

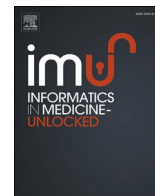


Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Informatics in Medicine Unlocked

journal homepage: [www.elsevier.com/locate/imu](http://www.elsevier.com/locate/imu)

# Prediction of generalized anxiety levels during the Covid-19 pandemic: A machine learning-based modeling approach

Faisal Mashel Albagmi<sup>a,\*</sup>, Aisha Alansari<sup>b</sup>, Deema Saad Al Shawan<sup>c</sup>, Heba Yaagoub AlNujaidi<sup>c</sup>, Sunday O. Olatunji<sup>b</sup>

<sup>a</sup> College of Applied Medical Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

<sup>b</sup> College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

<sup>c</sup> College of Public Health, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

## ARTICLE INFO

## Keywords:

COVID-19  
Pandemic  
Anxiety  
Machine learning  
Saudi Arabia

## ABSTRACT

The rapid spread of the Covid-19 outbreak led many countries to enforce precautionary measures such as complete lockdowns. These lifestyle-altering measures caused a significant increase in anxiety levels globally. For that reason, decision-makers are in dire need of methods to prevent potential public mental crises. Machine learning has shown its effectiveness in the early prediction of several diseases. Therefore, this study aims to classify two-class and three-class anxiety problems early by utilizing a dataset collected during the Covid-19 pandemic in Saudi Arabia. The data was collected from 3017 participants from all regions of the Kingdom via an online survey containing questions to identify factors influencing anxiety levels, followed by questions from the GAD-7, a screening tool for Generalized Anxiety Disorders. The prediction models were built using the Support Vector Machine classifier for its robust outcomes in medical-related data and the J48 Decision Tree for its interpretability and comprehensibility. Experimental results demonstrated promising results for the early classification of two-class and three-class anxiety problems. As for comparing Support Vector Machine and J48, the Support Vector Machine classifier outperformed the J48 Decision Tree by attaining a classification accuracy of 100%, precision of 1.0, recall of 1.0, and f-measure of 1.0 using 10 features.

## 1. Introduction

One of the shared global experiences of the COVID-19 pandemic is the experience of “lockdown.” However, such strict restrictions vary from country to country and change over time. The consequences of lockdowns on mental health have been substantial [1–3]. Lockdown conditions lead to social isolation and confinement, which can impact the population’s mental health. Furthermore, this crisis can have a broader impact on education, work, everyday life, and implications for mental health services. The Lancet psychiatry has highlighted the need for mental health services during lockdowns, especially for the most vulnerable groups such as students, medical professionals, and women [4]. As precautionary measures impact large portions of the population, it is expected that mental health problems will be on the increase globally [1,5,6]. According to the Anxiety and Depression Association of America, the commonest mental health problem in the United States was Generalized Anxiety Disorder (GAD). Approximately one-third of the

American population suffers from GAD, but only less than half of them have access to mental healthcare [7].

Despite the extensive research, the magnitude and the underlying factors of GAD during lockdown are unknown. However, there has been evidence to suggest that early identification and access to mental health treatment may help mitigate the impact of mental health, especially GAD. Using technology such as telepsychiatry for frontline medical workers and vulnerable populations during lockdown times could alleviate the effects of mental health in the population affected by the COVID-19 pandemic [2]. Nevertheless, research could guide mental health services such as telepsychiatry to target the most vulnerable population groups, especially during pandemic lockdowns [3]. For that reason, machine learning can be a powerful tool to enable decision makers to customize mental health services depending on the predicted needs of different of these subpopulations.

The early preparation for the potential mental health needs is crucial to prevent a mental health crisis. According to Thompson and his

\* Corresponding author.

E-mail addresses: [fmalbagmi@iau.edu.sa](mailto:fmalbagmi@iau.edu.sa) (F.M. Albagmi), [2180004329@iau.edu.sa](mailto:2180004329@iau.edu.sa) (A. Alansari), [dshawan@iau.edu.sa](mailto:dshawan@iau.edu.sa) (D.S. Al Shawan), [hnujaidi@iau.edu.sa](mailto:hnujaidi@iau.edu.sa) (H.Y. AlNujaidi), [osunday@iau.edu.sa](mailto:osunday@iau.edu.sa) (S.O. Olatunji).

<https://doi.org/10.1016/j.imu.2022.100854>

Received 2 November 2021; Received in revised form 3 January 2022; Accepted 8 January 2022

Available online 19 January 2022

2352-9148/© 2022 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Table 1**  
Overview of the reviewed sources arranged by data of publication.

| Author (s)<br>Citation | Title of article or chapter   | Objective   | Method  | Findings  | Limitations  |
|------------------------|---|---|---|---|--|
| [17]                   | Modeling anxiety and fear of COVID-19 using machine learning in a sample of Chinese adults: associations with psychopathology, sociodemographic, and exposure variables | To examined vulnerability factors associated with increased anxiety and fear.   | The researchers used R caret package for machine learning, with packages for specific algorithms of glmnet (lasso, ridge, and elastic net regression), rf (random forest), xgbTree (extreme gradient boosted regression), and svmRadial (support vector machine with a radial basis function kernel).   | Stress and rumination were the most relevant variables in modeling COVID-19-related anxiety intensity, according to shrinkage machine learning methods. The most powerful predictor of perceived COVID-19 death threat was health anxiety.  | Data was from one geographical area china. They only included self-report measures of psychopathology  |
| [18]                   | Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence                        | To identify important predictors for GAD and MDD risk using artificial intelligent  | A novel machine learning process was used to re-analyze data from an observational study to tackle the problem of predicting MDD and GAD. The pipeline is an algorithmically diverse collection of machines learning approaches, including deep learning.   | Being comfortable with living conditions and having public health insurance were the two most important factors in predicting MDD. Up-to-date vaccinations and marijuana usage were the two most powerful predictors of GAD. Our findings show that machine learning algorithms for detecting GAD and MDD based on EHR data have a moderate predictive performance. | The original screening for MDD and GAD outcomes may not have identified all cases in the community. The research originates from French college students, who are likely to have different baselines than other psychiatric populations. |
| [19]                   | Predicting generalized anxiety disorder among women using Shapley value   | To predict GAD among women using Shapley value  | On the mental health data set, the Shapley value was used as the feature selection for the data mining classifier.  | The finding has been improved using feature selection among the prediction's models (Naïve Bayes, Random Forest and J48).   | Small sample size 180 participants   |
| [15]                   | Toward Robust Anxiety Biomarkers: A Machine Learning Approach in a Large-Scale Sample.  | To predict trait anxiety from neuroimaging measurements in humans.  | They compared a suite of neuroimaging-based machine learning models using Python to predict anxiety within a discovery sample (n = 531, 307 women) via k-fold cross-validation. The final model using (a stacked model incorporating region-to-region functional connectivity, amygdala seed-to-voxel connectivity, and volumetric and cortical thickness data) in a held-out, unseen test sample (n = 348, 209 women). | Stacked model was able to predict anxiety within the discovery sample. But failed to test the generalizability in the holdout sample.   | The researchers studied a limited set of brain phenotypes and applied a circumscribed set of approaches. They didn't analyze a clinical sample. The imaging sequences used lack the spatial and temporal precision of current approaches |
| [20]                   | Assessment of Anxiety, Depression and Stress using Machine Learning Models  | To predict anxiety, depression, and stress using 8 algorithms.  | Using data from the online DASS42 tool, eight machine learning algorithms were used to predict the occurrence of psychological issues such as anxiety, depression, and stress.  | The prediction accuracy obtained by utilizing the hybrid algorithm was higher than that obtained by using single methods, although the radial basis function network, which falls within the category of neural networks, yielded the highest accuracy.   | NA   |
| [6]                    | Learning the Mental Health Impact of COVID-19 in the United States with Explainable Artificial Intelligence   | To focus on learning a ranked list of factors that could indicate a predisposition to a mental disorder during the COVID-19 pandemic. | They surveyed 17,764 adults in the United States using Bayesian network inference, they have identified key factors affecting mental health during the COVID-19 pandemic.   | They discovered that patients with a chronic mental disease were more susceptible to mental problems during the COVID-19 pandemic using the Bayesian network model.   | The data analyzed is limited to one geographical area (united stated)  |
| [21]                   | Screening of anxiety and depression among seafarers using machine learning technology   | To compare performance of different machine learning algorithms for screening of anxiety and depression among the seafarers.          | After obtaining the required approval and ethical clearance, 470 sailors were interviewed at the Haldia Dock Complex in India. Five machine learning classifiers i.e., CatBoost, Logistic Regression, Naïve Bayes, Random Forest, and Support Vector Machine, were evaluated using the Python programming language.   | They found that Catboost appeared to be the best one for predicting anxiety and depression with accuracy and precision 82.6% and 84.1% respectively.  | The study emphasized the application of machine learning technology in the field of automated screening for mental health illness.   |
| [22]                   | Detecting anxiety on Reddit   | To detect anxiety related posts from Reddit using   | study anxiety disorders through personal narratives collected through the   | apply N-gram language modeling, vector embeddings, topic  | They achieve an accuracy of 91% with vectorspace word embeddings, and an   |

