Journal of Engineering Science and Technology
Special Issue on ICITE20222, November (2022) 32 - 47

MEMORES: A WEB-BASED INTELLIGENT SCREENING TOOL FOR PREDICTING SOCIAL ANXIETY DISORDER USING A MACHINE LEARNING MODEL

HANS ADEY CESA[1,*], JOSHUA DANIEL LAWSIN[1],
CARLOS JULIAN CHIONGBIAN[1], MARIA ANGELICA ZARAGOZA[1],
GLENN PEPITO[1], RHODA ALVAREZ[2]

[1]Department of Computer, Information Sciences, and Mathematics, School of Arts and Sciences, University of San Carlos, Talamban Campus, Cebu City, Philippines, 6000
[2]Department of Psychology, School of Arts and Sciences, University of San Carlos, Talamban Campus, Cebu City, Philippines, 6000
*Corresponding Author: hanselcesa@gmail.com

Abstract

Social Anxiety Disorder is one of the most commonly diagnosed anxiety disorder that can disrupt a person's work and social life. People with social anxiety disorder experience the symptoms of fear and anxiety in situations where there is a possibility of being scrutinized or judged by others. Clinicians conduct clinical assessment through clinical interviews, screening and diagnostic testing when the patient seeks professional help. However, the process usually takes two to four hours to complete to avoid false positive impressions leading to misdiagnosis. Thus, this paper proposes a novel approach to streamline the screening process without compromising the accuracy rate of predicting a possible manifestation of social anxiety disorder within patients through an intelligent web-based screening tool that uses established machine learning algorithms to screen and evaluate a patient. The tool assesses the possibility of SAD based on the information gathered from the patient's demographic, physiological symptoms, affective stability, and feared situations. Four machine learning models namely: Decision Tree, Logistic Regression, Support Vector Machine, and K-Nearest Neighbours, were trained, tested, and cross-validated. Using K-Fold Cross-Validation, we evaluated four machine learning models based on accuracy, precision, recall, f1, and AUC – ROC curve. The SVM model performed the best among the other models and garnered the highest accuracy of 96.01%, with 97.13% precision, 95.33% recall, 96.13% f1 score, and an AUC score of 0.97 ± 0.05. The Support Vector Machine model is then integrated into the developed screening tool.

Keywords: Machine learning, Mental disorder, Predictive technology, Screening tool, Social anxiety disorder.

1. Introduction

Social Anxiety Disorder (SAD) is one of the most commonly diagnosed anxiety disorder that can disrupt a person's work and social life. People with social anxiety disorder experience the symptoms of fear and anxiety in situations where there is a possibility of being scrutinized or judged by others. Clinicians conduct clinical assessment through clinical interviews, screening and diagnostic testing when the patient seeks professional help. During clinical interviews, clinicians inquire about the patient's motivation for seeking medical or psychological assistance, references, background, age, frequency of symptoms, and severity of symptoms. During this phase, clinicians typically form an impression of the patient's condition. Based on the clinician's impression, patients undergo a screening procedure and are given psychometric tests for the specific disorder. The outcome of the screening procedure not only gives the clinician an insight about the patient's condition, but also enables clinicians to plan and recommend appropriate treatments. Nevertheless, issues typically occur during the course of the process. The procedure typically takes between two and four hours to complete as clinicians aim to avoid misdiagnosis.

A cross-sectional study shows that practitioners, even with doctorate degrees, often recognize these disorders at chance levels where 0.5% is the lowest chance for identifying social anxiety disorder [1]. The said study made use of the Mini International Neuropsychiatric Interview (MINI) and was administered to 840 clinical patients with medical charts containing evidence of previous diagnosis. It defined misdiagnosis as a situation where the MINI resulted to a positive case but not in the patient's medical chart. It was found that misdiagnosis rate of anxiety disorder reached up to 97.8%. The low detection rate for mental health disorders, in general, is reflected by the poor quality of care.

To tackle this problem, this study makes use of an intelligent web-based screening tool that uses established machine learning algorithms to streamline the screening process without compromising the accuracy rate of predicting a possible manifestation of social anxiety disorder within patients. Furthermore, data pre-processing procedures, as well as correct feature scaling and selection strategies, are used to increase the model's performance.

The rest of the paper is organized as follows: Section 2 presents the related literature of the study. Section 3 provides a detailed explanation of the proposed method. Section 4 presents the result and analysis in detail. Section 5 covers the conclusion and recommendations for future research.

2. Review of Related Literature

2.1. General screening tools

A screening tool is a quick questionnaire or method that assesses mental health or trauma symptoms, risk factors, or both to determine whether additional and more in-depth evaluation is required. A positive screening tool result indicates that a more thorough assessment is needed.

2.2. Related works

Liu et al. [2] investigated the use of functional connectivity to diagnose SAD. They utilized resting-state fMRI to scan twenty healthy and sick patients. They further used multivariate pattern analysis to differentiate between sick and the healthy controls. They employed Linear SVM to determine the pattern classifier, and the testing findings showed an 82.5 percent classification rate. Another study used a combination of fMRI and SVM to predict SAD among sixteen sick individuals and nineteen healthy individuals, where they obtained an AUC score of 0.89. Furthermore, with an AUC score of 0.82, they investigated a distinction between Parkinson's Disease (PD) and SAD patients [3]. In a similar study, Frick et al. [4] further investigated the possibility of distinguishing SAD patients using SVM, fMRI, and regional grey matter volume. Individuals are classified using SVM based on brain activation and structural patterns. With a balanced accuracy of 72.6%, the results demonstrated that SVM could be beneficial for identifying imaging biomarkers of SAD.

