International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Emotional Disorders Detection Using Machine Learning Algorithm

Callista Ivana Mogie[1 2]*, Tuga Mauritsius[3]

Submitted: 25/01/2024 **Revised**: 04/03/2024 **Accepted**: 11/03/2024

Abstract: Emotional disorders, namely anxiety and depression are the most debilitating mental illness. Early detection will minimize risks of developing complex disorders and suicidality. This study aimed to build and evaluate machine learning classification models in screening for anxiety, depression, and healthy control. Supervised machine learning algorithms including Random Forest, Artificial Neural Network, Support Vector Machine, and Naive Bayes were applied and compared to build superior models. The best algorithm for multiclass classification was Artificial Neural Network with an F1-score of 0.97. Additionally, for binary classification, the Support Vector Machine yielded the highest performance for both the 'Depression and Anxiety' class (F1-score: 0.99) and the 'Depression' class (F1-score: 0.98). For the 'Anxiety' class, the Artificial Neural Network exhibited the best performance with an F1-score of 0.99, while the Random Forest algorithm achieved the highest F1-score of 0.98 for the 'Healthy' class. These findings hold potential in assisting clinicians by providing more efficient treatment strategies.

Keywords: *Anxiety, Classification, Depression, Mental Health, Supervised machine learning*

1. Introduction

Emotional disorder is marked by frequent intense negative emotions, coupled with a diminished sense of control and efforts to avoid or dampen those emotions [1]. Anxiety and depression are grouped as emotional disorders due to their core disruption of feelings such as excessive fear and despair, diminishing emotional well-being, functioning, and overall life quality [1-3]. Anxiety is characterized by extreme fear along with avoiding situations that can trigger anxiousness, whereas depression is marked by chronic melancholy and anhedonia [1]. Despite the distinct symptoms, those disorders commonly co-occur [4].

As the most prevalent global mental illnesses, depressive and anxiety disorders were the second (5.6%) and sixth (3.4%) contributors to the global burden of years lived with disability, respectively [5]. Moreover, around 25% cases have spiked since COVID-19 pandemic [6]. Anxiety and depression can also be sign or may forms comorbidity of another mental disorder, such as Attention deficit hyperactivity disorder (ADHD), Borderline Personality Disorder (BPD), Bipolar, Post-Traumatic Stress Disorder (PTSD), Obsessive Compulsive Disorder (OCD) [7]. The disorders increase suicide risk, which an action which intentionally causing one's own death. It is one of depression disorder symptoms, and not uncommon to anxiety disorders, such as separation anxiety disorder, even 60% higher to people with specific phobia [2].

Nevertheless, barriers such as high diagnosis costs, scarcity of mental health professionals, geographic constraints, and the potential for misdiagnosis, discouraged individuals from being diagnosed [8-12]. Simultaneously, late detection will delay treatment, exacerbate symptom severity and comorbidity, and increase the risk of suicide [13]. Therefore, early detection is necessary to provide early intervention.

Machine learning (ML), a field within Artificial Intelligence, has gained exposure in the digital realm as an essential element of digitalization solutions [14]. It has been extensively utilized to predict and identify disease, including mental illnesses such as schizophrenia, PTSD, stress, anxiety, and depression using supervised learning algorithms [15]. In supervised learning, the machine learns from labeled data.

Mental disorders classification using ML has been a hot topic. Most prior works focused on single classification, such as classifying one mental illness, disregarding the coexistence of both disorders. For instance, solely classify depression disorder [16]. When in fact, there is a silver lining that can cause these disorders to overlap. Depression is caused by cognitive bias that contributes to ruminating (worrying things in the past), whereas anxiety is future-based excessive worrying [17]. Hence, it is necessary to detect both disorders as well. The screening result can be beneficial in personalized treatment and quicker diagnosis [18]

Integrated with ML, data mining holds a pivotal role in healthcare predictive analytics [19]. In building a reliable ML model, data mining technique is essential. Cross-Industry Standard Process for Data Mining (CRISP-DM) is a data mining framework, applicable to diverse domains, albeit still uncommonly applied in mental wellness. Its methods are performed iteratively to achieve excellent results [20]. The iterative experimentation and recurrent refining facilitate in optimizing model performance. Yet most prior studies in detecting mental disorders employed different methodology or none at all.

Therefore, the aim of this research is to build ML models using multiple supervised ML algorithms that had shown remarkable performance in former studies. This study employed a data mining process using the CRISP-DM approach. Secondly, to optimize and choose the superior ML model in detecting anxiety and depressive disorders, using the top-notch supervised algorithms from preceding research. Through building a ML

[1] *Information System Management Department, BINUS Graduate Program – Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia 11480*
[2]*Psychology Department, Esa Unggul University, Jakarta, Indonesia*
**Corresponding Author Email: callista.mogie@binus.ac.id*
ORCID ID : 0009-0004-6735-1881

model to detect anxiety and depression, this study aids in efficient screening. The findings may provide a model for early detection that can be cost and time efficient for patients and healthcare providers. Besides, contributing to the advancement of modern applications and innovative solutions to improve world mental health.

2. Related Work

Machine learning methods have shown promising potential in predicting various mental health disorders using different datasets and methodologies. The studies reviewed encompass a range of mental health conditions and employ diverse machine learning algorithms to achieve predictive accuracy. The studies covered mental disorders including anxiety, depression, schizophrenia, Post-Traumatic Stress Disorder (PTSD), mental illness in adolescence and suicidality.

Naïve Bayes (NB) is known to lack complexity yet performed highest accuracy in depression classification compared to Decision tree, RF, Support Vector Machine (SVM), and K-Nearest Neighbour (K-NN) [21]. The algorithm is applicable for binary and multi-class classification tasks [19]. It has been used in categorizing spam messages to classify physical ailments. Hence, it is suitable for classifying psychological disorders.