(continued on next page)

Table 1 (continued)

| Author (s) Citation | Title of article or chapter | Objective                    | Method                               | Findings   | Limitations   |
|---------------------|-----------------------------|------------------------------|--------------------------------------|--|---|
|                     |                             | various linguistic features. | popular social media website, Reddit | analysis, and emotional norms to generate features that accurately classify posts related to binary levels of anxiety. | accuracy of 98% when combined with lexiconbased features. |

colleagues investigated the delay in seeking treatment for anxiety and mood disorders. They found that people delay seeking help for around 8.2 years. Moreover, they reported two main indicators associated with this delay slower problem recognition and younger age at onset. As older people take a longer time to contact initial treatment. This could be effectively prevented by early prediction of anxiety using machine learning models [8].

Several studies aimed to assess the psychological impact of the pandemic on the Saudi population, which enforced a complete lockdown in March 2020 [9]. However, most of these studies lack modeling of the collective effects of GAD on the population in a pandemic. This study addresses this gap by using supervised machine learning algorithms, which is an explainable artificial intelligence approach to capture the joint multivariate distribution underlying extensive survey data collected across Saudi Arabia during a lockdown. The choice of machine learning algorithm selected paid succinct attention to models that have proved their success in various medical applications, namely, Support Vector Machine (SVM) and J48 Decision Tree (DT). Support Vector Machine learning is a well-established algorithm for both classification and regression with extensive successful applications in several fields, including medical and biomedical applications, while decision tree is well known for its clarity and easier understanding even to non-computer professionals, thereby making it appealing to the medical and public health professionals coupled with its excellent performance in various applications. Empirical results showed that SVM outperformed the J48 Decision Tree using the ten highest correlated features and the optimized hyperparameters, achieving a 100% accuracy in anxiety binary [10]. As for comparing Support Vector Machine and J48, the Support Vector Machine classifier outperformed the J48 Decision Tree by attaining a classification accuracy of 100%, precision of 1.0, recall of 1.0, and f-measure of 1.0 using 10 features. Although the J48 decision tree achieved lesser performance measures, the highest being 95% accuracy, nonetheless, it offers the possibility of having a better explainability to non-computer professionals in understanding how the developed models worked. In fact, the potentials of the proposed machine learning models in mitigating the late effect of anxiety cannot be overemphasized.

## 2. Review of related literature

The public health mental crisis during COVID-19 has been studied by several researchers worldwide. According to a study conducted in China in 2020, one-third of the participants reported moderate-to-severe anxiety, and more than half of the participants had a moderate-to-severe psychological impact [11].

There are several studies that aim to assess the psychological impact of the pandemic on the Saudi Population. For instance, Albagmi and his colleagues assessed the prevalence of anxiety and associated factors during the lockdown period at the peak of the outbreak in Saudi Arabia. A total of 3,017 respondents from all five main regions of Saudi Arabia completed the survey. The results indicated that 19.6% of the respondents had a moderate to severe level of anxiety during the COVID-19 pandemic [9]. The factors that were associated with a higher level of anxiety included being female, being a student, being single or divorced, and living with a family member who is vulnerable to COVID-19.

Similarly, another study conducted in Saudi Arabia measured the

impact of the pandemic on the psychological disposition of a total of 2081 Saudi residents and citizens. According to the results, 7.3% of the respondents had anxiety. Additionally, the researchers concluded that individuals are more likely to develop depression during the pandemic included non-Saudi, divorcees, the elderly, and university students. As for factors that correlated with a higher level of anxiety, they included "Saudi individuals, married people, the unemployed and those with a high income" [12]. Moreover, another study investigated the anxiety level across students in Saudi Arabia during the COVID-19 pandemic. The study revealed that 35% of students experienced moderate to severe anxiety. Female and fourth-year students were more anxious compared to their counterparts [13].

In recent years, there has been an increasing interest in using machine learning models in predicting anxiety disorders. These prediction models are appealing to decision-makers due to their ability to detect the potential outcomes of different courses of action. These tools are handy to assess the potential impact of public mental health crises and understand their associated factors' dynamics. Pintelas and colleagues., [14] conducted a systematic review of machine learning prediction methods for anxiety disorders. They concluded that the accuracy of these research rely on the type of prediction methods and data acquisitions as clinical data or self or screening tools. Out of the 16 studies examined, they found that the highly used method for predicting post-traumatic stress disorder (PTSD) and Seasonal affective disorder (SAD) were Hybrid methods and Support Vector Machine (SVM), respectively. Also, Artificial neural networks (ANNs) and ensemble methods achieve the highest prediction scores.

Boeke and his colleagues used neuroimaging measurements to predict traits of anxiety using a k-fold cross validation machine across 531, 307 women. They conclude that they did not find evidence of a generalizable anxiety biomarker using different method [15]. Other studies have also predicted GAD among women using data acquired from a self-screening survey. Husain et al. [16] found that the random forest approach showed high prediction accuracy (0.9). This was also investigated by Jothi et al., in 2021 as they used Shapley value as a feature selection to predict GAD among women in Malaysia.

Elhai et al. collected Cross-sectional data from 908 adults from Eastern China. The questionnaire was distributed between 24 February to 15 March 2020, when strict social distancing measures were in place [17]. The authors adopted several instruments to measure the Generalized Anxiety Disorder and other mental illnesses. These tools include the GAD-7, The Depression Anxiety Stress Scale-21 (DASS-21), and the Ruminative Responses Scale. Additionally, the participants were queried and the magnitude of their exposure to pandemic-related news. Furthermore, the researchers utilized multiple machine learning algorithms to customize their model to identify vulnerability factors for COVID-10-influenced anxiety and the perceived threat of death. The study's findings identified several predictors of anxiety severity such as stress, rumination, the threat of death from COVID-19, age, negative consequences of illness, news exposure to coronavirus, and the participant's sex.

Thompson and his colleagues investigated the delay in seeking treatment for anxiety and mood disorders. They found that people delay seeking help for around 8.2 years. Moreover, they reported two main indicators associated with this delay, slower problem recognition and younger age at onset, as older people take a longer time to contact initial

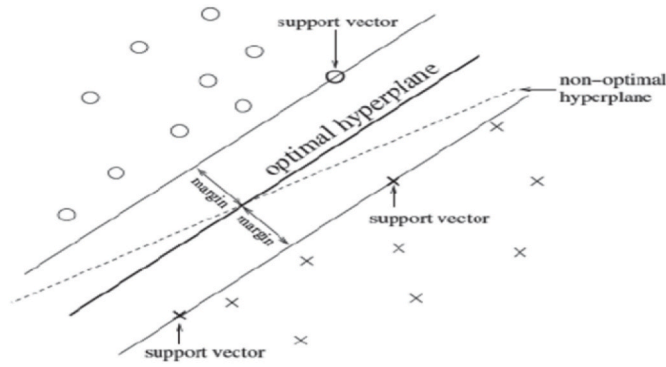


Fig. 1. Maximum hyperplane distance.

treatment, which could be effectively prevented by early prediction of anxiety using machine learning models [8]. Additional studies, methods, results and limitations are described in Table 1.