Asvestopoulou et al. [5] developed a screening tool for dyslexia using machine learning. The screening tool is called *DysLexML*. The screening tool applies various machine learning algorithms to analyse the fixation points during the silent reading of children. Eye tracking technology reads the fixation points. *DysLexML* achieved an accuracy of 97% using linear SVM. Also, the researchers analysed the impact of noise on the fixation positions and found that the screening tool is still accurate in the presence of noise. Therefore, the researchers concluded that based on the promising results, this would serve as a basis for cost-effectively developing screening tools in more extensive areas.

Chiu et al. [6] developed a screening tool for detecting moderate cognitive impairment and dementia. The machine learning tool helps neurologists and neuropsychologists screen mild cognitive impairment (MCI) and dementia. The screening model is called *NMD-12*. Overall, the model shows that it can aid healthcare professionals as a screening tool wherein it can accurately differentiate normal cognition (NC), mild cognitive impairment (MCI), and very mild dementia (VMD), and dementia.

Using the Indian Liver Patient Dataset (ILPD), Md et al. [7] suggested an ensemble-based model for diagnosing liver illness. They used six ensemble-based algorithms, such as Extra Tree Classifier, Bagging, Stacking, Random Forest, XGBoost, and Gradient boosting, and compared their performance with models used in other research works. They also performed data pre-processing procedures to improve the model's performance. To evaluate the performance of the models, they performed GridSearchCV using 10-fold cross validation. Extra Tree Classifier had the highest testing accuracy of 91.82% when compared to other trained models. This proposed model also outperformed the models of different research works.

Kumar et al. [8] proposed a novel method for predicting and curing depression based on the Global Vector and Bi-directional Long Short-Term Memory algorithms. In this study, these algorithms are integrated into smartwatches and fitness bands to collect heart rate data. This data is remembered for a long time using Bi-directional Long Short-Term Memory algorithm. The algorithm analyses the data to predict whether the person is depressed. Based on the result, the algorithm suggests curative actions for the person to do. The algorithm takes input from both directions and keeps the past and present data into memory to produce better outcome. In this study, the models Bi-LTSM with Global Vector and Bi-LTSM without Global Vector were compared, where the former model performed better with an accuracy of 86%.

3. Methods

3.1. Acquisition of research data

The study used an existing data set with 30 attributes divided into different categories, including demographic, emotional, and physical symptoms; results from the 24-item Liebowitz Social Anxiety Scale (LSAS) and 17-item Social Phobia Inventory (SPIN) tests; and classification of the specific instance - whether it has SAD or not.

The dataset was acquired from Mendeley Data, a secure cloud-based repository for domain-specific and cross-domain specific dataset, where it was published by Sina Fathi and Maryam Ahmadi on March 9, 2020.

The dataset contained initial attributes of SAD symptoms based on the Diagnostic and Statistical Manual of Mental Disorders - 5 (DSM-5) and International Classification of Disease, 10th Revision (ICD-10), Diagnostic Criteria for Research guidelines [9].

Furthermore, they employed a triangulation method to assess subjects for social anxiety disorder by administering two conventional measures, the Liebowitz Social Anxiety Scale (LSAS) and the Social Phobia Inventory (SPIN). After pre-processing the data, the final social anxiety disorder dataset was formed.

3.2. Clinical diagnostic process

A clinical diagnostic process starts with an interview conducted by a mental health professional to understand the situation of the patient. The mental health professional then performs screening assessment using a set of questionnaires to further investigate the patient's condition and symptoms. The screening includes gathering information about the patient's demographic, physiological symptoms, affective stability, feared situations; and using of clinician-rated psychometric instruments. Liebowitz Social Anxiety Scale (LSAS) and Social Phobia Inventory (SPIN) are the most commonly used psychometric measures for social anxiety disorder. The screening result will be the basis for whether a further patient diagnosis is required. Figure 1 illustrates the flow of a complete process that a patient undergoes when consulting a professional in the field.

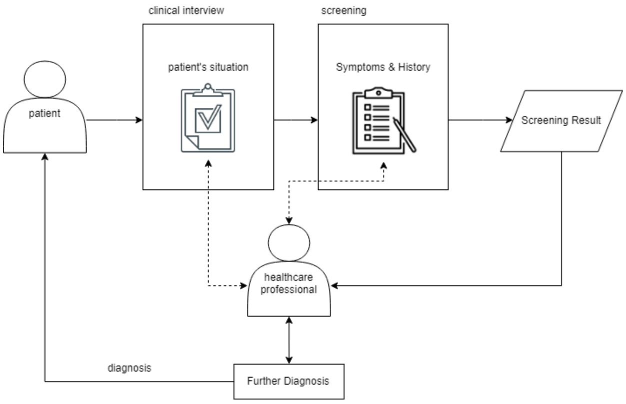

Fig. 1. Clinical diagnostic process.

3.3. Model selection process

The model selection process begins with data pre-processing where missing values are handled using an imputation technique. The rest of the dataset attributes are then standardized or normalized, depending on which machine learning algorithm the dataset is fed to. The relevant features are then selected using feature selection techniques. The dataset is split into training and testing sets *k* times using K-Fold

Cross Validation, and the machine learning model is trained each iteration. The resulting model is validated based on its accuracy, precision, recall, f1, and AUC scores in classifying whether the patient manifests social anxiety disorder. In addition, the model that produces the best performance concerning the specified performance metric is integrated into the tool.

Figure 2 illustrates the complete process of model selection.

