Alexandria Engineering Journal 84 (2023) 175–183

Contents lists available at ScienceDirect

Alexandria Engineering Journal

journal homepage: www.elsevier.com/locate/aej

Original Article

A novel approach to forecasting the mental well-being using machine learning

Alanazi Rayan *, Saad Alanazi

Department of Computer Science, College of Computer and Information Sciences, Jouf University, Sakaka, Saudi Arabia

ARTICLE INFO

Keywords:
Mental health problems
Machine learning
Neural networks
Random forest

ABSTRACT

Mental well-being is critical to an individual's health and quality of life. It encompasses emotional, psychological, and social dimensions, making it a complex and multifaceted construct. Traditionally, assessing and forecasting mental well-being has relied on self-report measures, clinical evaluations, and surveys, which can be subjective, time-consuming, and limited scope. To overcome these drawbacks and offer more precise and timely insights into mental well-being, this study presents a novel approach that uses the power of machine learning. This has been achieved by creating a comprehensive dataset containing various variables relevant to mental health. These variables include behavioral traits like exercise routines, sleep patterns, and social interactions and psychological traits like mood, stress levels, and emotional states. Environmental factors, such as geographic location, climate, and accessibility to mental health care, are also considered. A comprehensive understanding of the elements affecting mental well-being is made possible by this diversified dataset. The data is analyzed using machine learning models, such as deep learning neural networks, support vector machines, and random forests. These models were selected since they can capture intricate correlations and patterns in data, making them suitable for forecasting mental health.

1. Introduction

A person's mental health is a combination of his or her mental state and a valuation of his or her whole disposition. Brain chemistry abnormalities are the cause of depression. Identifying and giving treatment for people with aberrant mental behavior needs a detailed inspection of mental well-being [1]. An individual's psychological state is a barometer for how well they are being treated for their illnesses. A person's mental health is a combination of his or her state of mind and a valuation of his or her entire personality. Anomalies in serotonin levels characterize depression. Treating the underlying individuals with unstable mental conduct necessitates thoroughly investigating their mental health [2]. The psychosomatic disorder of a person serves as a barometer for how successfully their ailments are being treated [3]. Psychologists devote more time to therapy and use behavior interventions to assist patients with mental and emotional problems. Psychologists are also competent to provide psychological evaluations to diagnose a person's psychological disease and develop the most effective treatment strategy. The created prediction method will aid psychologists in psych evaluations and forecasting an individual's mental well-being. Both psychologists and psychiatrists collaborate to address the physician's behavioral and medical vital signs [4].

Professional organizations, agencies, or services, as well as the public at large, utilize a variety of languages about mental well-being. It is a common misunderstanding that mental well-being requirements and mental health issues are synonymous. While all children have mental health issues, not all adults do. Children suffer from depression. 'Emotional stability,' 'sentimental intelligence,' and 'mental, sentimental, and psychological health' are some of the terminologies used to characterize mental well-being [5]. Psychological abuse (PA) is an under-recognized and under-reported by adolescents and young according to their own families. It is regarded as the most challenging and widespread kind of childhood abuse and neglect. One PA is harder to define and quantify when compared to physical abuse. Verbal abuse, severe non-material punishments, and the threat of abusive behavior are examples. It is defined as the occurrence of elderly contact (usually including a parent) that makes the child feel worthless, defective, unwanted, undesired, threatening, or only valuable in servicing the needs of others [6].

Alcoholism and non-suicidal self-injury (NSSI) are important public

* Corresponding author.
E-mail addresses: sanazi@ju.edu.sa (A. Rayan), rmalanazi@ju.edu.sa (S. Alanazi).

https://doi.org/10.1016/j.aej.2023.10.060
Received 22 July 2023; Received in revised form 15 October 2023; Accepted 30 October 2023
Available online 9 November 2023
1110-0168/

health issues that predominantly affect children and young people. Alcohol intake is characterized as consuming enough alcohol in one sitting to raise one's blood alcohol level to a certain level (BAC) above 0.08 percent, which is especially dangerous when started during adolescence. Binge drinking begins in this important era, which influences neural development and is linked to many bad psychological outcomes later in life. Emotional, psychological, and social well-being are all influenced by how people expect, perceive, and act [7]. It also influences how people deal with stress, interact with others, and make decisions. According to the findings, mental well-being prediction is crucial at each phase of life, from pre-teens to adolescents to maturity. It is defined by a high level of impairment, such as mood and anxiety disorders caused by the affective disorder. At some point in their lives, one in every four people will undoubtedly suffer from mental or brain illnesses. Mental illness affects 25 % of persons in both developing and underdeveloped nations. According to a Health Organization estimate, 7.5 percent of the global population suffers from depression.