There are limited studies on the pandemic's impact and its associated factors on public mental health in Saudi Arabia. For that reason, this paper aims to use a carefully selected machine learning algorithm that includes SVM and Decision Tree for predicting anxiety using real-life data collected in Saudi Arabia during the lockdown due to the Covid-19 pandemic. In addition, feature selection was systematically carried out, which identified 10 best features that achieved the highest accuracy out of the 20 available features.

### 3. Description of the proposed techniques

The following subsections exhibit a brief description of the machine learning algorithms utilized in the proposed project for anxiety classification.

#### 3.1. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a promising supervised non-linear machine learning algorithm founded by Cortes, Vapnik, and Boser in the late nineties [23]. Many researchers have commonly adopted SVM for its unique ability to operate with linear and non-linear data and support diverse kernel functions. Nevertheless, SVM's main advantage is its ability to overcome the curse of dimensionality issues and operate successfully with few data through utilizing a generalization control technique [23].

SVM can be employed in both classification and regression problems. However, it is mainly adopted for binary classification applications [24], where it inspects the training instances and determines a hyperplane to classify two classes. The distance between the support vectors and the hyperplane must be maximized to obtain an optimal hyperplane [25], as shown in Fig. 1.

Equation (1) represents the mathematical formula for measuring the hyperplane, where  $w$  denotes the weight vector,  $x$  is the value from the set of labeled training pairs, and  $b$  is the bias.

$$w^T x_i + b = 0 \tag{1}$$

However, minimizing the weight vector is essential for finding the optimal hyperplane to obtain generalization control. Equation (2) represents the mathematical formula for finding the optimum hyperplane through employing the Lagrangian duality theory to give the function more degree of freedom [26].

$$\max a = \sum_{i=1}^n - \frac{1}{2} \sum_{i,j=1}^n y_i a_i y_j a_j \varphi(x_i)^T \varphi(x_j) \tag{2}$$

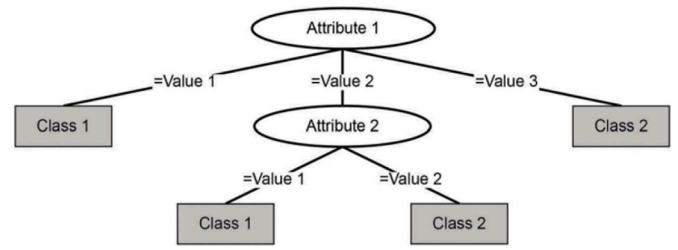


Fig. 2. Decision tree scheme.

$$\text{subject to } \begin{cases} \forall_i a > 0 \\ \sum_i y_i a_i = 0 \end{cases}$$

#### 3.2. J48 decision tree

J48, also known as C4.5, is a supervised machine learning algorithm developed by Ross Quinlan [27]. It is basically a decision tree algorithm extended from the ID3 algorithm [28]. The structure of the J48 tree is composed of three main components, the interior node, which denote the attributes, the branches that give information on the possible values a node can have, and the leaves that determine the final value of classification [29]. Fig. 2 shows the basic structure of a J48 decision tree [29].

J48 utilizes an enhanced procedure of the tree pruning method to overcome the misclassification error that a high noise training dataset can cause. It also uses the divide and conquer approach to partition the data into smaller subsets recursively [30]. As in other decision tree algorithms, the gain is calculated in each step to decide the best attribute in each upcoming node [29]. To calculate the gain, entropy is first computed to evaluate the uncertainty degree of an instance, as shown in Equation (3).

$$E(S) = - \sum_{i=1}^c p_i \log_2(p_i) \tag{3}$$

In Equation 3,  $S$  provides the set of samples,  $c$  corresponds to the number of classes, and  $P_i$  denote the most frequent probability of an element ( $i$ ) in the sample set. The entropy is null when all values are related to one class and is maximum when the sample is proportional [29].

Equation (4) shows how the information gain is calculated, where the biggest possible information gain is calculated [29].  $S_v$  contains the instances that have the value  $v$  in feature  $A$ , whereas  $V(A)$  contains the values of feature  $A$ .

$$IG(S, A) = E(S) - \sum_{v \in V(A)} \frac{S_v}{S} E(S_v) \tag{4}$$

## 4. Empirical studies

#### 4.1. Description of dataset

A cross-sectional study was conducted to assess generalized anxiety disorder (GAD) levels during the COVID-19 pandemic in Saudi Arabia. The data collection took place during the full lockdown from May 11 to May 26, 2020. The researchers adopted the GAD-7, which had been proven as a valid and efficient tool for screening for GAD. The current survey consisted of questions to identify demographic information and potential factors associated with anxiety levels, followed by the seven questions of the GAD-7 tool. The survey was developed using an online Question pro questionnaire. The survey was initially distributed through Sharek Health, an organization that aids in data collection in all Saudi Arabian regions, followed by a snowball sampling strategy to increase

**Table 2**  
Survey questions.

| Variable                 | Label  |
|--------------------------|--|
| Q3                       | Nationality  |
| Q18                      | Gender   |
| Q19                      | Age  |
| Q20                      | Marital status   |
| Q21                      | How many people are in the house? (Includes house workers and drivers)   |
| Q22                      | Are you or any of your household members at increased risk of contracting the coronavirus? (This includes anyone over the age of 60 or pregnant or having comorbidities) |
| Q24A1                    | Have you been tested positive for COVID-19 test?   |
| Q24A2                    | Have you been suspected of carrying the coronavirus?   |
| Q24A3                    | Have any member of your family have been diagnosed with coronavirus?   |
| Q25                      | Qualification  |
| Q26                      | Occupation   |
| Q28                      | What is the method followed by your employer, or academic institution during the pandemic? (Online or in person)   |
| Q30                      | Feeling nervous, anxious, or on edge   |
| Q31                      | Not being able to stop or control worrying   |
| Q32                      | Worrying too much about different things   |
| Q33                      | Trouble relaxing   |
| Q34                      | Being so restless that it's hard to sit still  |
| Q35                      | Becoming easily annoyed or irritable   |
| Q36                      | Feeling afraid as if something awful might happen  |
| Q37                      | How difficult have these problems made it to do work, take care of things at home, or get along with other people  |
| Georgian                 | Geographically region  |
| Anxiety (Two category)   | Anxiety two categories (Anxious and non-anxious)   |
| Anxiety (Three category) | Anxiety score three categories (Mild-Moderate-Severe)  |

**Table 3**  
Statistical analysis of the dataset.

| Attributes | Mean  | Median | Standard Deviation | Max. | Min. |
|------------|-------|--------|--------------------|------|------|
| Q3         | 1.063 | 1      | 0.242              | 2    | 1    |
| Q18        | 1.560 | 2      | 0.496              | 2    | 1    |
| Q19        | 3.307 | 3      | 1.300              | 6    | 1    |
| Q20        | 1.731 | 2      | 0.560              | 4    | 1    |
| Q21        | 6.733 | 7      | 3.026              | 30   | 0    |
| Q20        | 1.651 | 2      | 0.477              | 2    | 1    |
| Q24A1      | 0.002 | 0      | 0.048              | 1    | 0    |
| Q24A2      | 0.006 | 0      | 0.075              | 1    | 0    |
| Q24A3      | 0.010 | 0      | 0.099              | 1    | 0    |
| Q25        | 3.731 | 4      | 0.954              | 5    | 1    |
| Q26        | 2.730 | 2      | 1.487              | 6    | 1    |
| Q30        | 1.056 | 1      | 1.046              | 3    | 0    |
| Q31        | 0.638 | 0      | 0.910              | 3    | 0    |
| Q32        | 0.930 | 1      | 0.982              | 3    | 0    |
| Q33        | 0.700 | 0      | 0.941              | 3    | 0    |
| Q34        | 0.754 | 0      | 0.976              | 3    | 0    |
| Q35        | 0.768 | 0      | 0.966              | 3    | 0    |
| Q36        | 0.627 | 0      | 0.900              | 3    | 0    |
| Q37        | 1.696 | 2      | 0.712              | 4    | 1    |
| Geo-region | 1.022 | 1      | 0.989              | 4    | 0    |

the number of participants. The survey was shared via different social media platforms, including Twitter and WhatsApp.