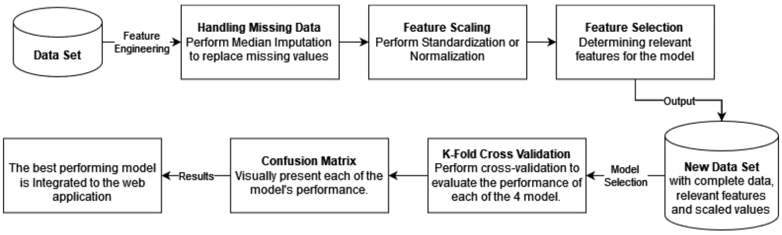

Fig. 2. Model selection process.

3.4. System process

The screening tool uses a machine learning algorithm to help predict a manifestation of social anxiety disorder within patients. The tool is meant to provide an initial diagnosis to aid mental health professionals in conducting further diagnosis or confirmation steps for social anxiety disorder.

Similar to the standard screening process administered by mental health professionals, the tool requires inputs from the patient's responses to the questions related to the attributes used by the model. The machine learning model analyses the responses and predicts the probability of the patient manifesting social anxiety disorder or the lack thereof. The result is then displayed on the screen for the mental health professional to see and decide should there be a need to conduct a further diagnosis. Figure 3 illustrates the system flow of how mental health professionals utilize the tool in conducting screening tests.

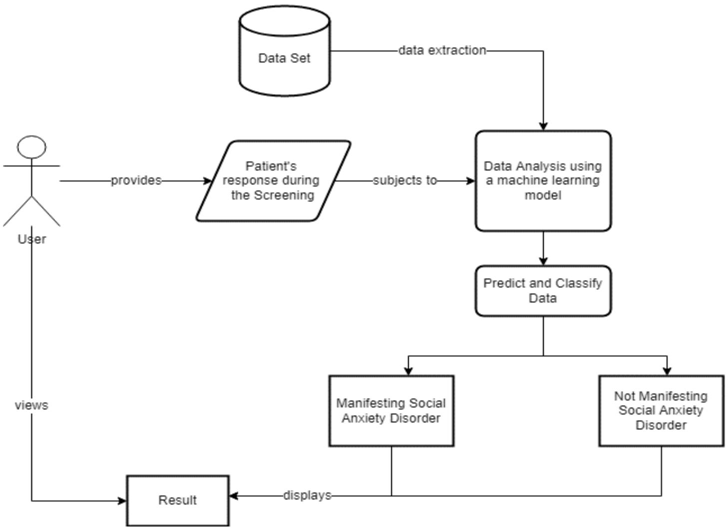

Fig. 3. System process diagram.

4. Results and Analysis

Initially, four classification algorithms were considered: Decision Tree, Logistic Regression, K-Nearest Neighbours (KNN), and Support Vector Machine (SVM). A class imbalance highly influences the performance of these models. Fortunately, the dataset used contains 52% positive and 48% negative cases, which suggests that the data is balanced. Therefore, there was no need to employ techniques such as Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic (ADASYN) to balance the dataset.

4.1. Feature engineering

This section discusses the result of extracting, organizing, transforming, and selecting the dataset’s essential features that the machine learning model used.

4.1.1. Handling Missing Data

During this stage, it was found that the column LSAS contained missing data or NaN values as seen in Fig. 4.

	id	Age	EducationLevel	Gender	ATF	EAF	TKF	CMT	...	CP	NS	DZ	UR	UB	MD	TG	hasSAD	SPIN	LSAS
4	9	33	5	1	5	1	5	1	...	0	0	0	0	0	0	0	0	16	NaN
7	17	27	2	1	7	5	7	7	...	0	0	0	1	0	0	0	0	18	NaN
19	68	27	6	1	7	2	8	2	...	0	0	0	0	0	0	0	1	26	NaN
22	73	28	5	0	6	3	7	4	...	0	1	0	0	1	0	0	0	20	NaN
23	76	38	5	0	0	0	0	0	...	0	0	0	0	0	0	0	1	26	NaN
27	89	27	5	0	2	1	6	3	...	0	0	0	0	0	0	0	1	31	NaN
28	102	44	4	0	4	1	3	0	...	1	1	0	1	0	0	1	1	25	NaN
37	142	35	5	1	1	2	0	0	...	0	0	0	0	0	0	0	1	29	NaN
40	157	28	5	0	2	2	3	3	...	0	1	1	1	0	0	1	1	26	NaN
48	183	21	3	1	1	1	5	3	...	0	1	0	0	1	0	0	1	32	NaN
49	187	40	4	1	6	0	2	4	...	0	0	0	0	0	0	0	1	35	NaN
53	205	27	5	1	8	2	2	2	...	1	0	0	0	0	0	0	0	7	NaN
56	214	32	5	1	5	0	8	2	...	0	0	0	0	1	0	0	0	9	NaN

Fig. 4. Dataset rows with NaN values.

Imputation technique was used to handle the missing data. Three imputation techniques for continuous variables exist: *Mean Imputation*, *Median Imputation*, and *Mode Imputation*. The study specifically used *Median Imputation* to replace missing values since the data is skewed as seen in Fig. 5.

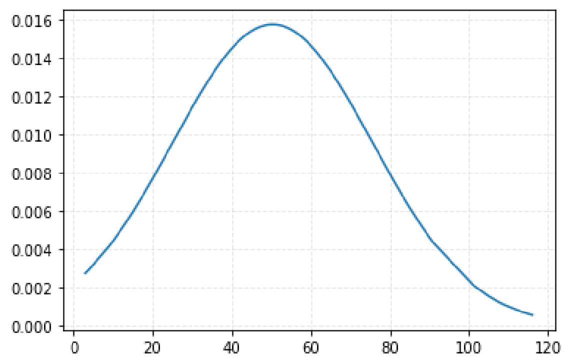

Fig. 5. LSAS column distribution.

4.1.2. Feature scaling

There were columns in the data set with continuous values, which do not contribute evenly to model fitting and may thus introduce bias. To solve the problem, feature scaling techniques such as Standardization and Normalization were performed.