Machine learning technologies are rapidly being employed in mental health for various objectives, including forming diagnoses, predicting illness progression, recommending the best course of therapy, processing data on community psychological health, and managing medical notes. Many mental diseases have been studied using machine learning, including depression, stress, sleep disorders, schizophrenia, and autism. Machine learning techniques can enhance clinical practice [8]. For instance, the psychotherapist was able to recognize the most effective discussion tactics using a mixture of image discussion modeling, language model comparison, communication clustering, and frequent word analyses. It's been proven that therapists who get better outcomes are more likely to exhibit positive views, discuss the future, identify the main issue faster, and summarize and understand the client's feelings.

Adolescence is a stage when the body is changing rapidly. It causes an increase in adipose tissue, especially in girls, which may be considered undesirable given the slender body ideal. Adolescence is also a stage of development when peer interactions, reputation, and membership in peer groups are extremely important, as well as a period when unfavorable peer assessments and peer victimization are common. Weight and form are qualities that can lead not only to bad peer judgments but also to negative self-perceptions and excessively anticipating unpleasant reactions from others [9]. As a result, adolescence is the developmental age during which weight-related worries, both perceived and actual overweight, are most closely linked to depressed mood. Behavioral (externalizing) and Emotional (internalizing) disorders and symptoms, on either side, frequently overlap in adolescence and share risk factors. Morphological behaviors should indeed be treated when investigating the correlates of emotional changes, and emotional behaviors should be controlled when studying the correlates of behavioral and psychological symptoms to properly comprehend the relationships between diseases and potential risk factors [10].

This study's research represents a ground-breaking method for the difficult challenge of predicting mental health utilizing cutting-edge machine learning algorithms. Mental well-being is a significant and diverse aspect of total health and quality of life. However, traditional approaches to assessing and monitoring it frequently rely on arbitrary self-reporting or clinical assessments, which have limitations in accuracy, speed, and scalability. In contrast, this study presents a novel framework for predicting and forecasting mental well-being that uses state-of-the-art machine learning models, such as deep learning neural networks and ensemble techniques. This study's broad and varied datasets, which cover various elements influencing mental health, from psychological and emotional states to behavioral patterns and environmental variables, are one of its most remarkable characteristics. These datasets provide a thorough understanding of the complex interactions between the many factors that affect a person's mental health. The main contribution of this study is the creation and assessment of predictive models that are especially suited to mental health. These models can spot early warning signs of mental health problems and find subtle patterns and connections in the data that more conventional approaches would miss. This study presents a data-driven and proactive method for anticipating mental well-being through these machine learning models, which could have broad ramifications. It makes prompt interventions and individualized care possible for people, perhaps improving outcomes for those dealing with mental health issues. The investigation further develops mental health research by offering important insights into the intricate interplay between psychological, behavioral, and environmental factors, illuminating the dynamics of mental well-being. The use of machine learning in a data-driven framework promises to improve the results for both individual's and communities' mental health, making this research a significant step forward in anticipating and improving mental well-being.

2. Related work

Research on behavioral modeling for mental health based on the machine learning approach deals with analyzing the changes in the person's behavior and recognition using machine learning techniques [11]. Social anxiety, stress, obsessive–compulsive disorder, depression, personality problems, and drug addiction are all aspects that relate to psychiatric conditions. Identifying the onset of mental disease has become progressively significant to maintaining a decent life equilibrium. ML algorithms and AI can be utilized to fully exploit the nature of machine learning approaches and AI for forecasting the onset of mental problems. When used in real-time, this technology can help society by acting as a surveillance tool for individuals who engage in questionable activity. This study proposes employing a variety of ML algorithms, including decision trees, support vector machines, K-nearest neighbor classifiers, regression models, and naive Bayes classifiers, to evaluate the mental well-being status of a target population. The Mean Opinion Score assessed the categories that emerged from the clustering process. These clustering tags were used again to produce categories that could predict a person's mental health. The participants were separated into three categories: high school students, university graduates, and working professionals. The study investigates the impact of the abovementioned ML algorithms on the target groups and gives development has increased.

The Adolescent mental health illness related to back and neck pain problems were discussed in the research proposed by [12]. Adolescents suffer from a high prevalence of mental health issues. Furthermore, this demographic has a significant prevalence of spinal discomfort. According to the research, these problems appear to be linked. This research aimed to expand on previous outcomes by looking at the link between mental health illness, as indicated by the Child Behavior Checklist (CBCL), and teenage back and neck discomfort. The West Australian Prenatal (Raine) Study included 1,580 individuals (mean age 14.1 years) who supplied cross-sectional spinal pain and CBCL data. This cohort had a significant back and neck pain prevalence, as expected. Females, on average, reported higher mental health issues than males. The preponderance of CBCL sensation scales had substantial links to back and neck pain. Higher CBCL scores were linked to a higher risk of comorbid back and neck discomfort. These outcomes highlight the need to consider pain and psychological symptoms when assessing and treating adolescents. More study is needed to develop a hypothesized relationship.