Meanwhile, Vaishnavi et al. [7] explored various classifiers to predict unspecified psychological disorders such as Regression, K-NN Classifier, Decision Tree Classifier, Random Forest, and Stacking. The dataset consisted of 27 columns and 1259 entries, resulted in stacking techniques getting the highest accuracy of 81.75%, followed by Random Forest of 81.22%, Decision Tree classifier 80.69%, Kneighbors Classifier 80.42%, Logistic Regression 79.63%.

Srinivasagopalan et al. [22] predicted Schizophrenia disorder through several algorithms such as Deep Learning (DL), Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR). The data consisted of 69 schizophrenia patients and 75 control patients, reported DL achieving the highest accuracy of 94.44%, while SVM yielded a commendable accuracy of 83%. In practice, SVM can be applied to both linear and non-linear data. For non-linear data, kernel technique is used to map the data into higher dimensions, where a linear hyperplane can be used to separate the classes [23]. In a systematic literature review of 48 articles classifying disease, support vector machines are the most frequently utilized, while random forest yielded the best accuracy among other supervised learnings in the matter [24].

Subsequently, various studies have demonstrated the superior performance of random forest (RF) in the classification of mental illnesses. As psychological disorders can affect individuals of any age, one study conducted in 2017 focused on classifying anxiety and depression in 520 elderly patients in a hospital using an HADS questionnaire [25]. The algorithms utilized were Bayesian Network (BN), logistic regression, multiple layer perceptron (MLP), Naïve Bayes (NB), random forest (RF), random tree (RT), J48, sequential minimal optimization (SMO), random sub-space (RS), and K-Star (KS) models. Among all algorithms, the result showed RF achieving the highest accuracy at 89%. A study investigated PTSD among 13,690 military subjects employing various algorithms, namely Random Forest (RF), Support Vector Machine (SVM), Artificial Neural Networks (ANN), and Bagging. Once again, RF obtained an accuracy of 97% [26].

Tate et al. [27] extended the exploration of RF's utility by examining mental health problems in adolescence. The study involved 7,638 participants using the Strengths and Difficulties Questionnaire (SDQ). Compared to Logistic Regression, XGBoost, Support Vector Machine, Neural Network, RF emerged as a standout performer with the highest Area Under the Curve (AUC) at 0.739.

Another study conducted by Jo [27], utilized brain imaging to predict Schizophrenia employing RF, Multinomial Naïve Bayes, XGBoost (XGB), and Support Vector Machine (SVM). Their study reported RF achieving the highest accuracy of 68.9%. Afterward, Kumar [29] used a hybrid approach alongside Random Forest, J48, KNN, Kstar, RFBN, MLP, Naïve Bayes, and Bayes Network to predict Depression, Anxiety, and Stress. As expected, the RF model yielded 100% accuracy in classifying anxiety and 93.1% in depression classification. One systematic literature review study found that random forest (RF) has been showing performance superiority in disease prediction. In predicting schizophrenia disorder, RF provided lower error rates compared to other methods [15].

Other studies have used RF algorithm alone, such as in Marmar et al. [30], utilized ML model in speech recognition to detect PTSD. They assessed PTSD prediction using Random Forest, achieving an accuracy of 89.1% and an AUC of 0.954 based on a smaller sample size of 52 PTSD subjects and 77 trauma-exposed controls. Similarly, Van Mens et al. [31] predicted suicidality on a vast database, reporting a PPV of 0.05, sensitivity of 0.39, and an AUC of 0.82, indicating high specificity but relatively lower sensitivity. In summary, Random Forest consistently demonstrates high accuracy across various mental health studies, making it a reliable and effective tool for disease prediction and classification. However, other algorithms also have shown superior performance such as Naive Bayes, ANN, and SVM must be considered.

Artificial neural networks (ANN) algorithms have been buzzing due to its capacity, including natural language processing, image, and speech recognition [15]. It consists of multiple hidden layers, and each unit within the network plays a role in determining the network's overall performance [14]. ANN obtained 89% accuracy in classifying PTSD of the military forces. However, the application in identifying anxiety and depression remains restricted and is yet to be enhanced [26].

Overall, these studies demonstrate the potential of machine learning algorithms in predicting various mental health disorders, albeit with variations in accuracy based on the specific disorder, dataset characteristics, and chosen methodologies.

3. Methodology

Based on the CRISP-DM framework, there are 6 iterative phases, starting from business understanding, data understanding, data preparation, modeling, evaluation, and ending in deployment [20]. Jupyter Notebook with Python 3.10.6 in Visual Studio Code was utilized for the whole methodology.

3.1. Business Understanding

The initial step is to understand the business' needs and solutions in terms of data mining projects [20]. Demand for anxiety and depression screening increases due to heightened cases and public awareness. It is forecasted that the market revenue for anxiety disorders will reach US$8.72bn by 2028, with annual growth (CAGR 2023-2028) of 0.49%. While depression is projected to reach US$21.09bn in 2023 and US$21.77bn by 2028, CAGR (2023-2028) of 0.64% [32].

Early detection is necessary since it can relieve the symptoms of mental disorders and increase overall productivity and quality of life. Hence, the data mining objective is to build a robust ML model for classifying the presence of anxiety and depression. This initiative holds immense potential for mental health businesses, encompassing healthcare providers and mental health professionals, by integrating the model into their practices and creating additional revenue streams. Beyond financial gains, the incorporation of such models is poised to enhance treatment efficiency, resulting in heightened patient satisfaction and improved overall outcomes.