Figures 6 and 7 show the data before and after feature scaling techniques are applied, respectively.

	Age	ATF	EAF	TKF	CMT	DEF	SMF	ERF	DAF	LSAS	SPIN
0	27	4	2	6	2	0	1	1	1	39.0	23
1	24	3	0	3	1	0	4	4	4	43.0	20
2	32	4	1	7	0	0	1	0	4	50.0	33
3	37	4	0	6	1	0	2	3	1	44.0	30
4	33	5	1	5	1	1	5	0	4	46.0	16
...	...	...	...	...	...	...	...	...	...	...	...
209	18	5	8	7	10	0	3	5	4	34.0	28
210	29	6	2	5	2	3	3	6	8	78.0	29
211	24	10	7	9	10	6	6	5	10	46.0	54
212	44	0	0	2	2	0	0	0	2	42.0	11
213	26	7	7	8	6	5	5	6	6	70.0	35

Fig. 6. Data before applying a feature scaling technique.

	Age	ATF	EAF	TKF	CMT	DEF	SMF	ERF	DAF	LSAS	SPIN
0	-0.412931	0.203774	0.007842	0.620108	-0.186235	-0.650294	-0.746638	-0.652997	-0.689983	-0.467432	0.052441
1	-0.725921	-0.165783	-0.831200	-0.375165	-0.555255	-0.650294	0.340304	0.441583	0.428626	-0.285144	-0.191525
2	0.108717	0.203774	-0.411679	0.951866	-0.924275	-0.650294	-0.746638	-1.017857	0.428626	0.033860	0.865664
3	0.630366	0.203774	-0.831200	0.620108	-0.555255	-0.650294	-0.384324	0.076723	-0.689983	-0.239572	0.621697
4	0.213047	0.573331	-0.411679	0.288350	-0.555255	-0.189490	0.702618	-1.017857	0.428626	-0.148428	-0.516814
...	...	...	...	...	...	...	...	...	...	...	...
209	-1.351899	0.573331	2.524966	0.951866	2.765928	-0.650294	-0.022010	0.806443	0.428626	-0.695292	0.459053
210	-0.204272	0.942888	0.007842	0.288350	-0.186235	0.732120	-0.022010	1.171303	1.920105	1.309875	0.540375
211	-0.725921	2.421117	2.105445	1.615382	2.765928	2.114534	1.064932	0.806443	2.665845	-0.148428	2.573431
212	1.360675	-1.274454	-0.831200	-0.706923	-0.186235	-0.650294	-1.108952	-1.017857	-0.317114	-0.330716	-0.923425
213	-0.517261	1.312445	2.105445	1.283624	1.289847	1.653729	0.702618	1.171303	1.174366	0.945299	1.028308

Fig. 7. Data after applying a feature scaling technique.

4.1.3. Feature selection

Feature selection is essential in feature engineering since it has a significant effect on the performance of the machine learning model where partially relevant features can have a negative impact on its performance. Three Feature Selection techniques were performed namely: *Univariate Selection*, *Feature Importance*, and *Correlation Matrix*, to reduce overfitting, improve accuracy, and reduce training time. The common features between the three techniques were used for the final dataset.

Figure 8 shows the relevant features in the data set after applying the feature selection techniques.

	DAF	TKF	LSAS	SW	ERF	ATF	SPIN	SMF	TR	EAF	CMT	DEF
0	-0.689983	0.620108	-0.467432	0	-0.652997	0.203774	0.052441	-0.746638	1	0.007842	-0.186235	-0.650294
1	0.428626	-0.375165	-0.285144	0	0.441583	-0.165783	-0.191525	0.340304	0	-0.831200	-0.555255	-0.650294
2	0.428626	0.951866	0.033860	0	-1.017857	0.203774	0.865664	-0.746638	1	-0.411679	-0.924275	-0.650294
3	-0.689983	0.620108	-0.239572	0	0.076723	0.203774	0.621697	-0.384324	1	-0.831200	-0.555255	-0.650294
4	0.428626	0.288350	-0.148428	0	-1.017857	0.573331	-0.516814	0.702618	0	-0.411679	-0.555255	-0.189490
...	...	...	...	...	...	...	...	...	...	...	...	...
209	0.428626	0.951866	-0.695292	1	0.806443	0.573331	0.459053	-0.022010	1	2.524966	2.765928	-0.650294
210	1.920105	0.288350	1.309875	0	1.171303	0.942888	0.540375	-0.022010	1	0.007842	-0.186235	0.732120
211	2.665845	1.615382	-0.148428	0	0.806443	2.421117	2.573431	1.064932	1	2.105445	2.765928	2.114534
212	-0.317114	-0.706923	-0.330716	0	-1.017857	-1.274454	-0.923425	-1.108952	0	-0.831200	-0.186235	-0.650294
213	1.174366	1.283624	0.945299	1	1.171303	1.312445	1.028308	0.702618	1	2.105445	1.289847	1.653729

Fig. 8. Relevant features in the dataset.

4.2. Model selection

K-Fold Cross-Validation (Fig. 9) was used to assess each machine learning model's performance by folding the given dataset *k* times. This approach is handy since it normally results in a less biased manner thereby building a more generalized model where it can perform well on unseen data.

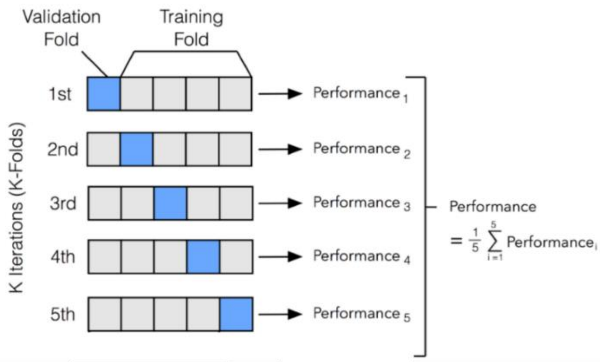

Fig. 9. K-fold cross validation.