To recognize the symptoms of depressive and anxiety disorder, machine learning in mental health is proposed by [13]. Machine learning algorithms can retrieve essential information about a person's mental well-being from unprocessed texts, such as social media posts and counseling session interview transcripts. Machines have been trained to detect the existence of mental disorders thus far, but they must still learn to recognize physical problems to make a proper diagnosis. This study aims to progress a machine learning model to distinguish individual anxiety and depressive illness symptoms. The researcher gathered 1065 notifications regarding depression and anxiety from online psychological forums, divided them into 7149 copies, and then categorized each replication using DSM-5 criteria. Due to an unbalanced dataset, researchers could not accurately distinguish the complete range of disorders. A two-stage model has been created, first recognizing big groups of depression, anxiety, and irritation. It defined sub-classes of symptoms in the second phase, such as low mood, suicidal intention, and negative self-talk in the depression category and excessive worry and social anxiety in the anxiety category. The study showed that diagnosis of mental diseases can be extracted from unstructured information in a huge dataset.

The research on the mental health problems for adolescents and children in school settings proposed by [10]. For this investigation, controlled studies, *meta*-analyses, and randomized and non-randomized randomized studies were found using a classification technique in the PubMed, PsycInfo, and Google Scholar databases. 1–6 % of the population may be affected by hyperkinetic syndrome. The most prevalent indications are agitation, focus issues, and lack of impulse control. Learning disabilities such as dyscalculia and dyslexia afflict 4–6 % of children, whereas depression affects 4–5 % of adolescents and teenagers, with women half as likely to be depressive as boys. Mental health worries raise the probability of completing a semester, dropping out of school, and being absent. Changes in the educational system and the implementation of proof-of-concept school programs can minimize the probability of an externalizing or internalizing mental disorder.

The refugee adolescents with mental health problems proposed by [14]. Because separation and migration from parents are generally acknowledged as the main risk factors for adolescent mental health, this research focuses on evaluating the mental health issues of split refugee adolescents to their accompanying peers in Belgium. A total of 1,294 adolescents—10 % of whom were refugee teenagers kept separate from their parents—completed three self-report survey questions on the pervasiveness of traumatic events, depression or anxiety symptomatology, externalizing problems, and post-traumatic stress disorder. Detached migrant teenagers are more likely to have traumatic experiences and build significant psychological problems, according to this research, which stresses the importance of fathers' accessibility to adolescents dealing with migratory experiences. Government actions should treat these teenagers primarily as "minors" instead of as "refugees," and receiving and care institutions should deliver more curative treatments and adequate preventive to these at-risk populations.

Because the COVID-19 epidemic began in early 2020, the significance of prompt and accurate mental health evaluation has significantly grown. This tendency is expected to persist in the post-pandemic timeframe due to enhanced mental disease chances. Few Asian populations have been studied using machine learning categorization of mental well-being. This research aims to make trustworthy ML classification algorithms based on health behavior markers that may be used with Southeast Asian college students. This study designs mental health using information from a sizable, multi-site cross-sectional questionnaire. It assesses the effectiveness of different ML algorithms, including generalized linear models, naive Bayes, k-nearest neighbor, random forest, neural networks, bagging, boosting, and recursive partitioning. Different criteria, including accuracy, sensitivities, failure rate, specificity, Gini Index, and AUC, were used to assess the forecasting model. The techniques of adaptable enhancing and random forest identified negative psychological well-being features with a higher degree of accuracy. Body mass index, the quantity of athletic activities a person participates in each week, sedentary hours, grade point average (GPA), and age are the five top characteristics most strongly linked to the prediction of poor mental health. Several particular comments and ideas for further research are explored in light of the provided results. These results could assist with providing affordable assistance and modernizing mental health assessment and supervising at the personal and institutional levels. The prediction accuracy in this method is a difficult challenge. [15].

The Internet of Things (IoT), a constantly connected and technologically integrated object network, has prohibited focused mental health studies. Decision-makers have looked to technology to see what possibilities it could present. They recognize that mental health problems are alarmingly increasing, impacting society and the individual gradually, and that existing HRs are insufficient to address the crisis. The COVID-19 epidemic, whose effects included not only severing the already brittle live and personal interactions among doctors and their patients but also the beginnings of a widespread mental well-being crisis as a result of the virus' effect on well-being and the developed measures to manage it, has given this endeavor higher significance than ever. The evaluation and management procedures may be thought of as the two complementing methods that comprise the function that IoT-enabled technologies perform in this new environment of digitized mental well-being. A person's physiological, behavioral, cognitive, and emotional levels, as well as the circumstances in which they reside, all play a role in tracking, knowing about, and identifying their difficulties with mental health. The next step is intervention, which adheres to the details of an evaluation and aims to improve a person's behavior and attitude. Artificial intelligence, in particular, allows for highly individualized evaluation and action due to technology. Omnipresent technology Smart wristbands enable ever-more precise evaluations, allowing for improved treatments and those that can be supplied instantly via a smartphone's intelligence cognition assistant, with the continual exchange of both being the largest advancement. This study is ineffective because it represents an early attempt to standardize this multidisciplinary research connection between technology and mental well-being. The scientific field remains young and has not yet been systematized into a consistent framework or even provided a summary of strategies, patterns, and directions [16].

