International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

REAL-TIME ANALYSIS OF STUDENT BEHAVIOR FOR IDENTIFYING SOCIAL ANXIETY IN HIGH SCHOOL USING MACHINE LEARNING

Ruqiya Fatima[1], Dr. I. Samuel Peter James[2]
[1]PG Scholar, Department of CSE, Shadan Women's College of Engineering and Technology, Hyderabad, ruqiyafatima16@gmail.com
[2]Associate Professor, Department of AI&DS, Shadan Women's College of Engineering and Technology, i.samuelpeterjames@gmail.com

Received: 24-07-2025 Accepted: 28-08-2025 Published: 05-09-2025

ABSTRACT: In this study, high school students at Little Scholars Matriculation Hr. Sec. School in Thanjavur, Tamil Nadu, India, are asked about the prevalence and effects of social anxiety. The 11-item Social Phobia Inventory (SPIN) questionnaire, which asks about social interactions, fear of being judged, and discomfort in different social settings, was used to collect data from students. The study analyses student replies and determines the degree of social anxiety using this dataset and a Random Forest machine learning technique. Through the identification of important characteristics that lead to higher degrees of discomfort, the model seeks to predict social anxiety levels. By use of feature selection and correlation analysis, the research reveals intricate connections among many facets of social interactions that impact anxiety. Accuracy and predictive power are used to assess the Random Forest model's performance, showing that it can accurately predict high school students' social anxiety. In order to improve mental health support systems for high school kids, the study suggests more research to improve predictive models and emphasises the potential of Random Forest for precisely pinpointing important elements linked to social anxiety.

1. INTRODUCTION

Teenagers who suffer from social phobia, also known as social anxiety disorder (SAD), face significant challenges in the classroom that affect their mental well-being. academic success and social ties [1]. A deep dread of social circumstances is a common issue that leads to avoidance behaviours that prevent positive social relationships. [2], [3] the importance of tackling. The study employs a novel strategy by combining state-of-the-art machine learning algorithms with real-time data collecting to examine social anxiety during a student's growth stage [4]. It aims to overcome biases and limitations associated with traditional self-report measures and retrospective data [5]. To properly diagnose and treat social phobia, it's imperative to differentiate between typical shyness and severe social anxiety [6]. The research explores the complex situation that the majority of students at Little Scholars Matriculation Higher Secondary School experienced social anxiety through the use of an integrated analytical approach that combines data exploration, correlation analyses, clustering techniques, and machine learning methodologies. The 11-question SPIN survey is a thorough tool for assessing students' experiences in a range of social contexts [7], [8], and [9]. Patterns in the degree of social anxiety can be found using plots that compare responses. Correlation studies highlight relationships between survey questions, contributing to our overall understanding of social anxiety. Clustering methods like k-means clustering contribute to the recognition of heterogeneity by identifying discrete subgroups within the student body [10]. Machine learning methods like decision trees and random forests [11], [12] that extract latent features offer a more thorough comprehension of the variables affecting distress levels [13], [14], and [15]. By exposing unique social phobia characteristics at Little Scholars Matriculation Higher Secondary School (LSMS), the study advances a thorough examination of the diverse experiences of students and sheds light on social anxiety. It emphasises the complex relationships between factors that impact the severity of social anxiety, which is essential for creating a supportive and safe learning environment.

OBJECTIVE

Through an analysis of responses to a 11-item Social Phobia Inventory (SPIN) questionnaire, the project aims to determine the prevalence and severity of social anxiety among high school students at Little Scholars Matriculation Hr. Sec. School in Thanjavur, Tamil Nadu, India. Using Random Forest, a machine learning algorithm, the research seeks to forecast students' distress levels based on their answers and pinpoint important elements causing social anxiety. Furthermore, the project aims to identify the connections among several social and emotional elements that impact social anxiety, including social avoidance, discomfort in social situations, and dread of receiving a poor grade. By employing data-driven techniques, the research also seeks to show how well Random Forest predicts social anxiety levels, offering a thorough grasp of the underlying causes. In the end, the study hopes to provide insightful information that will help teachers, therapists, and mental health specialists better assist kids who struggle with social anxiety and aid in the creation of more specialised interventions and support networks in educational settings.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

PROBLEM STATEMENT

The stigma around mental health conversations and the limits of conventional self-reporting techniques mean that social anxiety, a prevalent mental health issue among high school students, frequently remains undiagnosed. Students' general well-being, peer relationships, and academic achievement may all suffer as a result of this untreated worry. A lack of sophisticated data-driven methods for the real-time detection and analysis of social phobia tendencies in educational settings persists despite the availability of diagnostic instruments such as the Social Phobia Inventory (SPIN). An objective, scalable, and trustworthy approach to evaluating and forecasting pupils' levels of social anxiety is desperately needed. By using machine learning techniques—more especially, the Random Forest algorithm—to evaluate SPIN questionnaire data and pinpoint the main social and emotional elements that contribute to anxiety, this study tackles this problem. In order to improve early identification, enable focused interventions, and eventually help high school students' mental health in an educational setting, the study intends to reveal hidden patterns and correlations.