3.2. Data Understanding

The dataset was retrieved from the Kaggle repository uploaded by Nicolò Malatesta. The data consists of Generalized Anxiety Disorder-7 (GAD-7) and Patient Health Questionnaire-9 (PHQ-9). The GAD-7 was invented by Spitzer et al. (2006), consists of 7 items about experiencing generalized anxiety disorder symptoms. It has been widely used in clinical practice and research to monitor individuals' anxiety disorder [33-35]. Whereas PHQ-9 was made by Kroenke et al. in 2001 consists of 9-items experiencing depressive disorder symptoms. Participants were asked to complete the self-administered questionnaire. The dataset has 1225 instances with 17 columns.

3.3. Data Preparation

In this phase, the dataset was prepared for modeling, including feature selection, data transformation, and checking for inconsistent or empty values. In feature selection, the unnecessary feature was dropped. The mutual information (MI) tool was utilized for gaining insights on most informative features in class.
The first step was feature selection. Next, labeling where each instance is labeled into one of four categories, namely anxiety, depression, 'both', and healthy. The labeling process is done by a mental health professional. Afterwards, the class distribution was found, consisting of 'Both' 78.7% of total datasets, 'Depressed' class over 13%, 'Anxiety' was 1.7%, and 'Healthy' was 6.6% of the total dataset. Though severely imbalanced, there were no missing values in the dataset.
Afterward, the labels were encoded using `LabelEncoder()`. The goal is to replace the categorical value of the label to number (integer), since ML only can process integer data. The labels were encoded as follows, class 'Both' converted to 0, 'Depressed' converted to '1', 'Anxiety' to '2', and 'Healthy' to '3'.
The next step was to split data into training (80%) and testing data (20%), as presented in Table 1 by using Pareto principle [36].
Subsequently, the imbalanced training dataset through Synthetic Minority Over-sampling Technique (SMOTE). SMOTE has been widely utilized to address imbalanced datasets in clinical settings and successfully aided in increasing performance on imbalanced data model performance [27, 37). SMOTE has shown a higher F1-score in classifying E-coli compared to ADASYN (Halim et al., 2023). SMOTE also performed better compared to Random Over-sampling technique [38].
Instead of directly duplicating minority samples like ROS, SMOTE creates synthetic instances within the training dataset through connecting existing minority instances and produces new synthetic minority points along these connections [38].

3.4. Modeling

In modeling, the supervised ML algorithms utilized are Gaussian Naïve Bayes (Gaussian NB), RF, ANN, and SVM to build the classification model. The training used the `fit` method from Python. The train data teaches the classifier how to make predictions based on the features. There were four iterative experiments carried out to reach the optimum classification model for detecting depression, anxiety, and healthy classes. The first experiment used an imbalanced data train set. Next, the second experiment utilized a balanced data train set through SMOTE. Afterward, a hyperparameter is being tuned to yield better performance. Similar to experiment 2, the third experiment utilized `GridSearchCV`, as it has been a broadly utilized tool in trying out different settings and picking the best one based on how the model performs in cross-validation. The final experiment is feature selection, selecting the most contributing and informative features. The aim of this experimentation is to maximize predictive performance metrics (F1-score, precision, recall, and accuracy) with minimum features. The process involves using SelectKBest, which tests different numbers of features ('k') and evaluates precision, recall, and F1-score for each individual class.

3.5. Evaluation

Evaluation is carried out to assess whether the results of experiments or the model performance are aligned with the initial objectives. Furthermore, this stage will include analyzing the result of each model. There are several metrics that will be measured including accuracy of each model, precision, recall, F1-Score in macro and weighted average. Macro-average calculates the metric independently for each class and then takes the average across all classes. It gives equal weight to each class, regardless of class imbalance to assess performance across all classes without considering class imbalance. Weighted average calculates the metric for each class and then takes the average, weighted by the number of true instances in each class, giving more weight to classes with more instances. This can provide a more representative measure of overall model performance when classes are imbalanced. F1 score is mainly focused to assess the model's ability to correctly identify true positive cases while minimizing false positives.

3.6. Deployment

This stage is to plan the deployment, monitoring, and maintenance of the machine learning model. The deployment is planned to be applied in the mental wellness businesses, as an early anxiety and depression screening tool for anxiety and depression disorder.

4. Result

The model resulted in multiclass and binary classification. In multiclass classification with four labels (0 both, 1 depression, 2 anxiety, 3 healthy), the goal is to classify instances into one of the four categories. Each instance can only belong to one class, and the model's output is one of the four class labels. For instance, if the model predicts an instance as "2" it means the model has classified that instance as belonging to the anxiety class. In contrast, binary classification involves classifying instances into one of two classes. For example, in the context of mental health screening, binary classification might involve predicting whether an individual has a mental health condition (e.g., depression or anxiety) or not (healthy).

4.1. Multiclass Classification

In the first experiment, the dataset utilized for modeling was the dataset without SMOTE procedure. Random Forest has the highest precision, recall, F1-score, and accuracy both weighted and macro average, indicating it performs better in terms of precision and recall across all classes equally. This is because Random Forest inherently provides mechanisms to deal with class imbalance, such as bootstrapping and aggregation. Moreover, Random Forest generally works well with both numerical and categorical features, and it's less sensitive to outliers and noise in the data. Also, it can provide insights into feature importance, which can be helpful for understanding the underlying patterns in the data and feature engineering. However, other algorithms were not yielding excellent results, such as ANN that gained the lowest scores, followed by Naive Bayes, and Support Vector Machine. ANN is a form of deep learning models with complex architectures that may require careful tuning of hyperparameters, such as the number of layers, neurons per layer, and batch size, to achieve optimal performance. If ANN is not properly configured, it may not generalize well to the dataset, leading to lower performance. Hence, further experimentation was conducted, including SMOTE and hyperparameter tuning to each algorithm. Similar to SVM, SVM has hyperparameters such as the choice of kernel function and regularization parameter (C), which need to be carefully tuned for optimal performance. Suboptimal choices of hyperparameters can lead to reduced performance.