The study examines how the HR environment, namely employment independence, career prospects, active communication, and effective action, supports the mental well-being of public workers. The study extends the literature by establishing organizational identification (OID) as the primary component by which the human resource environment might promote workers' well-being. OID is referred to as a "public remedy" due to its close connection to worker well-being and health. The findings of the structural model indicate a favorable relationship between HR practices and OID, which consequently improves the well-being of workers in the public sector. Even though effect and cause linkages cannot be completely examined, the cross-sectional methodology reduces the trust in the results. Secondly, there is also worry about common-method bias raised by the self-reporting measurements of the HR context and OID [17].

The expansion of psychology studies is concentrating its attention on studying psychological wellness. The commonly used Psychological Well-Being Scales (PWBS), which has frequently come under fire for incompatibilities between the conceptual and the empirical model, report measurement issues as one of the major difficulties. Network models, which describe psychological processes as implementing technology of constantly related variables, may be a viable option to comprehend the structure of mental health. The study investigated the network model of the Spanish 29-item PWBS in randomly selected specimens of 1,404 individuals using exploration graph modeling. The graphic LASSO approach calculated a regularized partial correlation structure at the element and dimensional levels. The research examines the robustness of both networks and determines the key network parameters. According to the PWBS network structure, there are four aspects: environmental mastery, reason for living, and self-acceptance. According to node strength centralization, Self-acceptance was shown to be the most central consideration in the PWBS-measured mental well-being framework. The network structure of mental well-being offers a fresh concept of mental well-being. It suggests target markers for psychosocial interventions, although it does not replicate the theoretical basis of Ryff's framework. This research was ineffective because it only looked at the system structure of one mental well-being measurement that had been translated into one language, used in one cross-sectional, non-clinical group, and had a bias toward women [18].

The mental health issues in adolescence and childhood were developed by [19]. Detached migrant teenagers are more likely to have traumatic experiences and build significant psychological problems, according to this research, which stresses the importance of fathers' accessibility to adolescents dealing with migratory experiences. Government actions should treat these teenagers primarily as "minors" instead of as "refugees," and receiving and care institutions should deliver curative treatments and more adequate preventive to these at-risk populations. The mental well-being and most wells of children and young adults are critical. Unmet mental health needs in preschool contribute to adolescent issues and adult issues. Providing thorough prevention, early detection, and intervention strategies is critical. A comparison of various mental health systems is presented in Table 1 below.

3. Methodology

Researchers employed supervised learning, in which a subset of the information was labeled to build the machine, and the labeled data was used to forecast new information classes. This technique takes longer and requires more effort, but it is more efficient than unsupervised classification, which involves the system starting to learn to distinguish classes from large datasets. They gathered 1065 posts from a Russian psychology forum to construct a dataset. Forum admins have previously written the messages into sections such as depressive illnesses (459 messages) and anxiety symptoms (444 messages) (606 messages). The common experience length for mood was 205 characters, and 161 characters for stress. Each replication indicated a whole thought, phrase, or sentence, and each reference was broken into 7149 replicates [25]. Extracted features and labeling were carried out manually by psychiatrists. They were given tasks that required them to categorize the characteristics of major depressive disorder and anxiety disorders using the statistical and diagnostic Manual of Mental Disorders categories (DSM-5). They were also invited to propose new categories of commonly discussed topics. As a result, classes for postnatal depression, loneliness, and confidence issues have been introduced. The following Table 2 Symptom Frequency mentioned in Depression Category and the Table 3 Symptoms Frequency in the Anxiety Category are mentioned below with clear details.

Table 2
Symptom frequency mentioned in depression category.

Depression (D_x)	Symptom	No of Specification	Specification Share
D_1	Hopeless and Sadness	234	15,49 %
D_2	Pleasure or Loss of interest	158	10,57 %
D_3	Lack of energy or Tiredness	143	9,45 %
D_4	Negative talk and worthless feelings	126	8,32 %
D_5	Thought of suicide or self-harm	107	7,19 %
D_6	Lonesomeness	101	6,65 %

Table 3
Symptoms frequency mentioned in anxiety category.