The following tables and figures present the model training and testing results using the K-Fold Cross Validation strategy.

4.2.1. Accuracy

Table 1 shows the accuracy score of the machine learning models after cross-validation.

Table 1. Accuracy score of the machine learning models.

ML Model	Accuracy (%)
Decision Tree	94.06
Logistic Regression	95.05
K-Nearest Neighbours (KNN)	87.27
Support Vector Machine (SVM)	96.01

Accuracy, being the most used metric for evaluating the performance of a machine learning model, was obtained after cross-validation. The KNN model showed the worst accuracy score of 87.27% among the other models, followed by the Decision Tree model, with an accuracy score of 94.06%. On the other hand, the

Logistic Regression model showed a generally impressive accuracy of 95.05%. However, the SVM model, one of the most commonly known models to perform well in classification problems, garnered the highest accuracy score of 96.01%.

4.2.2. Precision

Table 2 shows the precision score of the machine learning models after cross-validation.

Table 2. Precision score of the machine learning models.

ML Model	Precision (%)
Decision Tree	94.16
Logistic Regression	96.00
K-Nearest Neighbuors (KNN)	91.01
Support Vector Machine (SVM)	97.13

Like accuracy, precision is one of the indicators of a machine learning model's performance. After cross-validation, the KNN model showed the lowest precision score of 91.01%, followed by the Decision Tree model, with a precision score of 94.16%. Meanwhile, the Logistic Regression and SVM models performed well by obtaining 96.00% and 97.13% precision scores, respectively.

4.2.3. Recall

Table 3 shows the recall score of the machine learning models after cross-validation.

Table 3. Recall score of the machine learning models.

ML Model	Recall (%)
Decision Tree	96.24
Logistic Regression	96.00
K-Nearest Neighbours (KNN)	88.16
Support Vector Machine (SVM)	95.33

Recall is one of the metrics for evaluating a machine learning model that tells how good the model is in classifying actual positive cases. After cross-validation, the KNN model showed the worst recall score of 88.16%. The SVM model, having the best scores from both accuracy and precision, actually underperformed with a recall score of 95.33%. On the other hand, the Logistic Regression model showed a consistent performance with a recall score of 96.00%. However, the Decision Tree model performed the best, with a recall score of 96.24%.

4.2.4. F-measure or F1

Table 4 shows the f1 score of the machine learning models after cross-validation.

Table 4. F-measure score of the machine learning models.

ML Model	F1 (%)
Decision Tree	94.04
Logistic Regression	96.00
K-Nearest Neighbours (KNN)	89.22
Support Vector Machine (SVM)	96.13

Comparing models' performance based only on precision or recall seems inappropriate since it is not possible to maximize these two metrics simultaneously due to the trade-off between precision and recall. To solve the problem, F1 metric was obtained. This metric sums the performance of a model by combining the otherwise competing metrics – precision and recall.

After the cross-validation, the KNN model showed the worst score of 89.22% for F1, followed by the Decision Tree model with an F1 score of 94.04%. Logistic Regression model, on the other hand, showed a generally impressive performance by garnering an F1 score of 96.00%. However, the SVM model again showed the best score of 96.13% for F1.

4.2.5. Area under the curve (AUC)

Table 5 shows the AUC score of the machine learning models after cross-validation.

Table 5. AUC score of the machine learning models.

ML Model	AUC
Decision Tree	0.93 ± 0.06
Logistic Regression	0.96 ± 0.08
K-Nearest Neighbours (KNN)	0.88 ± 0.19
Support Vector Machine (SVM)	0.97 ± 0.05

AUC indicates how well the machine learning model can distinguish between classes. To determine the better score for AUC, the mean value and the width of the confidence interval should be considered. In this case, the highest mean value and the narrowest confidence level would be considered better.

After cross-validation, the KNN model scored the lowest with 0.88 ± 0.19 AUC score. It was followed by the Decision Tree model with 0.93 ± 0.06 AUC score. Meanwhile, the Logistic Regression and Support Vector Machine models performed well by obtaining 0.96 ± 0.08 and 0.97 ± 0.05 AUC scores, respectively.

Figure 10 shows the ROC curve of the SVM model during all the iterations of cross-validation.

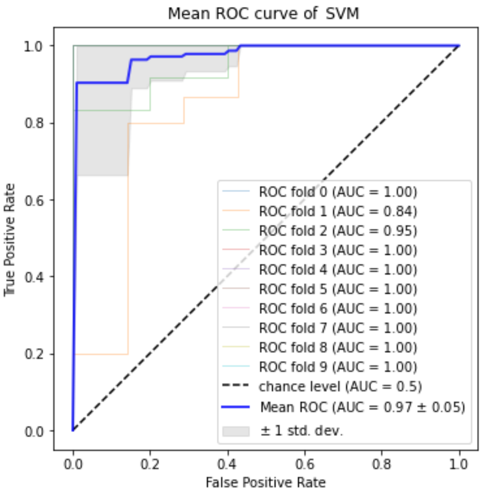

Fig. 10. ROC curve of SVM during cross-validation.

4.2.6. Performance summary

Table 6 shows the summary of the results of the different models according to their accuracy, precision, recall, f1 and AUC score.

Table 6. Performance summary of the machine learning models.