Table 1. Multiclass Classification Result

Experiments	Algorithm	Macro Average			Weighted Average			Accuracy
		Precision	Recall	F1-score	Precision	Recall	F1-score	
1: Imbalance Dataset	ANN	0,34	0,31	0,30	0,68	0,71	0,68	71%
	NB	0,51	0,60	0,47	0.86	0.76	0.79	76%
	SVM	0,68	0,70	0,69	0,92	0,95	0,94	95%
	RF	0,93	0,72	0,74	0,95	0,95	0,94	95%
2: SMOTE Dataset	ANN	0,53	0,55	0,47	0,78	0,78	0,76	78%
	NB	0,71	0,68	0,56	0,91	0,84	0,85	84%
	RF	0,90	0,83	0,85	0,95	0,95	0,95	95%
	SVM	0,88	0,86	0,87	0,96	0,96	0,96	96%
3: Hyperparameter Tuning	NB	0,81	0,89	0,85	0,92	0,90	0,91	90%
	RF	0,90	0,84	0,86	0,96	0,96	0,95	96%
	SVM	0,88	0,82	0,84	0,96	0,96	0,96	96%
	ANN	0,93	0,95	0,94	0,97	0,97	0,97	97%

In experiment 2 after class imbalances were addressed through SMOTE, the result showed increased performance for each algorithm, except for Random Forest, the F1-score ranked three and showed insignificant changes. On the other hand, SVM achieved the highest score among all showing the default parameter ‘rbf’ is sensitive to class imbalances. The balanced class distribution provides a more representative training dataset, enabling the models to learn from a broader range of instances across all classes. As a result, the models are better able to generalize and make accurate predictions for all classes, leading to improved performance metrics.

Based on the provided data from Experiment 3, each algorithm displayed improvement after parameters tuned with GridSearchCV. Artificial Neural Network (ANN) was the top-performing algorithm, achieving the highest precision, recall, accuracy and F1-score among all algorithms evaluated. The best parameters found were 100 epochs, 0.1 validation split, 1 hidden layer, 18 neurons in hidden layer, and 10 batch sizes. At the same time, Random Forest outperformed the SVM’s macro-average score, which suggests that the model performs well on average across all classes, regardless of their individual sizes or characteristics. SVM slightly outperformed Random Forest F1-score weighted average, indicating the model had better balance between correctly identifying positive instances (precision) and capturing all positive instances (recall) on both majority and minority classes. RBF kernel was deemed most suitable during hyperparameter tuning; it suggested that the model requires a nonlinear decision boundary to effectively separate the classes in the dataset.

Whereas Naive Bayes was the lowest performing algorithm in experiment 3, it demonstrated respectable performance with an accuracy of 90%. The score also increased after the hyperparameter was tuned. In Gaussian Naive Bayes (GNB), when calculating the probabilities of features belonging to different classes, the mean and variance of each feature for each class need to be estimated. This is because GNB assumes that the features are normally distributed (follow a Gaussian distribution). The optimal var_smoothing parameter found 0.0534. This parameter controls the amount of smoothing applied to the variances of features in GNB. By tuning var_smoothing, the balance between relying on observed data and introducing smoothing to stabilize the estimation process is adjusted, particularly useful for features with low variance or small datasets.

4.2. Binary Classification

In binary classification, the machine only predicts two classes. For binary classification class 0 (Both Anxiety and Depression), the model learnt to predict between class 0 and class other than Both. As seen in table 2, SVM achieved superior performance in terms of precision, recall, F1-score and accuracy in predicting class 0 and 1. Followed by ANN, which only had 1-2% differences of each metric. SVM achieved the highest precision, indicating that it had the lowest rate of false positives among the algorithms. High precision means it was effective in correctly identifying instances belonging to the 'both' class while minimizing misclassifications. Binary classification task is inherently simpler as it involves only two categories, making it potentially easier for algorithms to achieve higher accuracy. The highest recall, indicating that it had the lowest rate of false negatives among the algorithms. This means it effectively captured a high proportion of actual 'both' instances. And the F1-score successfully portrayed the harmonic balance of both precision and recall scores. Random forest achieved the lower but still respectable scores across all metrics. It demonstrated good performance in distinguishing Class 0 and Class 1 instances, albeit slightly lower than SVM and ANN. While NB exhibited lowest scores, it still demonstrated decent performance in classifying Class 0 and 1 instance, with acceptable precision, recall, F1-score, and accuracy.

Meanwhile in class 2 and 3 where the test set instances were low due to minority class, SVM showed a declining result in macro average, indicating it is sensitive to imbalance. However, the weighted average is still performed as well as class 0 and 1 classification. Whereas in class 3 Random Forest was the best model compared to SVM, ANN, RF. Class 3 has a smaller number of instances compared to other classes, RF inherent ability to handle class imbalance can be advantageous. The ensemble of decision trees can adapt to the imbalance in the data and make accurate predictions for the minority class.