Existing System

A popular unsupervised machine learning method called clustering puts related data points in one category according to specific characteristics or patterns. Clustering algorithms can be used to detect social anxiety by identifying discrete student groupings based on how they answer survey questions on social interactions, fear of being judged, or avoidance behaviours. In numerous research, popular clustering algorithms such as K-Means, Hierarchical Clustering, and DBSCAN have been used to group people into clusters based on shared traits, improving our comprehension of the wide range of social anxiety symptoms and experiences. The K-Means clustering algorithm, which is widely used, divides the data into a predetermined number of clusters by minimising the variation within each cluster.

Disadvantage of Existing System

- Depending on the starting point, clustering algorithms such as K-Means may produce various results since they are sensitive to the initial selection of centroids. Difficult to predict
- Several clustering techniques, like K-Means, need a predetermined number of clusters, which might be challenging to figure out without domain expertise or trial-and-error.
- Problems with High-Dimensional Data: Because of the curse of dimensionality, which states that the more dimensions there are, the less significant the distance between points becomes, clustering high-dimensional data—like survey replies with lots of features—can be challenging.

PROPOSED SYSTEM

Random Forest, a potent machine learning algorithm, is used in this project's suggested approach to assess and forecast high school pupils' degrees of social anxiety. A 11-item Social Phobia Inventory (SPIN) questionnaire is used to gather data at the start of the system. This questionnaire is intended to capture a variety of social anxiety symptoms, including discomfort in social circumstances, fear of being judged, and avoidance of social situations. These answers are then processed and entered into the Random Forest model for analysis, representing various aspects of social anxiety. Random Forest is an ensemble learning technique that builds several decision trees during training and uses each tree to determine the class mode (in classification problems).

This method can detect complicated patterns in the data, even when there are many interdependent features, and is especially well-suited for processing high-dimensional, complex datasets. The system may determine which features—or combinations of features—are most closely linked to higher levels of social anxiety by training the model on the response to the questionnaire. The Random Forest model then uses each student's questionnaire responses to estimate how severe their social anxiety will be. One of the main benefits of employing Random Forest is its capacity to process vast volumes of data accurately and offer insight into the relative value of each attribute.

Advantages of Proposed System

- Because Random Forest combines several decision trees, it can reduce the impact of individual model faults and improve overall performance, resulting in high prediction accuracy.
- Random Forest is less likely to overfit because of its ensemble nature, which combines the predictions of multiple trees and aids in good generalisation to new, unseen data.
- Random Forest can manage missing data by employing surrogate splits in decision trees, which guarantees that the model's performance is not greatly impacted by incomplete data.

2. RELATED WORKS

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

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

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. **[1]**

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 article was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (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. **[2]**

Food. The early diagnosis and classification of social anxiety disorder (SAD) are crucial clinical support tasks for medical practitioners in designing patient treatment programs to better supervise the progression and development of SAD. This paper proposes an effective method to classify the severity of SAD into different grading (severe, moderate, mild, and control) by using the patterns of brain information flow with their corresponding graphical networks. We quantified the directed information flow using partial directed coherence (PDC) and the topological networks by graph theory measures at four frequency bands (delta, theta, alpha, and beta). The PDC assesses the causal interactions between neuronal units of the brain network. Besides, the graph theory of the complex network identifies the topological structure of the network. Resting-state electroencephalogram (EEG) data were recorded for 66 patients with different severities of SAD (22 severe, 22 moderate, and 22 mild) and 22 demographically matched healthy controls (HC). **[3]**

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 ($r_{criterion\,validity} = 0.30$, $r_{split-half\,reliability} = 0.76$). 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. **[4]**

Anxiety disorder (AD) is a major mental health illness. However, due to the many symptoms and confounding factors associated with AD, it is difficult to diagnose, and patients remain untreated for a long time. Therefore, researchers have become increasingly interested in non-invasive bio-signals, such as electro encephalography (EEG), electrocardiogram (ECG), electrodermal response (EDA), and respiration (RSP). Applying machine learning to these signals enables clinicians to recognize patterns of anxiety and differentiate a sick patient from a healthy one. Further, models with multiple and diverse bio-signals have been developed to improve accuracy and convenience. This paper reviews and summarizes studies published from 2012 to 2022 that applied different machine learning algorithms with various bio-