Anxiety	Symptom	No of specification	Specification Share
A_1	Panic attack	114	7,31 %
A_2	Social anxiety	94	5,99 %
A_3	Health anxiety	87	5,56 %
A_4	Sleep problems	58	3,80 %
A_5	Irritability	57	3,63 %
A_6	Inability to control worry	56	3,56 %

An overview of symptom frequency for the anxiety category may be seen in Table 3. It displays a range of anxiety symptoms designated as A_1 to A_6, such as "panic attack," "social anxiety," "health anxiety," "sleep problems," "irritability," and "inability to control worry." The table lists the frequency of mention of each symptom, with "Panic attack" appearing 114 times. It also lists the proportion of all specifications that each symptom accounts for, with "Panic attack" making up 7.31 % of all specifications. This information is helpful for future analysis or study on anxiety-related topics since it provides a simple reference for determining the prevalence of various anxiety symptoms within the category.

The information was separated into a training dataset and a specimen in a 70/30 ratio to run the investigations. The data was represented using a bag of word models and SVM classifiers from the sci-kit-learn

Table 1
Comparison of existing methods for predicting mental well-being.

S. No	Reference	Dataset	Method/Tools	Results	Disadvantage
1.	[20]	840 adolescents from 6 public and private high schools in Spain	Pearson bivariate correlations, Descriptive analysis, and Hierarchical regression	Emotional competence is a protective factor for the best adaptation and well-being, and evidence suggests that self-esteem strengthens this relationship.	Because the specimen was selected through convenience sampling, it is challenging to generalize the findings.
2.	[21]	2835 academic students	Hierarchical regression analyses and Descriptive statistics	It improves students' satisfaction with their education and stops and lessens mental tension.	It requires a very time-consuming and difficult calculation and analysis procedure.
3.	[22]	340 college students, including 222 females and 118 males	Flourishing Scale, Satisfaction with UCLA Loneliness Scale and Life Scale, Rosenberg Self Esteem Scale,	Self-esteem is a mediator in the relationship between well-being and loneliness, with loneliness being a powerful indicator of well-being and self-esteem markers.	There aren't a lot of respondents, and demographic factors are not considered in the research.
4.	[23]	610 academic professionals from three institutions in Norway	Job demands-resources (JD-R) theory and Technology acceptance model (TAM)	Workers must be familiar with the technologies' features, giving the best tools for those who use them as their sole means of communication and productivity.	Using cross-sectional information to forecast a worker's well-being at work. As a result, it is not authorized to demonstrate causal relationships among variables in this study, and just one employee is the subject of this investigation.
5	[24]	70.5 % females and 29.5 % of males from four teacher education institutions in Sabah, Malaysia	partial least square-structuralequation modeling	Self-efficacy, subjective well-being, and emotional intelligence are all important personal resources that can forecast resilience.	The structured model utilized in this research had a poor predictive accuracy, at 19.5 %, for describing practicum stress. This research focused mostly on personal resources, which are insufficient to identify practicum stress.

module so the algorithms could identify similarities. Researchers used the term frequency technique to evaluate the texts after splitting them into unigrams and bigrams. The technique allows you to rank each token's relevance to other tokens for a given class [13]. The main premise is that if a word appears frequently in one class but infrequently in another, it is extremely important for this sample. As a result, regularly used terms are given less weight, while uncommon ones are given more. The algorithm was tested to see if it could provide false positives and false-negative outcomes. The tests were conducted on word embedding learned in the same literature using the pre-trained word technique. The same results were obtained, and accuracy did not improve.

3.1. Mental health based on Machine learning

Mental illnesses affect people of all ages and localities, significantly contribute to the global disease burden, and have substantial economic, environmental, and human liberties ramifications. The greatest disparities, however, are cross-national: while getting just 10 % of global funding for mental health, 80 percent of people with mental diseases live in low- and middle-income nations. The goal of national and community initiatives is to improve the availability, accessibility, and quality of mental health therapies for individuals around the globe. Diagnosing classifications and classifications of mental disorders are needed by multiple stakeholders to achieve access to mental health-specific goals: for medical providers to make treatment decisions and execute clinical practice guidelines [26], for lawmakers to make strategic choices, and for patients and families to develop an understanding of one's disorders. The following is Fig. 1 Categorization of Mental Health Problem.

Based on the above Fig. 1, mental health problems could be caused mainly by anxiety, and they could be categorized into three types based on mood, and three diagnoses could be determined. To make the prediction easier, a Machine learning approach is used so the mental health problem can be easily predicted and treated to prevent problems due to mental health. Machine learning is considered the best tool for classification or prediction modeling. Supervised machine learning presents the benefit of compensating for intricate interactions between parameters that may not have been detected earlier. As information gets larger and characteristics become more complicated, machine learning techniques may become an important tool in psychiatry for correctly disentangling variables linked to health outcomes [14].

