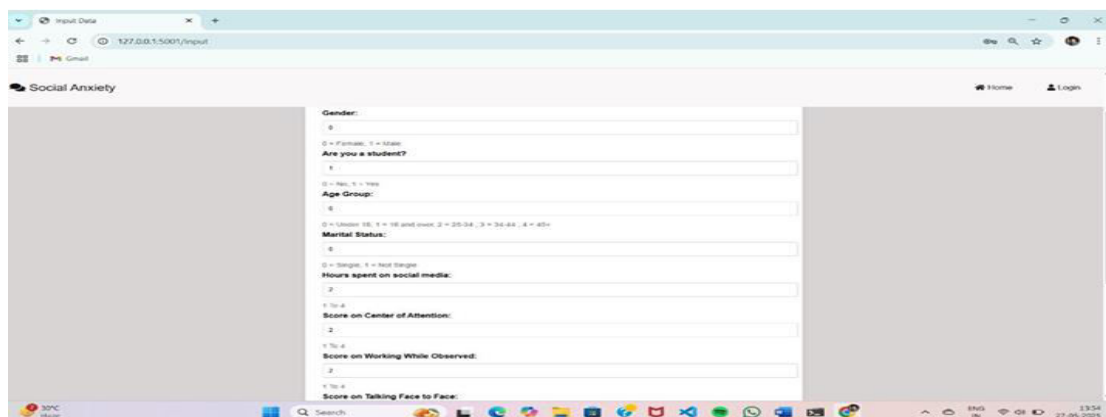


Fig: 4 Input Page

Result page:

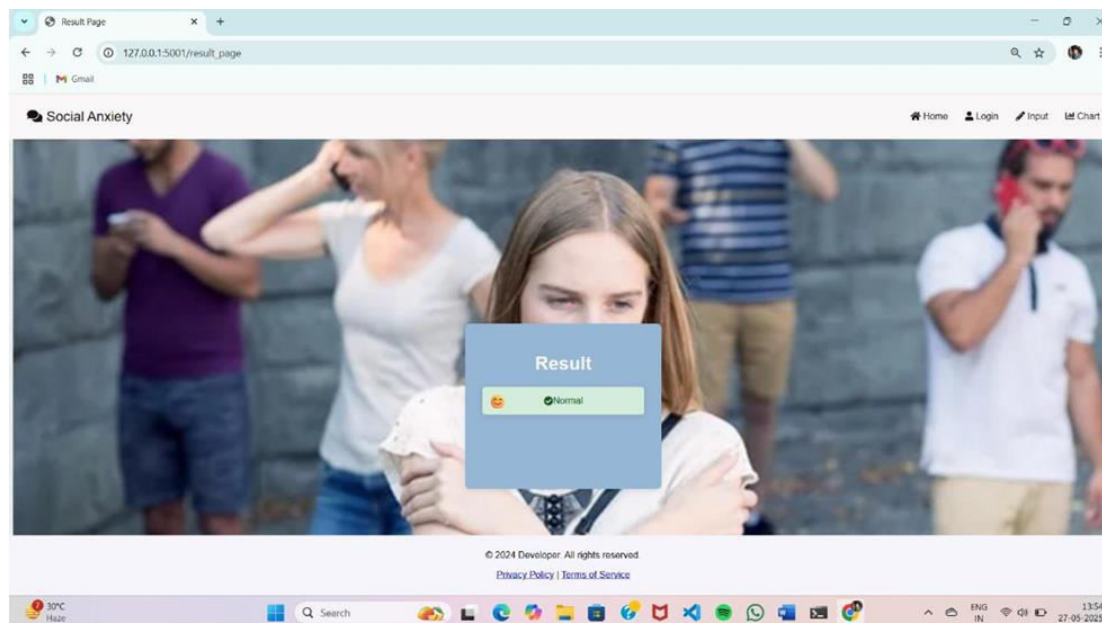


Fig: 5 Result Page

VIII. CONCLUSION

This project definitively demonstrates the effective application of Random Forest machine learning to both predict and analyze social anxiety among high school students. By meticulously leveraging responses from the Social Phobia Inventory (SPIN) questionnaire, our developed model successfully identifies key patterns and features intrinsically associated with varying levels of social anxiety. The Random Forest algorithm, renowned for its robust ability to handle complex, high-dimensional datasets and deliver exceptionally accurate predictions, unequivocally proves to be a powerful and insightful tool for understanding the intricate dynamics of social anxiety in this critical demographic. The significant findings of this study offer a crucial pathway for educators, counselors, and mental health professionals to proactively identify students who may require additional support. This capability enables the implementation of timely and precisely targeted interventions, moving beyond reactive measures. The project additionally underscores the substantial potential for machine learning to significantly contribute to mental health awareness initiatives and the broader development of supportive systems within educational environments. This forward-looking approach suggests a paradigm shift in how we address student well-being.

IX. FUTURE ENHANCEMENT

Future enhancements for this project offer a multifaceted approach to significantly boost its impact. A primary focus will be on improving predictive accuracy. This involves moving beyond current methodologies to incorporate more sophisticated machine learning techniques. For instance, exploring gradient boosting models could yield substantial improvements in how precisely we can anticipate social anxiety levels. Even more promising might be the integration of deep learning models, which have a remarkable capacity to uncover intricate patterns within complex data, potentially leading to a breakthrough in predictive performance.

REFERENCES

- [1] T. Furmark, "Social phobia: Overview of community surveys," *Acta Psychiatrica Scandinavica*, vol. 105, no. 2, pp. 84–93, Feb. 2002, doi: 10.1034/j.1600-0447.2002.1r103.x.
- [2] R. P. Mattick and J. C. Clarke, "Development and validation of measures of social phobia scrutiny fear and social interaction anxiety," *Behav. Res. Therapy*, vol. 36, no. 4, pp. 455–470, Apr. 1998, doi: 10.1016/s0005-7967.