Table 2. Binary Classification Result

Class	Algorithm	Macro Average			Weighted Average			Accuracy
		Precision	**Recall**	**F1-score**	**Precision**	**Recall**	**F1-score**	
0	SVM	0,99	0,98	0,99	0,99	0,99	0,99	99%
	ANN	0,96	0,98	0,97	0,98	0,98	0,98	98%
	RF	0,96	0,96	0,96	0,97	0,97	0,97	97%
	NB	0,89	0,93	0,91	0,94	0,93	0,94	93%
1	SVM	0,96	0,93	0,94	0,98	0,98	0,98	98%
	ANN	0,97	0,9	0,93	0,97	0,97	0,97	97%
	RF	0,91	0,91	0,91	0,96	0,96	0,96	96%
	NB	0,71	0,78	0,74	0,89	0,87	0,88	87%
2	ANN	0,94	0,94	0,94	0,99	0,99	0,99	99%
	SVM	0,89	0,93	0,91	0,99	0,99	0,99	99%
	RF	0,99	0,75	0,83	0,98	0,98	0,98	98%
	NB	0,66	0,91	0,73	0,97	0,94	0,95	94%
3	RF	0,99	0,88	0,93	0,98	0,98	0,98	98%
	ANN	0,94	0,79	0,85	0,97	0,97	0,96	97%
	SVM	0,94	0,79	0,85	0,97	0,97	0,96	96%
	NB	0,82	0,92	0,86	0,97	0,96	0,96	96%

4.3. Feature Importance

As seen on table 3, there were five most contributing features for each class. Feature importance was measured with Mutual Information score. The higher the mutual information score of a feature, the more informative the features regarding the target variable. The most important features have shown there is an alignment in depression and anxiety symptoms.

4.3.1. Class 0: Both (Depression and Anxiety)

In the 'Both' or class 0, the most contributing features consisted of 4 depression symptoms and 1 anxiety symptom (table 3). Despite the distinct symptoms, those disorders commonly co-occur [4]. The most prominent feature in determining depression and anxiety is feeling down and hopeless (PHQ2), suggesting feelings of sadness and hopelessness, which are common symptoms of both depression and anxiety. When someone experiences both depression and anxiety, they are more likely to have negative feelings regarding the future and ruminate on the past [17]. In depression, people will also experience hopelessness about the future.

The second prominent feature was having trouble concentrating (PHQ7), also associated with both anxiety and depression disorders. This is because anxiety and depression lower people's cognitive abilities, affecting their ability to concentrate, make decisions, and perform everyday tasks.

Table 3. Feature Importance

Class	Rank	Features	Information
0	1	PHQ2	Feeling down and hopeless
	2	PHQ7	Trouble concentrating
	3	GAD1	Feel anxious
	4	PHQ3	Sleep problem
	5	PHQ8	Unusual movement
1	1	GAD7	Trouble concentrating
	2	GAD3	Excessive worry
	3	PHQ3	Sleep problem
	4	PHQ4	Feeling tired
	5	PHQ5	Appetite problem
2	1	GAD6	Irritable
	2	GAD2	Uncontrollable worry
	3	PHQ3	Sleep problem
	4	GAD4	Trouble relaxing
	5	PHQ6	Feeling bad or failure of oneself

	1	GAD6	Irritable
	2	GAD7	Trouble concentrating
3	3	PHQ9	Suicidal or self-harm thoughts
	4	PHQ4	Feeling tired
	5	PHQ5	Appetite problem

Meanwhile, feeling anxious in GAD1 is related to anxiety symptoms, which involve extreme fear that something awful might happen and an intense fear of the future. These fears can also manifest in physiological symptoms such as lightheadedness, shaking, and muscle tension. In contrast, depression symptoms involve being preoccupied with negative thoughts about oneself, the world, or the future.

Sleep problems are also among the main symptoms of both illnesses and can manifest as oversleeping or insomnia. Oversleeping or sleeping too much can occur as a coping strategy for mental and physical fatigue or as a means to escape reality through sleep. Due to negative perceptions and feelings, individuals with these illnesses may struggle to relax and fall asleep.

The fifth main feature, unusual movement (PHQ8), indicates that depression and anxiety can cause a person's movements to become unusual. Individuals may become restless or move too little or slowly. In response to the symptoms of illness, they might try to ease the mental burden through excessive movement or fidgeting. Another impact may include moving very little and speaking slowly. These symptoms must be assessed and prioritized in individuals with depression and anxiety.

4.3.2. Class 1: Depression

In class 1 (Depression), the most contributing features consisted of 2 anxiety symptoms and 3 depression symptoms. Difficulty concentrating (GAD7) is a common symptom of both depression and anxiety. Specifically, individuals with depression often experience negative distortions and feelings that make it difficult for them to focus on the present moment, thereby interfering with their daily activities. Clinicians must pay close attention when someone exhibits symptoms of difficulty concentrating, as this seemingly simple issue can have significant implications.

GAD3 (Excessive worry) is another prominent feature in depression. While excessive worry is more characteristic of anxiety, it can also be present in depression. Individuals with depression may experience intense worry and rumination, leading them to dwell on negative thoughts and predict future negative outcomes, thus exacerbating feelings of hopelessness [17]. Emotional disturbances are common in individuals with depression, leading to frequent experiences of negative emotions.

Disturbances in sleep patterns (PHQ3), such as insomnia or hypersomnia, are prevalent in individuals with depression. Insomnia, characterized by decreased sleep time, and hypersomnia, characterized by excessive sleep, are both indicators of depression. Individuals with depression may complain more about having difficulty falling asleep, which can contribute to their overall sense of fatigue and low energy levels (PHQ4). Insomnia can significantly impact energy levels, as defined by the DSM-5, resulting in decreased daytime functioning [2].

Changes in appetite, indicated as appetite problems, including overeating or loss of appetite (PHQ5), are also common in depression. These changes in eating habits can lead to weight gain or weight loss and are considered somatic symptoms. Additionally, individuals with depression often experience anhedonia, a diminished interest or pleasure in activities they once enjoyed, which may contribute to changes in appetite and eating behaviors.