The sample included 3017 participants who had completed the survey questionnaire with no missing data, as participants were required to complete all the questions. Some questions were selected from the survey to be included in this study as seen in Table 2, the remaining questions were omitted due to their irrelevance to this study. One-third (33%) were males, more than half were between the age group 20–39 years (n = 1689, 56%) and married (n = 989, 63.7%).

#### 4.2. Statistical analysis

In this study, the main statistical analysis methods were used to

**Table 4**  
Correlation between each Attribute and the First Experiment Target Attribute.

| Attributes | Target Attribute         | Correlation coefficient |
|------------|--------------------------|-------------------------|
| Q31        | Anxiety Two category (2) | 0.69032                 |
| Q32        | Anxiety Two category     | 0.68472                 |
| Q30        | Anxiety Two category     | 0.68466                 |
| Q33        | Anxiety Two category     | 0.67673                 |
| Q36        | Anxiety Two category     | 0.65965                 |
| Q35        | Anxiety Two category     | 0.58508                 |
| Q34        | Anxiety Two category     | 0.54546                 |
| Q37        | Anxiety Two category     | 0.48791                 |
| Q19        | Anxiety Two category     | 0.14877                 |
| Q22        | Anxiety Two category     | 0.11936                 |
| Q26        | Anxiety Two category     | 0.09987                 |
| Q20        | Anxiety Two category     | 0.08589                 |
| Q18        | Anxiety Two category     | 0.06622                 |
| Q24A2      | Anxiety Two category     | 0.05201                 |
| Georegion  | Anxiety Two category     | 0.05052                 |
| Q3         | Anxiety Two category     | 0.02726                 |
| Q24A3      | Anxiety Two category     | 0.02619                 |
| Q21        | Anxiety Two category     | 0.01486                 |
| Q25        | Anxiety Two category     | 0.01195                 |
| Q24A1      | Anxiety Two category     | 0.00648                 |

**Table 5**  
Correlation between each Attribute and the Second Experiment Target Attribute.

| Attributes | Target Attribute       | Correlation coefficient |
|------------|------------------------|-------------------------|
| Q31        | Anxiety Three category | 0.64316                 |
| Q30        | Anxiety Three category | 0.63942                 |
| Q32        | Anxiety Three category | 0.63888                 |
| Q33        | Anxiety Three category | 0.63119                 |
| Q36        | Anxiety Three category | 0.61451                 |
| Q35        | Anxiety Three category | 0.54564                 |
| Q34        | Anxiety Three category | 0.50835                 |
| Q37        | Anxiety Three category | 0.45479                 |
| Q19        | Anxiety Three category | 0.13954                 |
| Q22        | Anxiety Three category | 0.11146                 |
| Q26        | Anxiety Three category | 0.09348                 |
| Q20        | Anxiety Three category | 0.08045                 |
| Q18        | Anxiety Three category | 0.06233                 |
| Q24A2      | Anxiety Three category | 0.04852                 |
| Georegion  | Anxiety Three category | 0.04767                 |
| Q3         | Anxiety Three category | 0.02526                 |
| Q24A3      | Anxiety Three category | 0.02427                 |
| Q21        | Anxiety Three category | 0.01428                 |
| Q25        | Anxiety Three category | 0.01328                 |
| Q24A1      | Anxiety Three category | 0.00807                 |

analyze the dataset attributes. Before analyzing the dataset, three attributes were removed, which are sector, whether the participant was learning/working online or in person (Q28), and anxiety score. The mean, median, standard deviation, maximum, and minimum values of the numerical attributes were calculated and recorded in Table 3. Furthermore, the correlation coefficient between each attribute and the target class was computed, and the values were ranked in descending order, as shown in Tables 4 and 5.

#### 4.3. Experimental setup

This experiment was carried out using open-source software, called Weka, that affords machine learning algorithms to build an anxiety prediction model. The dataset was used to classify two-class and three-class anxiety problems. The attributes “sector” and “Q28” were excluded from both experiments since they contain missing values that can negatively affect the classification accuracy. The attribute “Anxiety score 1” was also removed from both experiments, as it can directly contribute to classifying anxiety in patients on its own. Additionally, the target class of the first experiment was excluded from the second and vice versa. Afterward, the nominal features were converted to numerical using Excel software.

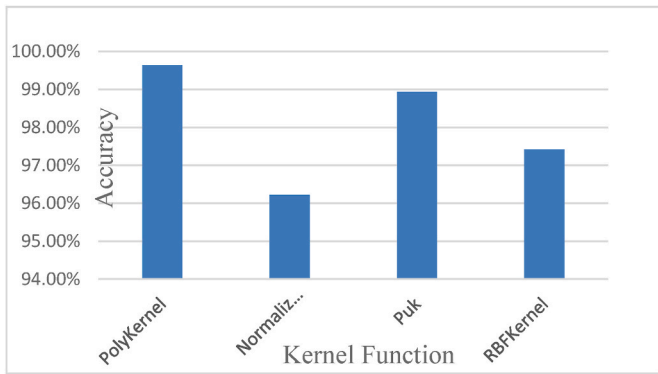


Fig. 3. Tuning Kernel function.

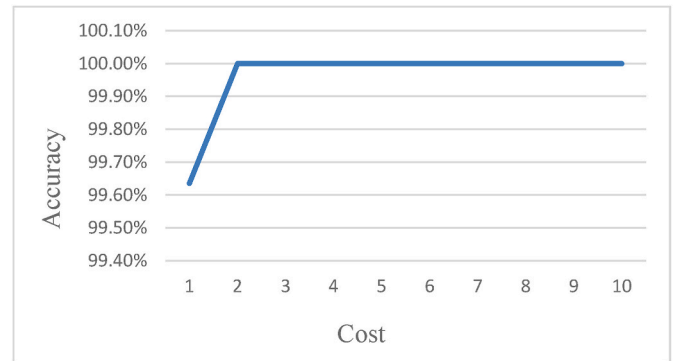


Fig. 4. Tuning the cost.

Two supervised machine learning algorithms were employed to build the models in both experiments: SVM and J48 decision tree. Hyperparameter tuning was performed to optimize the classifiers. The Correlation Ranking Filter provided by Weka called ‘CorrelationAttributeEval’ was utilized to obtain the best feature subset that results in attaining the highest average accuracy for both experiments, that are anxiety two-class and three-class. Then, 10-fold cross-validation was used to partition the dataset and evaluate the accuracy. Furthermore, to determine the best models for classifying anxiety, confusion matrices were constructed to compare the accuracy, recall, precision, and the f–Measure of the proposed models.

4.4. Performance measure

Four primary performance measures were utilized in this study: classification accuracy, f–Measure, precision, and recall. Equation (5) shows how the classification accuracy that is responsible for calculating the precisely classified instances is calculated. Equations (6) and (7) show how precision and recall are computed, where precision calculates the amount of true positive prediction belonging to the positive class, and recall calculates true positive prediction belonging to all positive samples. Equation (8) shows how the f–Measure is calculated, which estimates the performance of each class [31].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{6}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{7}$$

$$f\Delta Measure = \frac{2 \times Precision \times Recall}{(Precision + Recall)} \tag{8}$$

In the above equations, TP denotes true positive, TF denotes true negative, whereas FP denotes false positive, and FN denotes false negative.

4.5. Optimization strategy

Developing an optimum model for medical applications is essential to avoid further complications. Therefore, parameter tuning is a crucial step that must be performed effectively. Regarding the SVM hyperparameters for classifying the two-class and three-class anxiety problems, only the kernel function and cost (C) were tuned, as the other hyperparameters did not positively affect the accuracy. In SVM, the cost hyperparameter was fixed to its default value in Weka, which is 1, and the kernel functions (Poly Kernel, Normalized Poly Kernel, PUK, and RBF Kernel) were individually experimented. The kernel function that

Table 6

Optimum hyperparameters for the proposed SVM model.

| Parameters | Optimal value chosen |
|------------|----------------------|
| Kernel     | Poly Kernel          |
| C          | 2                    |
| Epsilon    | 1.0E-12              |



Fig. 5. Tuning the confidence factor.