In psychiatry, the preponderance of machine learning researchers has concentrated on categorization or prognosis. However, critics have claimed that these research findings are likely to underachieve, leading to a shortage of analysis of the different machine learning tricks of the trade's basic assumptions and mental conditions and diagnosing procedures, trying to highlight the difficulties in developing and validating such models. Nonetheless, the sector has made progress, with tree-based algorithms used to anticipate suicide in teenagers and the United States Army [19]. Tree-based methods offer data on how widely a factor is used for analysis or differentiation, giving considerable insight into the models' categorization process and their sufficient reliability. This suggests that, while the road ahead is difficult, correctly implemented machine learning approaches can increase the usefulness of medical decision-making. This research aims to create an archetypal that can forecast mental health issues in adolescents. Researchers also want to look into several machine learning approaches and conventional regression models to see which performs best with mixed survey and registration information. As per the "No Free Lunch Theorem," researchers anticipate that the approaches deployed will produce similar results.

3.2. Measures

The Strengths and Difficulties Questionnaire was used to gather information on adolescent mental health concerns under fifteen. Parent-reported psychological conditions, conduct difficulties, prosocial behavior, hyperactivity, and peer relationship difficulties were all assessed using the SDQ. A binary attribute was created using an amalgamation of the parent mentioned, excluding prosocial behavior, subscale, and a cut-off score validating for the country population, which is needed to be taken into account, corresponding to approximately 10 % of the population able to score far above cut-off and thus being classified as having mental health issue [27]. Forecasters were gathered from a questionnaire distributed via CATSS and registrations on 9/12 or later. Researchers include a wide range of variables based on past results of connection with teenage psychological health and/or adolescent mental well-being. Physical disease, environmental factors, birth information, and mental health symptoms like neighborhood and family income were all used as predictors. Both registry and parental recorded data were used as respondents.

A crucial component of creating a mental health forecasting model is the database employed in this study. Beyond the accepted constraints, it is crucial to examine the statistical analysis of the dataset's parameter values carefully. Descriptive statistics are often used to analyze the features of the data, correlation analysis is used to find possible links between parameters, feature relevance is determined, and multivariate interactions are explored. However, the existence of non-linear correlations between the chosen parameters is a crucial factor that the authors should consider, particularly in mental health. Mental well-being is intrinsically complicated since emotional, psychological, and environmental components frequently have non-linear interactions. The authors should use the proper techniques, such as non-linear models,

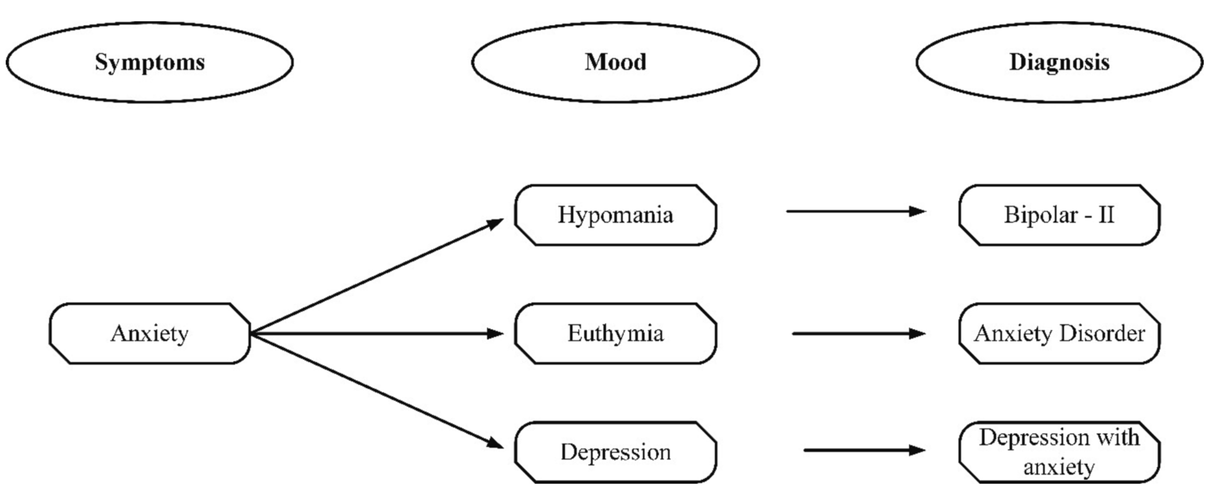

Fig. 1. Categorization of mental health problems.

feature engineering, or non-parametric testing, to identify and capture these non-linear dependencies to deal with this complexity. The ability of a predictive model to effectively predict mental well-being may be compromised if non-linear correlations are not considered. So, for the research to be effective, non-linear correlations must be fully considered.

The above Table 4 Prediction Technique shows the different techniques for predicting mental health problems and other health issues.

3.3. Data processing

The analysis was omitted for data with a missing rate greater than 50 % (202 variables). Unnecessary variables were also eliminated (134 excluded). Factors with no variation (32 eliminated) were also deleted, but those with variability approaching zero were consolidated into one variable if practicable; for example, mold, dust, and pollen allergies were compressed into allergy [28]. After all was said and done, 85 factors were found to be eligible for the study. Because most ML approaches need to complete datasets, missed values were imputed using the R package mice's tree-based impute.