ML Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	AUC
Decision Tree	94.06	94.16	96.24	94.04	0.93 ± 0.06
Logistic Regression	95.05	96.00	96.00	96.00	0.96 ± 0.08
K-Nearest Neighbors (KNN)	87.27	91.01	88.16	89.22	0.88 ± 0.19
Support Vector Machine (SVM)	96.01	97.13	95.33	96.13	0.97 ± 0.05

4.2.7. Confusion matrix

A confusion matrix allows one to visually assess the performance of the machine learning model. It also indicates how many correct and wrong predictions the model made during cross-validation. Figure 11 depicts the SVM model's confusion matrix.

The Confusion Matrix showed that out of the 103 negative instances (labelled as 0, or a person with no SAD), there were 99 cases for True Negative where the model correctly predicted a negative case. In contrast, there were only four cases for False Positive where the model incorrectly predicted a negative case.

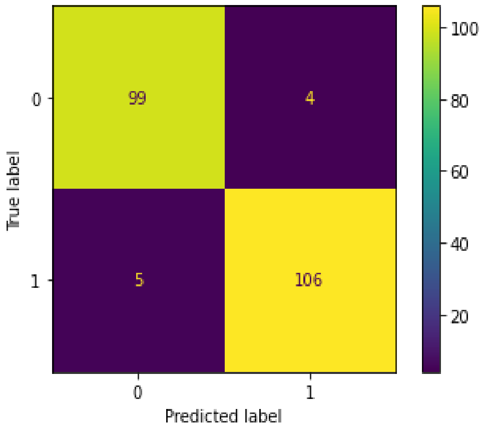

Fig. 11. Confusion matrix for the SVM model.

In addition, out of the 111 positive instances (labelled as 1, or a person with SAD), there were 106 cases for True Positive where the model correctly predicted a positive case. In contrast, there were only five cases for False Negative where the model incorrectly predicted a positive case.

Moreover, the confusion matrix visually represents how impressive the SVM model performed under the testing conditions. It also demonstrates that the model does a great job of correctly classifying people with or without a social anxiety disorder.

Furthermore, the confusion matrix showed that correctly classifying negative cases is as important as correctly classifying positive ones. In this case, the SVM model showed that we cannot miss negative cases, nor should we diagnose negative ones as positive. Doing so would put healthy people through serious treatments while it jeopardizes sick people into thinking they are well, and eventually undermine trust in the diagnostic process.

4.3. System validation

The tool was tested by five licensed psychologists. These professionals were invited for the system validation. The tool was tested based on the Functionality metric containing Data Validity and Accuracy sub-metrics. Accuracy, in this metric, measures the adequacy of the system to meet its objectives. Meanwhile, Data Validity tells whether the system contains input validation to avoid erroneous data entry.

Figure 12 shows the ratings of the professionals based on the functionality metric.

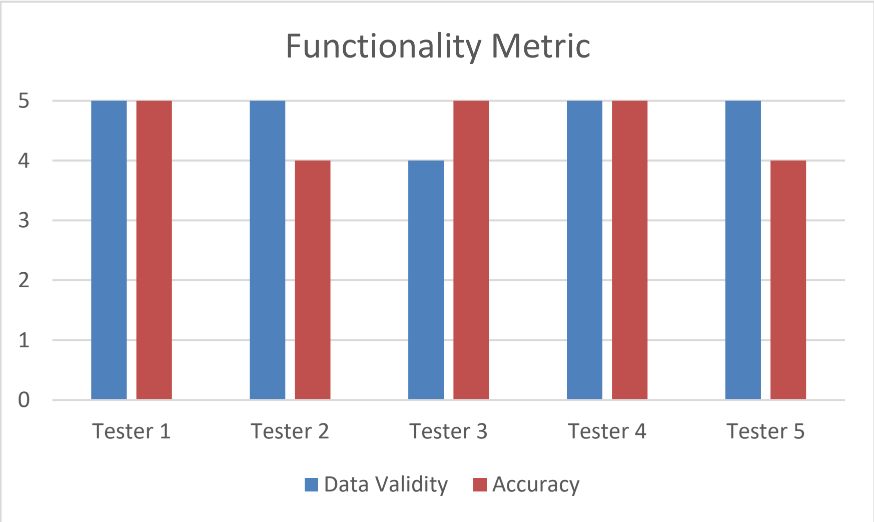

Fig. 12. Functionality metric.

4.4. System implementation

The system is a web-based application developed using React.js and Flask. The method comprises the screening tool with the machine learning model, and a patient management system. Figure 13 shows the starting page with instructions when a clinician uses the screening tool in the system. Meanwhile, Fig. 14 shows a table where the clinician can choose which patient to screen.

The tool aims to be able to determine whether the patient is possibly manifesting a social anxiety disorder or not. For the tool to perform predictions, clinicians will ask a series of questions to the patient, and the clinician is responsible for asking the questions. Figure 15 shows a particular section during the screening process.

The SVM model processes the array of responses provided by the user to predict whether the patient has a social anxiety disorder or not. Figure 16 shows the model’s prediction based on the array of responses provided.

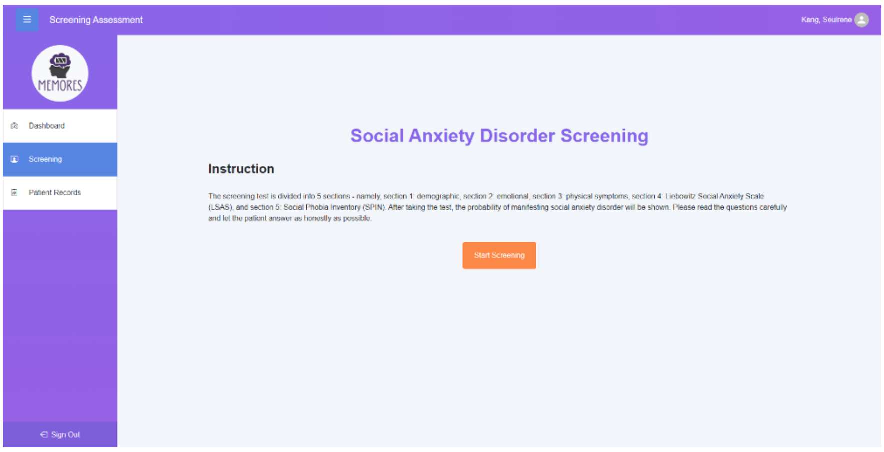

Fig. 13. System interface for the screening tool.