Table 7

Optimum hyperparameters for the proposed J48 model.

| Parameters        | Optimal value chosen |
|-------------------|----------------------|
| Confidence Factor | 0.45                 |
| MinNumObj         | 2                    |

achieved the highest accuracy was then tried on a cost range from 1 to 10 to gain the optimum accuracy. For the J48 decision tree, only the confidence factor hyperparameter was altered within a range from 0.15 to 0.95.

4.5.1. Two-class anxiety classification

Figs. 3 and 4 show the results of manipulating SVM’s kernel functions and costs to find the optimum binary-class classification accuracy, where the poly kernel with cost 2 to 10 achieved the best result. The cost value 2 was chosen as the optimal hyperparameter.

Table 6 shows the best hyperparameter combinations of SVM that achieved an accuracy of 100% when applied to the whole dataset for binary-class classification.

Fig. 5 shows the results of adjusting J48’s confidence factor with different values, where the confidence factor 0.45 achieved the best

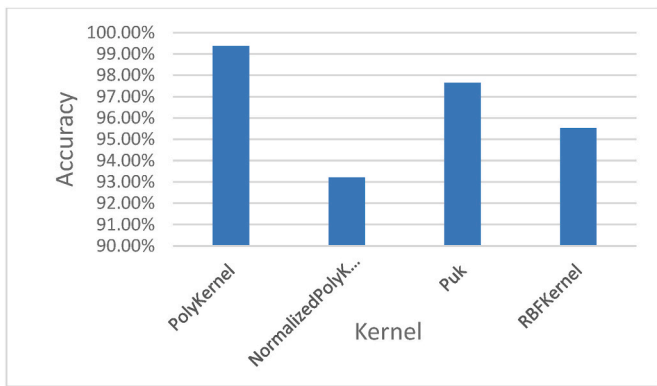


Fig. 6. Optimizing Kernel functions.

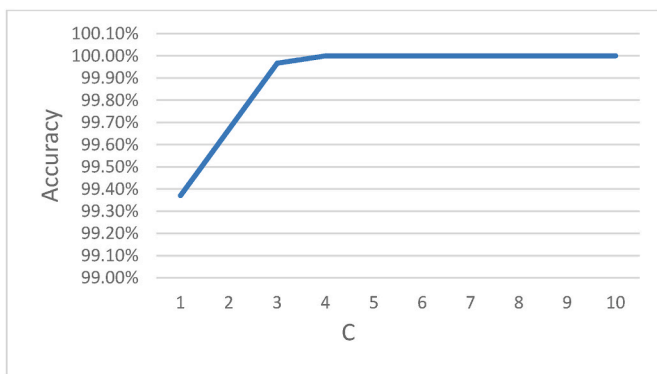


Fig. 7. Optimizing the cost.

Table 8 Optimum hyperparameters for the proposed SVM model.

| Parameters | Optimal value chosen |
|------------|----------------------|
| Kernel     | Poly Kernel          |
| C          | 4                    |
| Epsilon    | 1.0E-12              |



Fig. 8. Tuning the confidence factor.

outcome for classifying the two-class problem.

Table 7 shows the best hyperparameter combinations of the J48 decision tree that achieved an accuracy of 95.79% when applied to the whole dataset when classifying the two-class problem.

#### 4.5.2. Three-class anxiety classification

Figs. 6 and 7 show the results of manipulating SVM’s kernel

Table 9 Optimum hyperparameters for the proposed J48 model.

| Parameters        | Optimal value chosen |
|-------------------|----------------------|
| Confidence Factor | 0.15                 |
| MinNumObj         | 2                    |

Table 10

Average accuracy of different feature subsets of the two-class classification experiment.

| Number of features | Accuracy of SVM | Accuracy of J48 | Average accuracy of each set of features |
|--------------------|-----------------|-----------------|--|
| Using 20 Features  | 100%            | 95.79%          | 97.90%                                   |
| Using 10 Features  | 100%            | 95.96%          | 97.98%                                   |
| Using 5 Features   | 95.76%          | 95.00%          | 95.38%                                   |
| Using 3 Features   | 92.97%          | 93.27%          | 93.12%                                   |
| Using 2 Features   | 91.95%          | 91.51%          | 91.73%                                   |
| Using 1 Feature    | 90.19%          | 90.19%          | 90.19%                                   |

functions and costs to find the optimum three-class classification accuracy, where the poly kernel with costs 4 to 10 achieved the best result. The cost value 4 was chosen as the optimal hyperparameter.

Table 8 shows the best hyperparameter combinations of SVM that achieved an accuracy of 100% when applied to the whole dataset for classifying the three-class problem.

For the J48 decision tree, Fig. 8 shows that the confidence factor 0.15 achieved the best result when classifying the three-class problem.

Table 9 shows the best hyperparameter combinations of the J48 decision tree that achieved an accuracy of 92.81% when applied to the whole dataset for classifying the three-class problem.

## 5. Results and discussion

The parameter tuning succeeded in promoting SVM’s classification accuracy to 100% in both two-class and three-class classification problems. It also enhanced the accuracy of J48 to 95.79% for the two-class problem and 92.81% for the three-class problem. After performing feature selection, the J48’s classifier accuracy was further enhanced using the best feature subset that offered the best performance. Considering SVM, which already achieved an accuracy of 100% after parameter tuning, feature selection is applied to it to gain the same accuracy with fewer features. This facilitates the process of classifying anxiety for medical teams, as they will need to collect fewer attributes from patients.

This section will discuss the results of classifying the binary-class and three-class anxiety problems after performing feature selection using the performance measures listed previously. The results were evaluated using 10-fold cross-validation.

### 5.1. Feature selection

The ‘CorrelationAttributeEval’ tool in Weka was employed to the whole dataset to rank the attributes based on their correlation to the target attribute in descending order, as shown previously in Tables 2 and 3. A recursive feature elimination procedure was applied to divide the features in half in each iteration until a single feature remains. The highest correlated V/2 features were further experimented with, whereas the lowest correlated V/2 features were discarded.



**Table 11**

Average accuracy of different feature subsets of the three-class classification experiment.

| Number of features | Accuracy of SVM | Accuracy of J48 | The average accuracy of each set of features |
|--------------------|-----------------|-----------------|--|
| Using 20 Features  | 100%            | 92.81%          | 96.40%                                       |
| Using 10 Features  | 100%            | 93.50%          | 96.75%                                       |
| Using 5 Features   | 93.14%          | 91.48%          | 92.31%                                       |
| Using 3 Features   | 89.63%          | 89.96%          | 89.79%                                       |
| Using 2 Features   | 87.11%          | 88.66%          | 87.89%                                       |
| Using 1 Feature    | 85.18%          | 86.77%          | 85.98%                                       |

**Table 12**

Results of classifiers after optimization and feature selection of the two-class classification experiment.

| Performance Measure | SVM | J48    |
|---------------------|-----|--------|
| Accuracy (%)        | 100 | 95.96% |
| Precision           | 1   | 0.974  |
| Recall              | 1   | 0.975  |
| f-measure           | 1   | 0.975  |

**Table 13**

SVM Confusion matrix after Optimization and Feature Selection of the Two-class Classification Experiment.

|        |             | Predicted |             |
|--------|-------------|-----------|-------------|
|        |             | Anxiety   | Non-Anxiety |
| Actual | Anxiety     | 2425 (TP) | 0 (FN)      |
|        | Non-Anxiety | 0 (FP)    | 592 (TN)    |

5.1.1. Two-class anxiety classification

As shown in Table 10, the highest average accuracy achieved after classifying the two-class problem is 97.98%, where 10 features were used. The top 10 features include Q31, Q32, Q30, Q33, Q36, Q35, Q34, Q37, Q19, and Q22 in descending order.