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

signals. In doing so, it offers perspectives on the strengths and weaknesses of current developments to guide future advancements in anxiety detection. Specifically, this literature review reveals promising measurement accuracies ranging from 55% to 98% for studies with sample sizes of 10 to 102 participants. On average, studies using only EEG seemed to obtain the best performance, but the most accurate results were obtained with EDA, RSP, and heart rate. Random forest and support vector machines were found to be widely used machine learning methods, and they lead to good results as long as feature selection has been performed. Neural networks are also extensively used and provide good accuracy, with the benefit that no feature selection is needed. This review also comments on the effective combinations of modalities and the success of different models for detecting anxiety**.[5]**

Social phobia, also known as social anxiety disorder, is a prevalent and debilitating condition characterized by an intense fear of social situations where one might be scrutinized by others. It affects a significant portion of the population, with lifetime prevalence rates estimated between 7% and 13% in Western countries. The disorder typically begins in adolescence and can lead to chronic impairment in social and vocational functioning if left untreated. Despite its prevalence, social phobia is often misunderstood and trivialized, even by mental health professionals. The following sections provide a detailed overview of social phobia based on recent research findings. **[6]**

The development and validation of measures for social phobia, particularly the Social Interaction Anxiety Scale (SIAS) and the Social Phobia Scale (SPS), have significantly advanced the understanding and assessment of social anxiety disorders. These scales are essential tools for identifying individuals with social phobia and tailoring treatment approaches. **[7]**

Social anxiety disorder (SAD), also known as social phobia, is characterized by an intense fear of negative evaluation in social situations, leading to significant impairment in daily functioning. The understanding and classification of SAD have evolved, particularly with the DSM-5, which broadened diagnostic criteria to include fears of offending others and removed the generalized subtype(Rose & Tadi, 2021). Recent research emphasizes cognitive-behavioral models, highlighting factors such as attention bias, emotion regulation, and self-focused attention that contribute to the maintenance of social anxiety(Morrison & Heimberg, 2013). Effective treatments include cognitive-behavioral therapy (CBT), social skills training, and acceptance and commitment therapy, which have shown empirical support(Wong et al., 2012)(Wong et al., 2012). **[8]**

The study by Månsson et al. (2015) investigates the long-term outcomes of internet-delivered cognitive behavior therapy (iCBT) for social anxiety disorder (SAD) using fMRI and machine learning techniques. The research highlights the potential of neural predictors in forecasting treatment responses, emphasizing the role of specific brain regions in determining long-term efficacy**.[9]**

The study by Kampmann et al. (2018) investigates the predictive validity of self-report questionnaires, behavioral assessment tasks (BATs), and implicit measures for social anxiety in daily life. The findings suggest that self-report measures, particularly those assessing fear of negative evaluation, are more effective predictors of social anxiety than behavioral tasks or implicit measures. This conclusion is supported by the observation that maximum anxiety levels during BATs did not correlate with daily social anxiety outcomes, highlighting the limitations of these methods in capturing the nuances of social anxiety in real-world contexts. **[10]**

3. METHODOLOGY

Using machine learning techniques, specifically the Random Forest algorithm, the methodology used in this study is a systematic and structured approach intended to predict and analyse social anxiety levels among high school pupils. Several crucial steps are involved in this process to guarantee precise data analysis and trustworthy forecasts. A 11-item Social Phobia Inventory (SPIN) questionnaire is used to collect student responses in the first step, data collection. Several aspects of social anxiety are evaluated by this questionnaire, such as social avoidance, discomfort in social settings, and fear of being judged. Students from Little Scholars Matriculation Hr. Sec. School in Thanjavur, Tamil Nadu, India, provide the data, guaranteeing a representative sample of the student body.

MODULE DESCRIPTION:

Data Collection:

Using the 11-item Social Phobia Inventory (SPIN) questionnaire, data collection is the first step in the approach. The purpose of this questionnaire is to evaluate a number of social anxiety symptoms, including social avoidance, discomfort in social settings, and fear of judgement. High school pupils from Little Scholars Matriculation Hr. Sec. School in Thanjavur, Tamil Nadu, India, provide the data, guaranteeing a variety of answers to reflect the spectrum of social anxiety experiences.