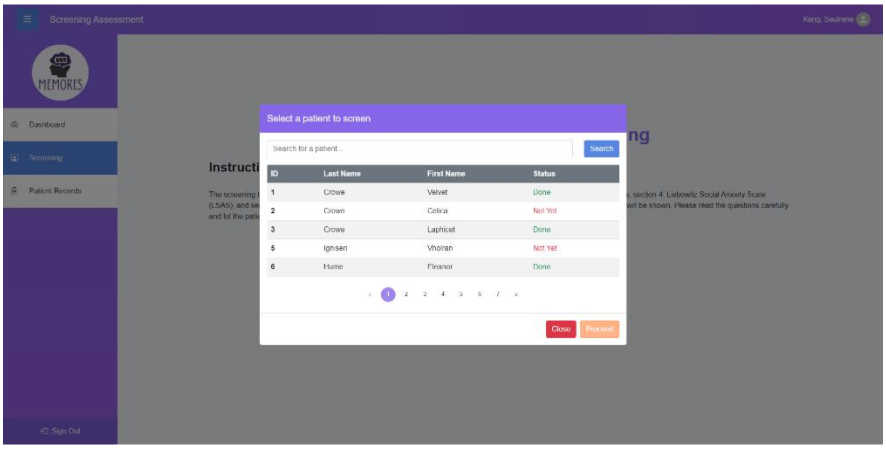

Fig. 14. System interface (Patient selection).

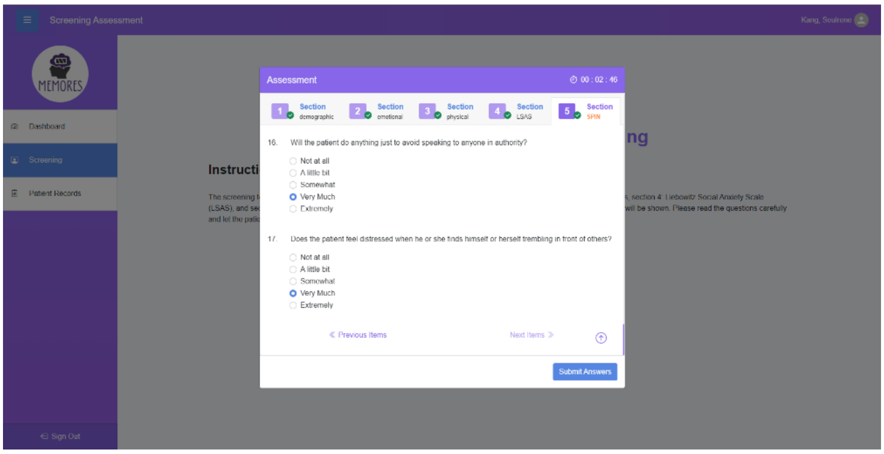

Fig. 15. System interface (Screening section).

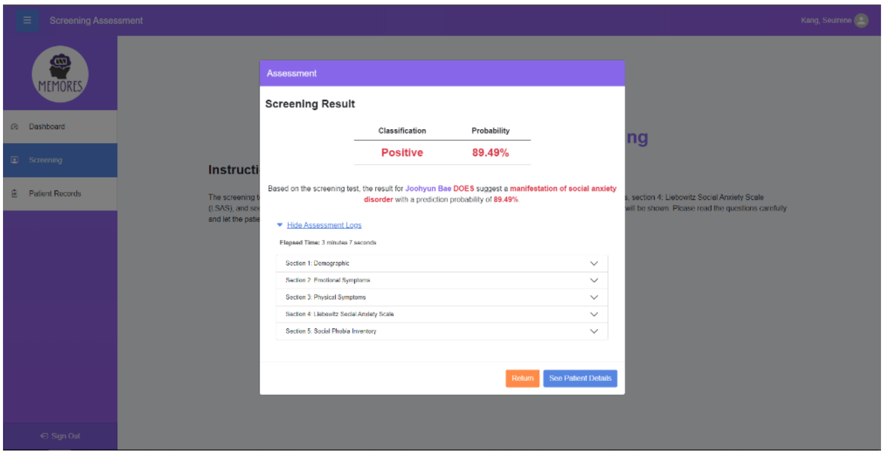

Fig. 16. System interface (Screening result).

5. Conclusion and Recommendations

The authors developed a screening tool to streamline the screening process without compromising the accuracy rate of predicting a possible manifestation of social anxiety disorder within patients. The tool assessed the possibility of SAD based on the information gathered about the patient's demographic, emotional symptoms, physical symptoms, and social symptoms.

Before the system development, four machine learning models – Decision Tree, Logistic Regression, Support Vector Machine, and K-Nearest Neighbours – were trained to predict SAD. K-Fold Cross Validation evaluated these four machine learning models based on accuracy, precision, recall, f1, and AUC scores. The Decision Tree model obtained 94.06%, 94.16%, 96.24%, 94.04%, and 0.93 ± 0.06 for accuracy, precision, recall, f1, and AUC, respectively. Meanwhile, the Logistic Regression model obtained 95.05%, 96.00%, 96.00%, 96.00%, and 0.96 ± 0.08 for accuracy, precision, recall, f1, and AUC, respectively. On the other hand, the K-Nearest Neighbours (KNN) model obtained a significantly low score of 87.27%, 91.01%, 88.16%, 89.22%, and 0.88 ± 0.19 for accuracy, precision, recall, f1, and AUC, respectively. Finally, the SVM model got the highest score overall with 96.01%, 97.13%, 95.33%, 96.13%, and 0.97 ± 0.05 for accuracy, precision, recall, f1, and AUC, respectively - hence, used as the machine learning model behind the screening tool.