4.3.3. Class 2: Anxiety

In class 2, the most contributing features consisted of 3 anxiety symptoms and 2 depression symptoms as illustrated in table 3. Irritability or being easily annoyed (GAD6) was found to be the most contributing feature or symptom of anxiety disorders. It refers to being easily agitated, annoyed, or even displaying a tendency to become angry compared to others in the same age group [39]. Adolescents who experience irritability may have difficulties regulating their emotions due to abnormal processing of emotional stimuli in the rostro-medial prefrontal cortex, which makes anxiety more challenging to control [40].

GAD2, excessive and uncontrollable worry, is a central feature of anxiety disorders. Individuals commonly experience persistent and intense worry ranging from minor to major occurrences, including everyday concerns such as job duties, health, finances, and the well-being of family members, as well as potential disasters. In children, excessive worry often revolves around concerns about their abilities or performance standards. Over time, the focus of this worry may shift [2].

As a result of excessive worry, PHQ3, defined as 'trouble falling or staying asleep, or sleeping too much', is common in individuals with anxiety disorders. Individuals may find it difficult to sleep soundly at night, leading to daytime fatigue. Additionally, individuals may compensate for fatigue by sleeping excessively, further disrupting their sleep patterns, and exacerbating feelings of anxiety.

Therefore, PHQ3, defined as 'trouble falling or staying asleep, or sleeping too much', is common in individuals with anxiety disorders. Due to excessive worrying (GAD2), individuals may find it difficult to sleep soundly at night, leading to daytime fatigue. Additionally, individuals may compensate for fatigue by sleeping excessively, further disrupting their sleep patterns, and exacerbating feelings of anxiety.

Next, GAD4 is defined as trouble relaxing, where individuals find it hard to relax because they are constantly in an anxious state as explained in GAD2 feature excessively worrying. This feature is also related to PHQ3 which has trouble sleeping.

PHQ6 is defined as feeling negative of oneself, feeling as if one has disappointed others or oneself, or feeling like a failure. Individuals with anxiety disorders are prone to worrying about work or school performance, and feelings of incompetence are common among people with anxiety. This is because symptoms of anxiety interfere with clarity of mind, cognitive abilities, sleep schedules, and emotion regulation, which may disrupt various aspects of life. For instance, a person with anxiety may become irritable and easily annoyed, potentially affecting interpersonal relationships with peers, colleagues, or family members. Additionally, individuals with anxiety often experience intense worry that disturbs sleep and consequently impacts cognitive abilities, putting them at risk of academic failure.

4.3.4. Class 3: Healthy

In class 3, the most contributing features were 2 anxiety symptoms and 3 depression symptoms. Firstly, GAD6 is irritability, which is the symptom of anxiety. This feature suggested, the presence or absence of irritability has a significant relationship with being classified as healthy. Since this feature is the highest in contributing to the Anxiety class, this feature might be less prominent in healthy individuals compared to those with depression or anxiety disorders.

The second feature was GAD7 or trouble concentrating. The mutual information score for this feature indicates the strength of its association with being healthy. A higher score suggests that individuals with little to no trouble concentrating are more likely to be classified as healthy.

PHQ9 means having suicidal or self-harming thoughts is the third most correlated features with healthy class. It indicated how much information the presence or absence of suicidal or self-harm thoughts provides about an individual being healthy. On the other hand, it depends on the intensity of wanting to kill oneself and other existing features or symptoms of mental illness as well.

PHQ4 or feeling tired might be less frequent in healthy individuals compared to those experiencing depression or anxiety. A higher mutual information score implies that individuals without feelings of tiredness are more likely to be considered healthy.

Finally, appetite issue or PHQ5's mutual information score indicates its association with being healthy. A higher score suggests that individuals without appetite problems are more likely to be classified as healthy.

5. Conclusion

This research provides distinctive features that contribute significantly to classifying mental states—depression, anxiety, both, and a healthy state which provides crucial insights for healthcare practitioners. After resampling with SMOTE and doing hyperparameter tuning, it was found the best algorithm for multiclass classification was ANN with an F1-score of 0.97. Furthermore, the best algorithm for binary classification for class 0 and 1 was SVM with F1 score of 0.99 and 0.98, class 2 was ANN with F1-score of 0.99, and class 3 was Random Forest with F1-score of 0.98. Binary classification showed slightly greater performance than multiclass classification. The most informative features of both depression and anxiety were feeling down and hopeless, trouble concentrating, feeling anxious, sleep problems, and unusual movements. Whereas the most informative features for depression disorder were trouble concentrating, excessive worry, sleep problem, feeling tired, and appetite problem. While the most informative features for class anxiety were irritability, uncontrollable worry, sleep problems, trouble relaxing, and feeling bad or feeling like a failure. Eventually, informative features for a healthy class were irritability, trouble concentrating, suicidal or self-harm thoughts, feeling tired, and appetite problem. These features can help provide early identification and targeted intervention for individuals exhibiting signs of depression, anxiety, or both. By leveraging machine learning models, healthcare professionals can utilize these predictive insights as a supplementary tool in their diagnostic process, potentially enhancing efficiency of mental health treatments. The emphasis on specific symptoms or behavioral patterns associated with each class enables a more nuanced approach towards personalized treatment strategies tailored to address the unique needs of individuals experiencing diverse mental health conditions.

While Artificial Neural Networks (ANN) and Support Vector Machines (SVM) demonstrated strong performance in this study, future research could explore additional models or ensemble methods to improve classification accuracy further. Additionally, integrating these predictive models into clinical settings by creating user-friendly interfaces that provide decision support for healthcare providers could enhance their ability to make informed decisions. Implementing these models in mental health applications or e-health platforms could increase user engagement and provide trustworthy psychometric assessments for screening anxiety and depression.