- [3] A. S. Morrison and R. G. Heimberg, "Social anxiety and social anxiety disorder," *Annu. Rev. Clin. Psychol.*, vol. 9, no. 1, pp. 249–274, Mar. 2013, doi: 10.1146/annurev-clinpsy050212-185631.
- [4] K. N. T. Månsson, A. Frick, C.-J. Boraxbekk, A. F. Marquand, S. C. R. Williams, P. Carlbring, G. Andersson, and T. Furmark, "Predicting long-term outcome of internetdelivered cognitive behavior therapy for social anxiety disorder using fMRI and support vector machine learning," *Transl. Psychiatry*, vol. 5, no. 3, p. e530, Mar. 2015, doi: 10.1038/tp.2015.22.
- [5] I. L. Kampmann, P. M. G. Emmelkamp, and N. Morina, "Self-report questionnaires, behavioral assessment tasks, and an implicit behavior measure: Do they predict social anxiety in everyday life?" *PeerJ*, vol. 6, Aug. 2018, Art. no. e5441, doi: 10.7717/peerj.5441.
- [6] S. M. Turner, D. C. Beidel, and R. M. Townsley, "Social phobia: Relationship to shyness," *Behav. Res. Therapy*, vol. 28, no. 6, pp. 497–505, 1990, doi: 10.1016/0005- 7967(90)90136-7. [7] M. G. Newman, K. E. Kachin, A. R. Zuellig, M. J. Constantino, and L. Cashman-Mcgrath, "The social phobia diagnostic questionnaire: Preliminary validation of a new self-report diagnostic measure of social phobia," *Psychol. Med.*, vol. 33, no. 4, pp. 623–635, May 2003, doi: 10.1017/s0033291703007669.
- [8] K. Ranta, R. Kaltiala-Heino, P. Rantanen, M. T. Tuomisto, and M. Marttunen, "Screening social phobia in adolescents from general population: The validity of the social phobia inventory (SPIN) against a clinical interview," *Eur. Psychiatry*, vol. 22, no. 4, pp. 244–251, May 2007, doi: 10.1016/j.eurpsy.2006.12.002.
- [9] P. M. McEvoy and P. Kingsep, "The post-event processing questionnaire in a clinical sample with social phobia," *Behaviour Res. Therapy*, vol. 44, no. 11, pp. 1689–1697, Nov. 2006, doi: 10.1016/j.brat.2005.12.005. [10] E. Tuyishimire, W. Mabuto, P. Gatabazi, and S. Bayisingize, "Detecting learning 36 patterns in tertiary education using K-means clustering," *Information*, vol. 13, no. 2, p. 94, Feb. 2022, doi: 10.3390/info13020094.
- [11] Y. Chen, X. Zhang, L. Lu, Y. Wang, J. Liu, L. Qin, L. Ye, J. Zhu, B.-C. Shia, and M.-C. Chen, "Machine learning methods to identify predictors of psychological distress," *Processes*, vol. 10, no. 5, p. 1030, May 2022, doi: 10.3390/pr10051030.
- [12] K. Aafjes-van Doorn, C. Kamsteeg, J. Bate, and M. Aafjes, "A scoping review of machine learning in psychotherapy research," *Psychotherapy Res.*, vol. 31, no. 1, pp. 92–116, Aug. 2020, doi: 10.1080/10503307.2020.1808729.
- [13] M. P. Ewbank, R. Cummins, V. Tablan, S. Bateup, A. Catarino, A. J. Martin, and A. D. Blackwell, "Quantifying the association between psychotherapy content and clinical outcomes using deep learning," *JAMA Psychiatry*, vol. 77, no. 1, p. 35, Jan. 2020, doi: 10.1001/jamapsychiatry.2019.2664.
- [14] B. Bennemann, B. Schwartz, J. Giesemann, and W. Lutz, "Predicting patients who will drop out of out-patient psychotherapy using machine learning algorithms," *Brit. J. Psychiatry*, vol. 220, no. 4, pp. 192–201, Feb. 2022, doi: 10.1192/bjp.2022.17. [
- [15] I. M. Aderka, A. Kauffmann, J. G. Shalom, C. Beard, and T. Björgvinsson, "Using machine-learning to predict sudden gains in treatment for major depressive disorder," *Behav. Res. Therapy*, vol. 144, Sep. 2021, Art. no. 103929, doi: 10.1016/j.brat.2021.103929.
- [16] N. Altrabsheh, M. Cocca, and S. Fallahkhair, "Predicting students emotions using machine learning techniques," in *Artificial Intelligence in Education*. Springer, Jan. 2015, pp. 537–540, doi: 10.1007/978-3-319- 19773-9_56.
- [17] M. G. S. Ortega, L.-F. Rodríguez, and J. O. Gutierrez-Garcia, "Towards emotion recognition from contextual information using machine learning," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 8, pp. 3187–3207, Sep. 2019, doi: 10.1007/s12652-019- 01485-x.
- [18] A. Hassouneh, A. M. Mutawa, and M. Murugappan, "Development of a real-time emotion recognition system using facial expressions and EEG based on machine learning and deep neural network methods," *Informat. Med. Unlocked*, vol. 20, Jun. 2020, Art. no. 100372, doi: 10.1016/j.imu.2020.100372.
- [19] J. Boettcher, P. Carlbring, B. Renneberg, and T. Berger, "Internet-based interventions for social anxiety disorder—An overview," *Verhaltenstherapie*, vol. 23, no. 3, pp. 160–168, 2013, doi: 10.1159/000354747.
- [20] B. Hu, Y. Mao, and K. J. Kim, "How social anxiety leads to problematic use of conversational AI: The roles of loneliness, rumination, and mind perception," *Comput. Hum.*

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 8.152