5.1.2. Three-class anxiety classification

As shown in Table 11, the highest average accuracy achieved after classifying the three-class problem is 96.75%, where 10 features were used. The top 10 features include Q31, Q30, Q32, Q33, Q36, Q35, Q34, Q37, Q19, and Q22 in descending order. It is concluded that the best feature subset in both models is the same but with a slight difference in the order.

5.2.1. Results of the two-class anxiety classification

Table 12 compares the performance of the classifiers SVM and J48 after parameter tuning and feature selection for classifying the two-class problem. The classifiers' performance is evaluated according to the performance measures mentioned earlier. As shown in Table 10, SVM achieved the most reliable performance with a classification accuracy of 100%, whereas J48 achieved 95.96% (see Table 13).

Tables 13 and 14 present the confusion matrix of the classifiers SVM and J48. Since the experiment is based on medical diagnosis, the FN rate is the most significant evaluator, as undiagnosed anxiety may cause insomnia (sleep disorder) and mental trouble. SVM succeeded in achieving a 0 FN rate, whereas J48 possessed a 60 FN rate. Therefore, it is concluded that SVM is more powerful for classifying the two-class anxiety problem than the J48 Decision Tree.

**Table 14**

J48 confusion matrix after optimization and feature selection of the two-class classification experiment.

|        |             | Predicted |             |
|--------|-------------|-----------|-------------|
|        |             | Anxiety   | Non-Anxiety |
| Actual | Anxiety     | 2365 (TP) | 60 (FN)     |
|        | Non-Anxiety | 62 (FP)   | 530 (TN)    |

**Table 15**

Results of classifiers after optimization and feature selection of the three-class classification experiment.

| Performance Measure | SVM | J48    |
|---------------------|-----|--------|
| Accuracy (%)        | 100 | 93.50% |
| Precision           | 1   | 0.933  |
| Recall              | 1   | 0.935  |
| f-measure           | 1   | 0.934  |

**Table 16**

SVM Confusion matrix after Optimization and Feature Selection of the Three-class Classification Experiment.

|        |          | Predicted |          |        |
|--------|----------|-----------|----------|--------|
|        |          | Mild      | Moderate | Severe |
| Actual | Mild     | 2425      | 0        | 0      |
|        | Moderate | 0         | 247      | 0      |
|        | Severe   | 0         | 0        | 345    |

**Table 17**

J48 confusion matrix after optimization and feature selection of the three-class classification experiment.

|        |          | Predicted |          |        |
|--------|----------|-----------|----------|--------|
|        |          | Mild      | Moderate | Severe |
| Actual | Mild     | 2375      | 0        | 76     |
|        | Moderate | 0         | 211      | 34     |
|        | Severe   | 50        | 36       | 235    |

**Table 18**

J48 TP, FP, FN, and TN rates of the Three-class Classification Experiment.

|      |    | Class |          |        |
|------|----|-------|----------|--------|
|      |    | Mild  | Moderate | Severe |
| Rate | TP | 2375  | 211      | 235    |
|      | FP | 50    | 36       | 110    |
|      | FN | 76    | 34       | 86     |
|      | TN | 516   | 2736     | 2586   |

5.2.2. Results of the three-class anxiety classification

Table 15 compares the performance of the classifiers SVM and J48 after parameter tuning and feature selection for classifying the three-class problem. The classifiers' performance is evaluated according to the performance measures mentioned earlier. As shown in Table 14, SVM achieved the most reliable performance with a classification accuracy of 100%, whereas J48 achieved an overall accuracy of 93.50% (see Table 16).

Tables 16 and 17 present the confusion matrix of the classifiers SVM and J48 for classifying the three-class anxiety problem (see Table 17). Unlike binary classification, where the TP, TN, FP, FN values can be viewed clearly from the tables, it must be calculated for easier interpretation in multi-class classification. As shown in Table 15, SVM succeeded in achieving 0 rates of false predictions. Table 18 presents the TP, TN, FP, FN for each class for the J48 classifier. As shown, J48 possessed a

**Table 19**  
Comparing the accuracies of classifiers in 2-class and 3-class Experiments.

| Classifier | Anxiety Two-class | Anxiety Three-class |
|------------|-------------------|---------------------|
| SVM        | 100%              | 100%                |
| J48        | 95.96%            | 93.50%              |

196 FN rate. Therefore, it is also concluded that SVM is more powerful for classifying three-class anxiety problems than J48.

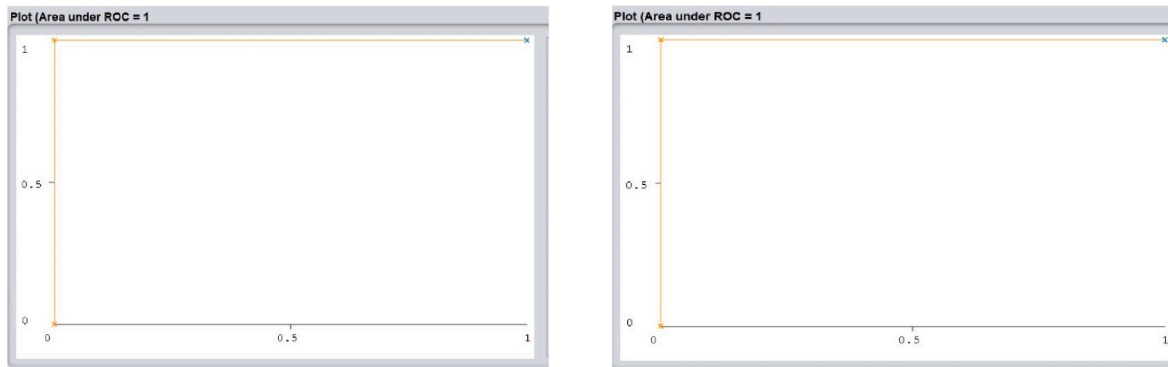
**5.3. Comparing the achieved result for classifying two-class and three-class anxiety problems**

From the experiment results stated above, it is concluded that SVM

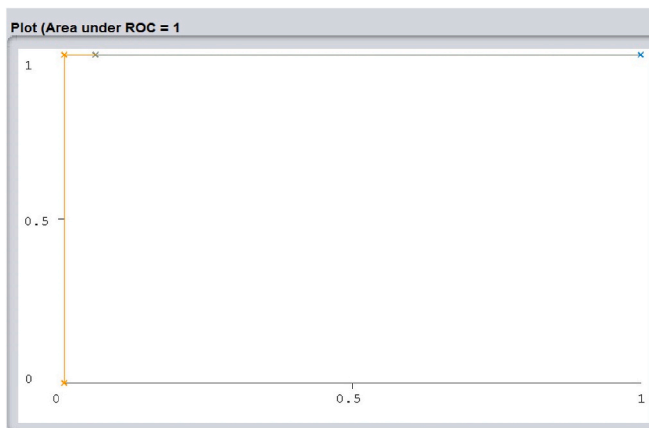
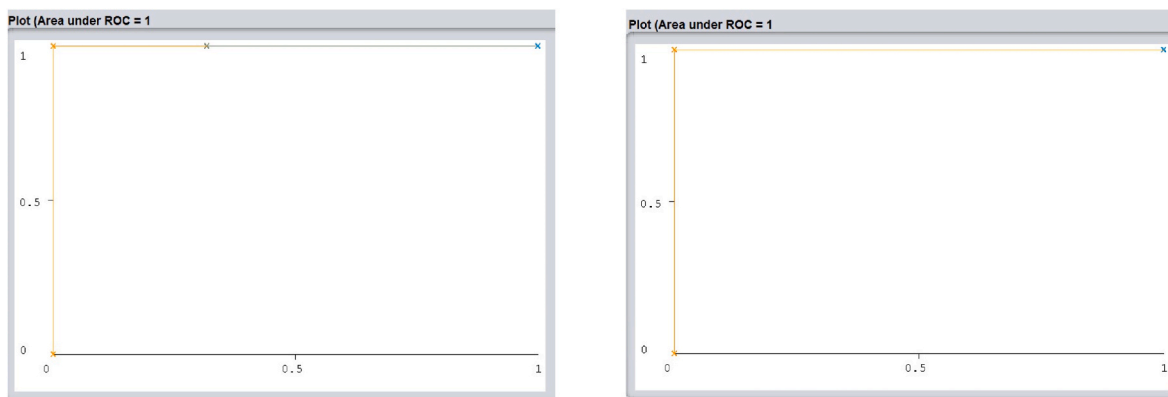
succeeded in classifying all test cases of both two-class and three-class problems. In contrast, J48 achieved an accuracy of 95.96% for two-class classification and 93.50% for three-class classification, which indicates that J48 performed better in the two-class classification experiment. The reason behind the outperformance of J48 in the two-class experiment is the slight difference in the correlation coefficient between the attributes and the two-class problem compared to the three-class problem listed in [Tables 3 and 4](#) [Table 19](#) compares the accuracies of the classifiers in classifying two-class and three-class experiments.