Acknowledgements

This research was supported by BINA Nusantara University. We thank Rezky Alundya, B. Psych for assistance with dataset labeling, who provided insight and expertise that greatly assisted the research and provided theoretical view in psychology.

Author contributions

Callista Ivana Mogie: Background, literature studies, data preprocessing, modeling, data analysis, and writing.

Tuga Mauritsius: Guidance in methodology and writing systematic.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] J. R. Bullis, H. Boettcher, S. Sauer-Zavala, T. J. Farchione, and D. H. Barlow, "What is an emotional disorder? A transdiagnostic mechanistic definition with implications for assessment, treatment, and prevention," *Clinical Psychology: Science and Practice*, vol. 26, no. 2, p. e12278, Mar. 2019, doi: 10.1111/cpsp.12278.

[2] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders: DSM-5.*, 5th ed. Arlington: American Psychiatric Publishing, 2013.

[3] D. H. Barlow, B. A. Harris, E. H. Eustis, and T. J. Farchione, "The unified protocol for transdiagnostic treatment of emotional disorders," World Psychiatry, vol. 19, no. 2, pp. 245–246, May 2020, doi: 10.1002/wps.20748

[4] R. N. Groen *et al.*, "Comorbidity between depression and anxiety: Assessing the role of bridge mental states in dynamic psychological networks," *BMC Medicine*, vol. 18, no. 1, Sep. 2020, doi: 10.1186/s12916-020-01738-z.

[5] World Health Organization, *World mental health report: Transforming mental health for all*. WHO, 2022. [Online]. Available: https://www.who.int/publications/i/item/9789240049338

[6] D. F. Santomauro *et al.*, "Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic," *The Lancet*, vol. 398, no. 10312, pp. 1700–1712, Nov. 2021, doi: 10.1016/s0140-6736(21)02143-7.J. Ingram, C. J. Hand, and G. Maciejewski, "Social isolation during COVID-19 lockdown impairs cognitive function," *Applied Cognitive Psychology*, vol. 35, no. 4, pp. 935–947, doi: 10.1002/acp.3821.

[7] K. Vaishnavi, U. Nikhitha Kamath, B. Ashwath Rao, and N. V. Subba Reddy, "Predicting Mental Health Illness using Machine Learning Algorithms," Journal of Physics: Conference Series, vol. 2161, no. 1, p. 012021, Jan. 2022, doi: 10.1088/1742-6596/2161/1/012021.

[8] R. S. McIntyre and J. R. Calabrese, "Bipolar depression: the clinical characteristics and unmet needs of a complex disorder," *Current Medical Research and Opinion*, vol. 35, no. 11, pp. 1993–2005, Aug. 2019, doi: 10.1080/03007995.2019.1636017.

[9] H. Brooks *et al.*, "Children and young people's beliefs about mental health and illness in Indonesia: A qualitative

study informed by the Common Sense Model of Self-Regulation," *PLOS ONE*, vol. 17, no. 2, p. e0263232, Feb. 2022, doi: 10.1371/journal.pone.0263232.

[10] G. Tampubolon and W. Hanandita, "Poverty and mental health in Indonesia," *Social Science & Medicine*, vol. 106, pp. 20–27, Apr. 2014, doi: 10.1016/j.socscimed.2014.01.012.

[11] M. Marthoenis, M. C. Aichberger, and M. Schouler-Ocak, "Patterns and Determinants of Treatment Seeking among Previously Untreated Psychotic Patients in Aceh Province, Indonesia: A Qualitative Study," *Scientifica*, vol. 2016, pp. 1–7, 2016, doi: 10.1155/2016/9136079.

[12] N. F. Praharso, H. Pols, and N. Tiliopoulos, "Mental health literacy of Indonesian health practitioners and implications for mental health system development," *Asian Journal of Psychiatry*, vol. 54, p. 102168, Dec. 2020, doi: 10.1016/j.ajp.2020.102168.

[13] E. Mahase, "Covid-19: Mental health consequences of pandemic need urgent research, paper advises," *BMJ*, p. m1515, Apr. 2020, doi: 10.1136/bmj.m1515.

[14] S. Ray, "A quick review of machine learning algorithms," in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Feb. 2019. Accessed: May 20, 2023. [Online]. Available: http://dx.doi.org/10.1109/comitcon.2019.8862451.

[15] J. Chung and J. Teo, "Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges," *Applied Computational Intelligence and Soft Computing*, vol. 2022, pp. 1–19, Jan. 2022, doi: 10.1155/2022/9970363.

[16] F. Cacheda, D. Fernandez, F. J. Novoa, and V. Carneiro, "Early Detection of Depression: Social Network Analysis and Random Forest Techniques," *Journal of Medical Internet Research*, vol. 21, no. 6, p. e12554, Jun. 2019, doi: 10.2196/12554.

[17] M. W. Eysenck and M. Fajkowska, "Anxiety and depression: Toward overlapping and distinctive features," *Cognition and Emotion*, vol. 32, no. 7, pp. 1391–1400, Jun. 2017, doi: 10.1080/02699931.2017.1330255.

[18] S. Venkatraman, S. Parvin, W. Mansoor, and A. Gawanmeh, "Big data analytics and internet of things for personalised healthcare: Opportunities and challenges," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 4, p. 4306, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4306-4316.