Data Preprocessing:

After being gathered, the data is pre-processed to adjust for missing values, standardise the responses, and transform categorical data into

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

numerical representations. In order to guarantee that the data is clean and prepared for model training, this step is crucial.

Feature Selection:

In order to determine the most pertinent factors that influence social anxiety, feature selection is essential. This stage involves identifying the salient characteristics from the SPIN questionnaire responses that have a substantial impact on social anxiety using correlation analysis and feature importance methodologies.

Model Training:

The pre-processed and feature-selected data is used in this stage to train a Random Forest model. To make predictions that are more accurate, Random Forest, an ensemble learning system, builds several decision trees and aggregates their output.

Model Evaluation:

A number of metrics, including accuracy, precision, recall, and F1-score, are used to assess the model's performance after training. In order to evaluate the model's resilience and avoid overfitting, cross-validation is employed.

Prediction and Interpretation:

In the last stage, the trained Random Forest model is used to forecast the social anxiety level of fresh student data. The features obtained from the SPIN questionnaire are used by the model to produce predictions. Following that, the results are analysed to pinpoint the main causes of anxiety levels, providing information that can direct student support programs and intervention efforts**.**

4. ALGORITHM

Random Forest:

Random Forest, a potent machine learning algorithm, is used in this project's suggested approach to assess and forecast high school pupils' degrees of social anxiety. A 11-item Social Phobia Inventory (SPIN) questionnaire is used to gather data at the start of the system. This questionnaire is intended to capture a variety of social anxiety symptoms, including discomfort in social circumstances, fear of being judged, and avoidance of social situations. These answers are then processed and entered into the Random Forest model for analysis, representing various aspects of social anxiety. Random Forest is an ensemble learning technique that builds several decision trees during training and uses each tree to determine the class mode (in classification problems).

When dealing with high-dimensional, complicated datasets, this method works very well since it can spot complex patterns in the data, even when there are a lot of interdependent aspects. The system may determine which characteristics (or feature combinations) are most closely linked to higher degrees of social anxiety by training the model on the questionnaire responses.

Precondition: A training set S := (x1, y1), . . . ,(xn, yn), features F, and number of trees in forest B.

```
1 function RandomForest(S , F)
2 H ← ∅
3 for i ∈ 1, . . . , B do
4 S (i) ← A bootstrap sample from S
5 hi ← RandomizedTreeLearn(S (i) , F)
6 H ← H ∪ {hi}
7 end for
8 return H
9 end function
10 function RandomizedTreeLearn(S , F)
11 At each node:
12 f ← very small subset of F
13 Split on best feature in f
14 return The educated tree
15 end function
```

Based on each student's questionnaire responses, the Random Forest model then forecasts the degree of social anxiety in each of them. Using Random Forest has several benefits, one of which is its capacity to process massive volumes of data accurately and offer insight into the significance of each attribute.

5. DATA FLOW DIAGRAM

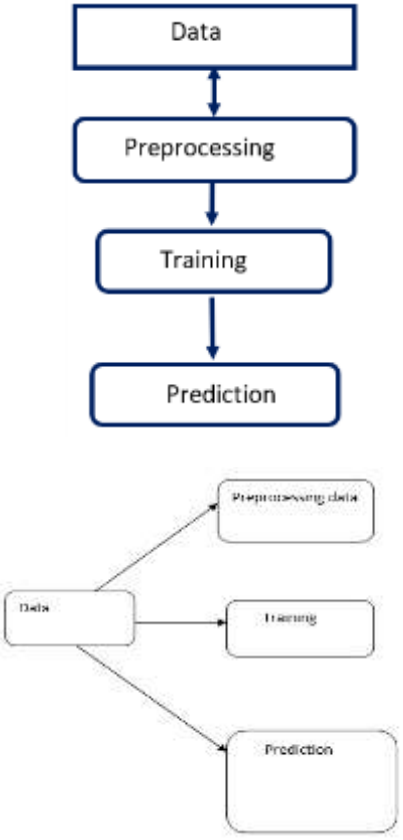

Fig 5: Flow Diagram

6. SYSTEM ARCHITECTURE

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

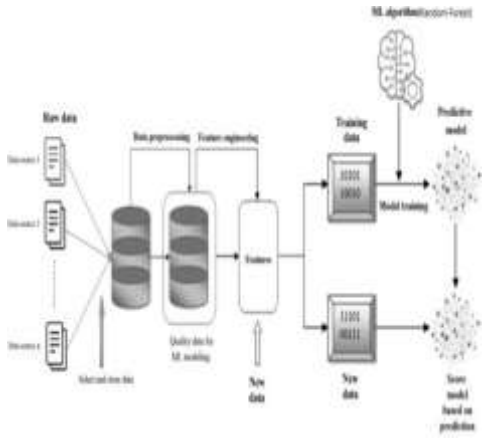

Fig 6: System Architecture Diagram

The above figure illustrates a machine learning system in which raw data is gathered, preprocessed, and key features are extracted. The handled data is active to train a Random Forest model. The trained model predicts levels of social anxiety on original data and generates scores based on the predictions.