**5.4. Further discussions**

This paper aims to predict anxiety using machine learning techniques to study the pandemic’s impact on Saudi Arabia’s society.



**Fig. 9.** SVM Roc curve for classifying two-class problem: (a) Class zero (b) class one.



**Fig. 10.** SVM Roc curve for classifying three-class problem: (a) Class zero (b) class one (c) class two.

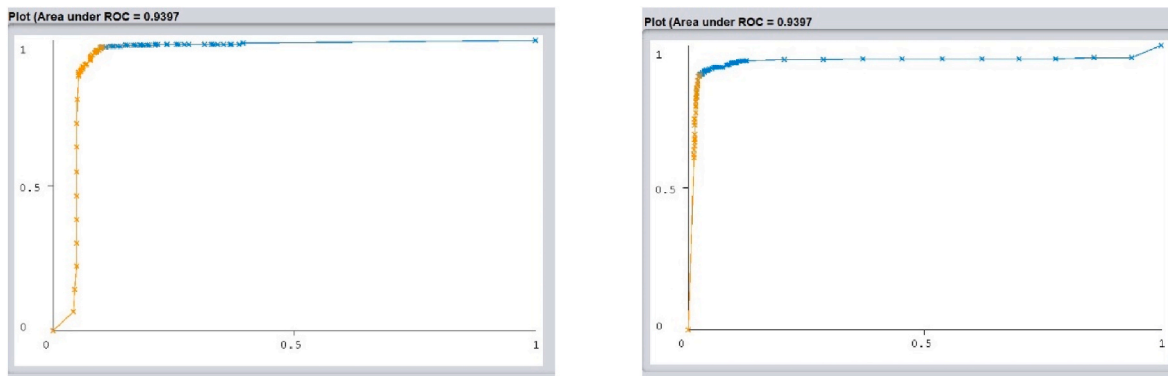


Fig. 11. J48 Roc curve for classifying two-class problem: (a) Class zero (b) Class one.

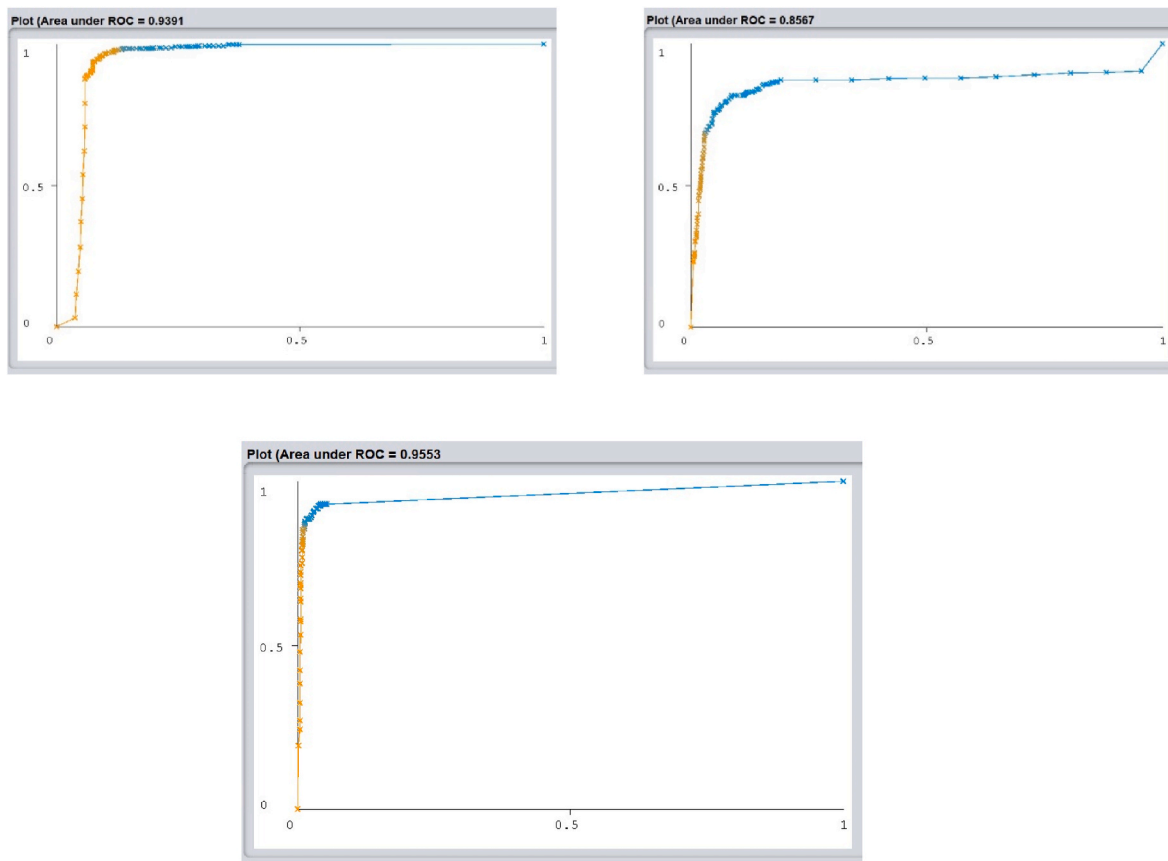


Fig. 12. J48 Roc curve for classifying three-class problem: (a) Class zero (b) Class one (c) Class two.

According to the previously discussed tables and figures, SVM outperformed the J48 Decision Tree attaining accuracy, precision, recall, and f-measure of 100%, 1.0, 1.0, and 1.0, respectively, in classifying both two-class and three-class problems. For further evaluation, the Area Under the Receiver Operating Characteristics (AUROC) was constructed to measure various confusion matrices that each threshold provided. Figs. 9 and 10 show the AUROC curve of SVM in classifying the two-class and three-class problems. From the figures below, it is concluded that SVM succeeded in providing a perfect prediction reaching an AUROC value of 1.0 in classifying two-class and three-class anxiety problems.

Figs. 11 and 12 show the AUROC curve of the J48 Decision Tree in classifying the two-class and three-class problems. From the figures below, it is concluded that the J48 performed better in classifying the two-class than the three-class problem reaching an AUROC value of

0.9397 against an average AUROC of 0.9170. The outcomes support the fact that increasing the number of output variable classes increases the complexity of the model, making it difficult to get good results. Hence, it is usually better to have fewer classes in the output variable to achieve better results.

### 6. Conclusion and recommendation

A Saudi Arabian dataset was utilized for the first time in this study to build a prediction model that categorizes two categories and three categories of anxiety during the COVID-19 pandemic. The authors utilized two classifiers, namely, the Support Vector Machine (SVM) and the J48 Decision Tree, due to their reliable outcomes in medical-related data. The optimal hyperparameters were obtained, and the effect of feature selection was examined to build the model with a reduced feature

subset. The empirical results attested to the fact that SVM outperformed the J48 Decision Tree with 100% accuracy against 95.96% for the three-class problem and 93.50% for the two-class problem when predicting anxiety for earlier diagnosis and timely intervention using ten features. Therefore, the researchers recommend that decision maker in Saudi Arabia adopt the prediction model produced by this study to strategically plan the distribution of both preventative and curative mental health care services.