[19] J. Maiga, G. G. Hungilo, and Pranowo, "Comparison of machine learning models in prediction of cardiovascular disease using health record data," in *2019 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Oct. 2019. Accessed: Aug. 22, 2023. [Online]. Available: http://dx.doi.org/10.1109/icimcis48181.2019.8985205

[20] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[21] A. Priya, S. Garg, and N. P. Tigga, "Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms," Procedia Computer Science, vol. 167, pp. 1258–1267, 2020, doi: 10.1016/j.procs.2020.03.442.

[22] S. Srinivasagopalan, J. Barry, V. Gurupur, and S. Thankachan, "A deep learning approach for diagnosing schizophrenic patients," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 31, no. 6, pp. 803–816, Jan. 2019, doi: 10.1080/0952813x.2018.1563636.

[23] M. T. R. Ullas, M. Begom, A. Ahmed, and R. Sultana, "A machine learning approach to detect depression and anxiety using supervised learning," 2019. [Online]. Available: http://hdl.handle.net/10361/12274

[24] S. Uddin, A. Khan, M. E. Hossain, and M. A. Moni, "Comparing different supervised machine learning algorithms for disease prediction," *BMC Medical Informatics and Decision Making*, vol. 19, no. 1, Dec. 2019, doi: 10.1186/s12911-019-1004-8.

[25] A. Sau and I. Bhakta, "Predicting anxiety and depression in elderly patients using machine learning technology," *Healthcare Technology Letters*, vol. 4, no. 6, pp. 238–243, Nov. 2017, doi: 10.1049/htl.2016.0096.

[26] D. Leightley, V. Williamson, J. Darby, and N. T. Fear, "Identifying probable post-traumatic stress disorder: applying supervised machine learning to data from a UK military cohort," *Journal of Mental Health*, vol. 28, no. 1, pp. 34–41, Nov. 2018, doi: 10.1080/09638237.2018.1521946.

[27] A. E. Tate, R. C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, and R. Kuja-Halkola, "Predicting mental health problems in adolescence using machine learning techniques," *PLOS ONE*, vol. 15, no. 4, p. e0230389, Apr. 2020, doi: 10.1371/journal.pone.0230389.

[28] Y. T. Jo, S. W. Joo, S. Shon, H. Kim, Y. Kim, and J. Lee, "Diagnosing schizophrenia with network analysis and a machine learning method," *International Journal of Methods in Psychiatric Research*, vol. 29, no. 1, Feb. 2020, doi: 10.1002/mpr.1818.

[29] P. Kumar, S. Garg, and A. Garg, "Assessment of Anxiety, Depression and Stress using Machine Learning Models," *Procedia Computer Science*, vol. 171, pp. 1989–1998, 2020, doi: 10.1016/j.procs.2020.04.213.

[30] C. R. Marmar *et al.*, "Speech-based markers for posttraumatic stress disorder in US veterans," *Depression and Anxiety*, vol. 36, no. 7, pp. 607–616, Apr. 2019, doi: 10.1002/da.22890.

[31] K. van Mens *et al.*, "Applying machine learning on health record data from general practitioners to predict suicidality," *Internet Interventions*, vol. 21, p. 100337, Sep. 2020, doi: 10.1016/j.invent.2020.100337.

[32] Statista, "Mental health - Worldwide," *Statista*, 2023. https://www.statista.com/outlook/hmo/mental-health/worldwide (accessed Jun. 01, 2023).

[33] A. Teymoori *et al.*, "Factorial structure and validity of depression (PHQ-9) and anxiety (GAD-7) scales after traumatic brain injury," *Journal of Clinical Medicine*, vol. 9, no. 3, p. 873, Mar. 2020, doi: 10.3390/jcm9030873.

[34] M. Shevlin *et al.*, "Measurement invariance of the Patient Health Questionnaire (PHQ-9) and Generalized Anxiety Disorder scale (GAD-7) across four European countries during the COVID-19 pandemic," *BMC Psychiatry*, vol. 22, no. 1, pp. 1–9, Mar. 2022, doi: 10.1186/s12888-022-03787-5.

[35] A. Pranckeviciene *et al.*, "Validation of the patient health questionnaire-9 and the generalized anxiety disorder-7 in

Lithuanian student sample," *PLOS ONE*, vol. 17, no. 1, p. e0263027, Jan. 2022, doi: 10.1371/journal.pone.0263027.

[36] H. N. Cahyani and R. Arifudin, "Improving the Accuracy of Multinomial Naïve-Bayes Algorithm with Adaptive Boosting Using Information Gain for Classification of Movie Reviews Sentiment Analysis," *Journal of Advances in Information Systems and Technology*, vol. 4, no. 1, pp. 50–59, Dec. 2022, doi: 10.15294/jaist.v4i1.60267.

[37] M. Naseriparsa, A. Al-Shammari, M. Sheng, Y. Zhang, and R. Zhou, "RSMOTE: improving classification performance over imbalanced medical datasets," *Health Information Science and Systems*, vol. 8, no. 1, Jun. 2020, doi: 10.1007/s13755-020-00112-w.

[38] Lima, J. L. P., Macedo, D., & Zanchettin, C. (2019, July). Heartbeat Anomaly Detection using Adversarial Oversampling. *2019 International Joint Conference on Neural Networks (IJCNN)*. http://dx.doi.org/10.1109/ijcnn.2019.8852242

[39] E. Leibenluft, "Pediatric irritability: A systems neuroscience approach," *Trends in Cognitive Sciences*, vol. 21, no. 4, pp. 277–289, Apr. 2017, doi: 10.1016/j.tics.2017.02.002.

[40] K. I. Crum *et al.*, "Interaction of irritability and anxiety on emotional responding and emotion regulation: A functional MRI study," *Psychological Medicine*, vol. 51, no. 16, pp. 2778–2788, Jun. 2020, doi: 10.1017/s0033291720001397.