3.4. Statistical analysis

Research data was divided into three sets: a training set (60 percent of the specimen), a tuning set (10 percent), and a testing set (10 percent) (30 percent). Splitting data enables a more precise assessment of how the technique will perform in a new dataset and reduces overfitting, which occurs when the training set is fitted too closely to reliably predict additional information. Stratified random sampling was utilized to guarantee that the twin pairs were not dividing for both databases to reduce the dimensionality. In particular, with respect, the researcher made sure that each set had an equal distribution of results. Qualitative data were constructed for each set to determine the quality of the division [29]. The region underneath the receiver operating characteristic was used to assess the accuracy of predictions using the algorithms in question (AUC). To assess which machine learning technique produced some of the best suitable solutions for a testing dataset, the researcher used XGBoost, random forest, support vector machine (SVM), neural network, and logistic regression to develop a forecasting model. Considered to determine the partition's grade.

Each strategy used a classification technique to train different products with the training data before testing their effectiveness on a subset of the training set. The tune combination was then used to test the model with the lowest standard deviation [30]. The completed models were then attached to the testing sample if the behavior in the tune set was considered satisfactory.

4. Result and discussion

The datasets were judged to be sufficiently isolated, as mentioned in Table 5 Partitioned Data Information and the graphical representation are mentioned in Fig. 2 Partitioned Data. The researchers mitigated the effects of this by using SMOTEBoost to do various things on the training dataset. The investigation categories were fairly unbalanced because just 12 % of the study sample passed the cut-off. The training indicated that, except for the neural network that requires considerable data preprocessing, such as scaling and centering on the dependent variables, the algorithms now functioned effectively without any additional info or hyper-parameters.

Table 4
Prediction Technique.

Prediction Method	Detail
Logistic Regression	In epidemiological, regression analysis is the usual method for assessing binary outcomes.
Random Forest	Decision trees are a framework that organizes information into a tree-like architecture using if-then-else logic.
Neural Network	Multiple interacting processing, or "neurons," are structured in different levels of a neural network: input, hidden, and output.
Support Vector Machine	Support Vector Machines divide categories, such as instances and no cases, using a line called a higher-dimensional space.

4.1. Sensitive analysis

The SDQ has several proposed cut-offs based on various factors and sample populations, which the researcher utilized to assess the quantity of variation in the study. Because the 11-point cut-off was determined based on capturing the top 10 % of a Swedish sample, it's possible that this cut-off doesn't always reflect a distinct subset of psychology, causing the model's performance to decrease. To investigate if the used cut-off affected model performance, the researchers formed an innovative approach utilizing the highest performance strategy with a stricter cut-off from the original paper [31]. The 17-point cut-off was chosen to catch the top 10 % of scores in a UK sample in the first edition. The graphical representation of the AUC curve test dataset is shown in Fig. 3.

The chance factor was set at 0.8, meaning that when the likelihood of belonging to the class was higher than 0.2, the program classified them as exhibiting mental health difficulties. The research-specialized edition is available. The predicted accuracy was 96 percent, while the negative accuracy of the model was 15 percent. This translates to a sensitivity of 0.91 and specificity of 0.30, with 15 % of the test set classified as positive.

4.1.1. Tuning model

The researcher next fitted models using all of the strategies that were investigated; the AUCs from the tune-set of the model summary table for every approach are shown in Fig. 3 Tune Dataset for AUC Curve and Fig. 4 Test Dataset for AUC Curve. Table 6 Tune Dataset Performance of Model offers a comprehensive list of the best characteristics and their values for every design. There was no clear winner; moreover, support vector machine (SVM) and random forest had the highest AUCs of 0.754 (95 percent CI 0.700–0.800 and 95 percent CI 0.697–0.805, correspondingly). With an AUC of over 0.700, the rest of the models fared equally.

The performance of multiple machine learning models is shown in Table 6 using the Area Under the Curve (AUC) as the main performance metric on the Tune dataset. XGBoost, SVM, Random Forest, Logistic Regression, and Neural Network are the models being compared. SVM and Random Forest both get an AUC of 0.755, demonstrating good discriminatory ability, although AUC values vary from 0.714 to 0.755. Additionally, 95 % confidence intervals are given for the AUC values obtained by bootstrapping, assisting in determining how certain these performance estimations can be. This table is an invaluable resource for model selection and assessment on the Tune dataset. The graphical representation of the Tune Dataset for the AUC Curve is shown in Fig. 4.

4.2. Discussion

This report predicts adolescent mental well-being relatively effectively, with a maximal AUC of 0.739 on the testing set, employing a vast range of information from parent surveys and registration information from numerous Different state registrations. The AUC suggests a good model but isn't precise enough for therapeutic application. The negative prediction accuracy of this model is 95.5 percent, indicating clinical-level sensitivities. However, the positive predictive value is only 13 %. This means that only a tiny fraction of the youngsters who have been flagged will meet the Research mental health cut-off, which should be compared to the 10 % incidence in the population.