### Declaration of competing interest

The research was funded by Imam Abdulrahman Bin Faisal University, the grant number is Covid19-2020-024-CAMS.

### Acknowledgement

The authors would like to acknowledge the participants who participated in the study.

### References

- Ingram J, Hand CJ, Maciejewski G. Social isolation during COVID-19 lockdown impairs cognitive function. *Appl Cognit Psychol* 2021;35(4):935–47. <https://doi.org/10.1002/acp.3821>.
- Mahase E. Covid-19: mental health consequences of pandemic need urgent research, paper advises. *Br Med J* 2020;369:m1515. <https://doi.org/10.1136/bmj.m1515>.
- Pfefferbaum B, North CS. Mental health and the covid-19 pandemic. *N Engl J Med* 2020. <https://doi.org/10.1056/NEJMp2008017>.
- The Lancet Psychiatry. Mental health and COVID-19: change the conversation. *Lancet Psychiatr* 2020;7(6):463. [https://doi.org/10.1016/S2215-0366\(20\)30194-2](https://doi.org/10.1016/S2215-0366(20)30194-2).
- Dagklis T, Tsakiridis I, Mamopoulos A, Athanasiadis A, Pearson R, Papazisis G. Impact of the COVID-19 lockdown on antenatal mental health in Greece. *Psychiatr Clin Neurosci* 2020;74(11):616–7. <https://doi.org/10.1111/pcn.13135>.
- Jha IP, Awasthi R, Kumar A, Kumar V, Sethi T. Learning the mental health impact of COVID-19 in the United States with explainable artificial intelligence. *medRxiv* 2020. <https://doi.org/10.1101/2020.07.19.20157164>. 07.19.20157164, 2020.
- The Anxiety and Depression Association of America. Facts & statistics | anxiety and depression association of America. 2020. ADAA, <https://adaa.org/understanding-anxiety/facts-statistics>.
- Thompson A, Issakidis C, Hunt C. Delay to seek treatment for anxiety and mood disorders in an Australian clinical sample. *Behav Change* 2008;25(2):71–84. <https://doi.org/10.1375/bech.25.2.71>.
- Albagmi F, AlNujaidi H, AlShawan D. Anxiety levels amid the COVID-19 lockdown in Saudi Arabia. *Int J Gen Med* 2021.
- Yu JS, Xue AY, Redei EE, Bagheri N. A support vector machine model provides an accurate transcript-level-based diagnostic for major depressive disorder. *Transl Psychiatry* 2016;6(10). <https://doi.org/10.1038/tp.2016.198>. e931–e931.
- Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, Ho RC. Immediate psychological Responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *Int J Environ Res Publ Health* 2020;17(5). <https://doi.org/10.3390/ijerph17051729>.
- Alyami HS, Naser AY, Dahmash EZ, Alyami MH, Alyami MS. Depression and anxiety during the COVID-19 pandemic in Saudi Arabia: a cross-sectional study. *Int J Clin Pract* 2021:e14244. <https://doi.org/10.1111/ijcp.14244>. n/a(n/a).
- Khoshaim HB, Al-Sukayt A, Chinna K, Nurunnabi M, Sundarasan S, Kamaludin K, Baloch GM, Hossain SFA. Anxiety level of university students during COVID-19 in Saudi Arabia. *Front Psychiatr* 2020;11. <https://doi.org/10.3389/fpsy.2020.579750>.
- Pintelas EG, Kotsilieris T, Livieris IE, Pintelas P. A review of machine learning prediction methods for anxiety disorders. In: *Proceedings of the 8th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-Exclusion*; 2018. p. 8–15. <https://doi.org/10.1145/3218585.3218587>.
- Boeke EA, Holmes AJ, Phelps EA. Toward robust anxiety biomarkers: a machine learning approach in a large-scale sample. *Biol Psychiatr: Cognitive Neurosci. Neuroimaging* 2020;5(8):799–807. <https://doi.org/10.1016/j.bpsc.2019.05.018>.
- Husain W, Xin LK, Rashid NA, Jothi N. Predicting generalized anxiety disorder among women using random forest approach. 2016 3rd international conference on Computer and information sciences (ICCOINS), 37–42. 2016. <https://doi.org/10.1109/ICCOINS.2016.7783185>.
- Elhai JD, Yang H, McKay D, Asmundson GJG, Montag C. Modeling anxiety and fear of COVID-19 using machine learning in a sample of Chinese adults: associations with psychopathology, sociodemographic, and exposure variables. *Hist Philos Logic* 2021;34(2):130–44. <https://doi.org/10.1080/10615806.2021.1878158>.
- Nemesure MD, Heinz MV, Huang R, Jacobson NC. Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. *Sci Rep* 2021;11(1):1980. <https://doi.org/10.1038/s41598-021-81368-4>.
- Jothi N, Husain W, Rashid NA. Predicting generalized anxiety disorder among women using Shapley value. *J. Infect. Public Health* 2021;14(1):103–8. <https://doi.org/10.1016/j.jiph.2020.02.042>.
- Kumar Prince, Garga Shruti, Garg Ashwani. Assessment of anxiety, depression and stress using machine learning models. *Procedia Comput Sci* 2020;171:1989. <https://doi.org/10.1016/j.procs.2020.04.213>. 1998.
- Sau A, Bhakta I. Screening of anxiety and depression among seafarers using machine learning technology. *Inf. Med. Unlock*. 2019;16:100228. <https://doi.org/10.1016/j.imu.2019.100228>.
- Shen JH, Rudzicz F. Detecting anxiety through reddit. *Proceedings of the fourth workshop on computational linguistics and clinical psychology — from linguistic signal to clinical reality*, vols. 58–65; 2017. <https://doi.org/10.18653/v1/W17-3107>.
- Brereton RG, Lloyd GR. Support vector machines for classification and regression. *Analyst* 2010;135(2):230–67. <https://doi.org/10.1039/B918972F>.
- Drucker H, Kaufman L, Smola AJ, Drucker H, Burges CJC, Smola A, Vapnik V. *Support vector regression machines queueing view project smoothing based on data view project support vector regression machines*, vol. 2; 1997.
- Kononenko I, Kukar M. *Machine learning and data mining*. Elsevier Science; 2007.
- Ciobanu, D. (n.d.). *Mathematical and quantitative methods using SVM for classification 1 introduction*. 8(13), 207–222.
- C4.5: Programs for Machine Learning—J. Ross Quinlan—Google Books (n.d.). Retrieved June 4, 2021, from +J.R.:+C4.+5:+Programs+for+Machine+Learning.+Elsevier+(2014)&ots=sR3oZLBp3&sig=VSxWkObMqzyH6HenEH-saiZWICw&redir\_esc=y#v=onepage&q=Quinlan%2C J.R.%3A C4. 5%3A Programs for Machine Learning. Elsevier (2014)&f=false, <https://books.google.com.sa/books?hl=en&lr=&id=b3ujBQAQBAJ&oi=fnd&pg=PP1&dq=Quinlan>.
- Alam F, Pachauri S. Comparative study of J48 , naive bayes and one-R classification technique for credit card fraud detection using WEKA. *Adv Comput Sci Technol* 2017;10(6):1731–43.
- Bienvenido-Huertas D, Nieto-Julián JE, Moyano JJ, Macías-Bernal JM, Castro J. Implementing artificial intelligence in H-BIM using the J48 algorithm to manage historic buildings. *Int J Architect Herit* 2020;14(8):1148–60. <https://doi.org/10.1080/15583058.2019.1589602>.
- Zhao Y, Zhang Y. Comparison of decision tree methods for finding active objects. *Adv Space Res* 2008;41(12):1955–9. <https://doi.org/10.1016/j.asr.2007.07.020>.
- Olatunji SO, Alotaibi S, Almutairi E, Alrabae Z, Almajid Y, Altabee R, Altassan M, Basheer Ahmed MI, Farooqui M, Alhiyafi J. Early diagnosis of thyroid cancer diseases using computational intelligence techniques: a case study of a Saudi Arabian dataset. *Comput Biol Med* 2021;131(Febuary):104267. <https://doi.org/10.1016/j.combiomed.2021.104267>.