The Confusion Matrix for the tool's integrated model showed that out of 103 negative instances, 99 were True Negative, indicating that the model correctly predicted a negative case, while only four were False Positive, indicating that the model incorrectly predicted a negative case. Additionally, it also showed that out of 111 positive instances, there were 106 True Positive cases where the model correctly predicted a positive case, and only 5 False Negative cases where the model incorrectly predicted a positive case.

The tool was tested by five licensed psychologists for system validation based on (1) Accuracy, measures the adequacy of the system to meet its objectives, and (2) Data Validity, tells whether the system contains input validation to avoid erroneous data entry. For the tool's data validity, four professionals rated a score of 5.0 while one professional rated a score of 4.0. For the tool's accuracy, three professionals rated a score of 5.0 while the remaining two rated a score of 4.0.

The tool is made to help professionals screen patients with social anxiety disorder and potentially increase the chances of correctly classifying these patients before conducting further clinical diagnosis. This study is designed to lay the groundwork for future research in the same topic.

For future researchers, the authors of this study recommend having an additional data set for training and testing to improve the model's performance. The authors would also recommend conducting a hyperparameter tuning of these models as it may significantly improve their performance. Additionally, the authors also recommend performing external validation. Furthermore, future researchers may also use classification algorithms other than the ones presented in this study. Lastly, a technique called ensemble learning is encouraged to be utilized by future researchers.

Acknowledgement

The authors would like to dedicate an acknowledgment of gratitude towards the following individuals:

We want to thank our thesis adviser, Professor Glenn B. Pepito, for the guidance and support that he has provided in the making of this research. We thank Dr Rhoda Alvarez for the insightful discussions during the formulation stage. Without her, we would not possibly be able to understand the diagnostic process for diagnosing patients with social anxiety disorder. Lastly, we would like to thank our dear parents for their continuous support, which helped motivate the researchers to finish this research with flying colours.

Nomenclatures

Abbreviations

ADASYN	Adaptive Synthetic
AUC	Area Under the Curve
DSM-5	Diagnostic and Statistical Manual of Mental Disorders - 5
ICD-10	International Classification of Disease, 10^{th} Revision
KNN	K-Nearest Neighbours
LSAS	Liebowitz Social Anxiety Scale
MINI	Mini International Neuropsychiatric Interview
ML	Machine Learning
ROC	Receiver Operating Characteristic
SAD	Social Anxiety Disorder
SMOTE	Synthetic Minority Oversampling Technique
SPIN	Social Phobia Inventory
SVM	Support Vector Machine

References

1. Vermani, M.; Marcus, M.; and Katzman, M.A. (2011). Rates of detection of mood and anxiety disorders in primary care: a descriptive, cross-sectional study. *The Primary Care Companion for CNS Disorders,* 13(2).

2. Liu, F.; Guo, W.; Fouche, J.P.; Wang, Y.; Wang, W.; Ding, J.; Zeng, L.; Qiu, C.; Gong, Q.; Zhang, W.; and Chen, H. (2015). Multivariate classification of social anxiety disorder using whole brain functional connectivity. *Brain Structure and Function*, 220(1), 101-115.
3. Pantazatos, S.; Talati, A.; Schneier, F.R.; and Hirsch, J. (2014). Reduced anterior temporal and hippocampal functional connectivity during face processing discriminates individuals with social anxiety disorder from healthy controls and panic disorder, and increases following treatment. *Neuropsychopharmacology*, 39(2), 425-434.
4. Frick, A.; Gingnell, M.; Marquand, A.F.; Howner, K.; Fischer, H.; Kristiansson, M.; Williams, S.C.R.; Fredrikson, M.; and Furmark, T. (2014). Classifying social anxiety disorder using multivoxel pattern analyses of brain function and structure. *Behavioural Brain Research*, 259(100), 330-335.
5. Asvestopoulou, T.; Manousaki, V.; Psistakis, A.; Smyrnakis, I.; Andreadakis, V.; Aslanides, I.M.; and Papadopouli, M. (2019). DysLexML: screening tool for dyslexia using machine learning. *arXiv.org*. arXiv:1903.06274v1.
6. Chiu, P.-Y.; Tang, H.; Wei, C.-Y.; Zhang, C.; Hung, G.-U.; and Zhou, W. (2019). NMD-12: a new machine-learning derived screening instrument to detect mild cognitive impairment and dementia. *PLoS ONE*, 14(3): e0213430.
7. Md, A.Q.; Kulkarni, S.; Jackson, J.C.; Vaichole, T.; Mohan, S.; and Iwendi, C. (2023). Enhanced preprocessing approach using ensemble machine learning algorithms for detecting liver disease. *Biomedicines*, 11(2), 581.
8. Kumar, A.; Md, A.Q.; Jackson, J.C.; and Iwendi, C. (2023). Predicting and curing depression using long short term memory and global vector. *Computers, Materials & Continua*, 74(3), 5837-5852.
9. Fathi, S.; Ahmadi, M.; Birashk, B.; and Dehnad, A. (2020). Development and use of a clinical decision support system for the diagnosis of social anxiety disorder. *Computer Methods and Programs in Biomedicine*, 190, 105354.