7. RESULTS

The cluster methods divided students into various social anxiety levels, for instance, low, medium, and high. They also identified that some students possessed individual patterns that were different. When the models were compared, Random Forest provided the best results with the highest accuracy (0.92). That implies it is the best approach in detecting social anxiety and can assist in providing the appropriate support to students.

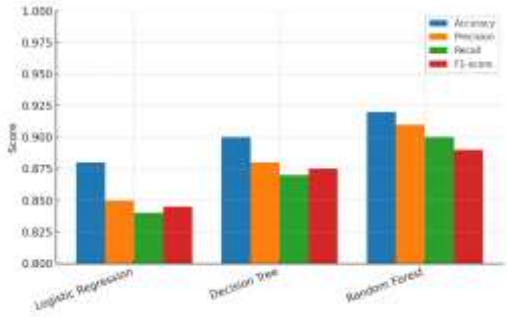

Fig 7: Accuracy

8. FUTURE ENHANCEMENT

In the future, this project might be improved by increasing the model's scope and forecast accuracy. Adding more sophisticated machine learning methods, like gradient boosting or deep learning models, could be one improvement as they may perform better at predicting social anxiety levels. To enhance comprehension of the elements driving social anxiety, the model could also be improved by adding additional characteristics, such behavioural or demographic data. Another potential improvement is using real-time data collecting via web platforms or mobile apps to continuously track students' mental health and enable prompt treatments.

To provide a more comprehensive tool for mental health support in schools, the method might also be extended to predict other mental health issues. In order to guarantee that the model is applicable to a larger population and to make it more generalisable and effective in various cultural and educational contexts, it could be enhanced using more varied datasets.

9. CONCLUSION

This project concludes by showing how Random Forest machine learning may be used effectively to predict and analyse social anxiety in high school pupils. Key patterns and characteristics linked to different degrees of social anxiety are successfully identified by the model using data from the Social Phobia Inventory (SPIN) questionnaire. A useful tool for comprehending social anxiety in children is the Random Forest algorithm, which can handle complex, high-dimensional data and produce precise predictions. In order to enable prompt and focused treatments, the results of this study can help educators, counsellors, and mental health specialists identify kids who might require extra support. Additionally, this initiative shows how machine learning may help raise awareness of mental health issues and provide support networks in school settings.

Additional features or real-time data integration are examples of future improvements that could increase the system's accuracy and usefulness and aid in the larger treatment of social anxiety.

REFERENCES

[1] “Anxiety in young people: Analysis from a machine learning model,” Marcela Tabares Tabares a,Consuelo Vélez Álvarez b,Joshua Bernal Salcedo c, Santiago Murillo Rendón d e, 2024.

[2] “Systematic review of machine learning techniques to predict anxiety and stress in college students.,”Alfredo Daza, Nemias Saboya , JorgeIsa ac NecocheaChamorro , Karoline Zavaleta Ramos [b], Yesenia del Rosario Vásquez Valencia [a],2023.

[3] Abdulhakim Al-Ezzi 1, Nidal Kamel 2,*,†, Amal A Al-Shargabi 3,*, Fares Al-Shargie 4, Alaa Al-Shargabi 5,†, Norashikin Yahya 1,*, Mohammed Isam Al-Hiyali 1, Machine learning for the detection of social anxiety disorder using effective connectivity and graph theory measures.

[4]” Prediction model of interaction anxiousness based on Weibo data.,” **:** Yilin Wang , Nan Zhao.,2022.

[5] “Lou Ancillon 1,2,†,Mohamed Elgendi” ,† andCarlo Menon 1 Machine Learning for Anxiety Detection Using Biosignals.;2022.

[6] T. Furmark, ‘‘Social phobia: Overview of community surveys,’’ Acta Psychiatrica

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

candinavica, vol. 105, no. 2, pp. 84–93, Feb. 2002, doi: 10.1034/j.1600-0447.2002.1r103 .x.
[7] 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.
[8] 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-clinpsy-050212-185631.
[9] 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 internet-delivered 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.
[10] 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.54