The random forest model's estimation method revealed that perhaps the program didn't trust any variable. Hence, the technique would be

Table 5
Partitioned Data Information.

	N	Year of Birth	Sex	Cutoff
		Standard deviation	Male %	Reached value of Cutoff
Training Dataset	4555	1995.6	48.5 %	12.2 %
Tune Dataset	805	1995.3	49.7 %	12.3 %
Test Dataset	2281	1995.5	48.2 %	11.6 %

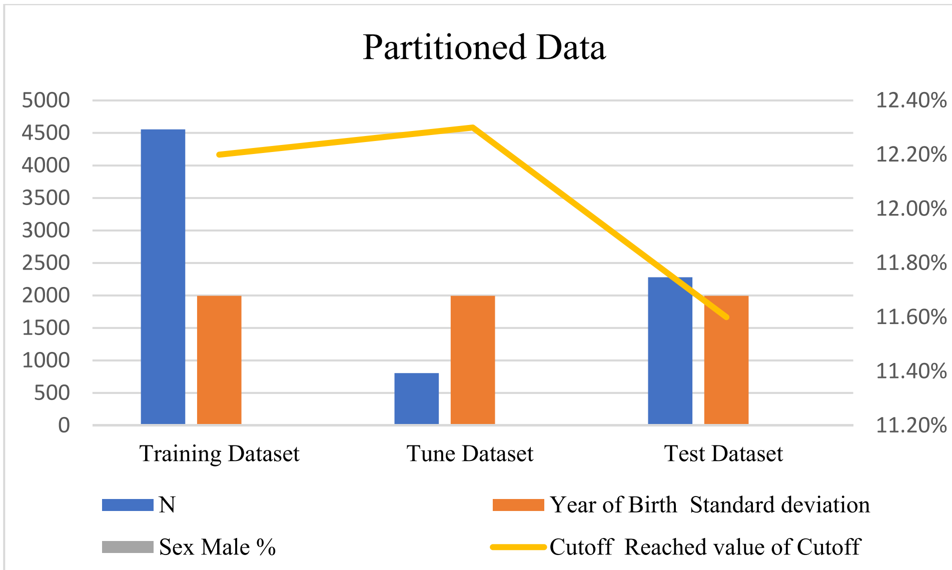

Fig. 2. Partitioned data.

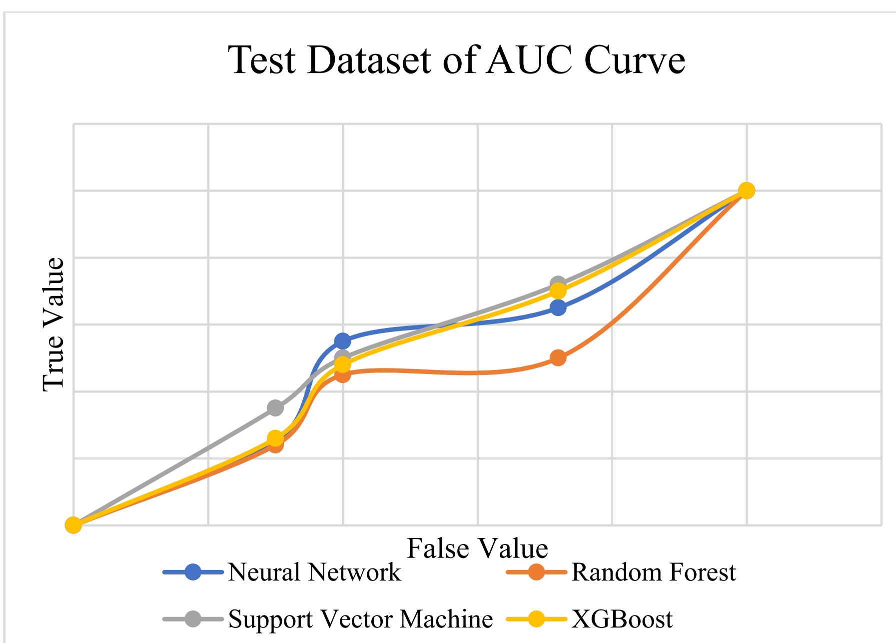

Fig. 3. Test Dataset of AUC Curve.

generally stable if any one parameter were removed, particularly steady ones over time. At the age of 15, parental-reported psychological healthcare outcomes such as impulsivity, inattention, and depressive problems were revealed to be strong predictors of psychological problems. Information on the neighborhood's healthiness, parity, and the maternal delivery age was also deemed necessary. These findings are from prior studies and could be utilized by therapists, families, or instructors to recognize at-risk children for therapy [32].

The greatest characteristics were either caregiver or easily recorded by families, implying that register data that can be complicated and costly for investigators might not have been required for a good mental model structure. As a result, future research forecasting teenage

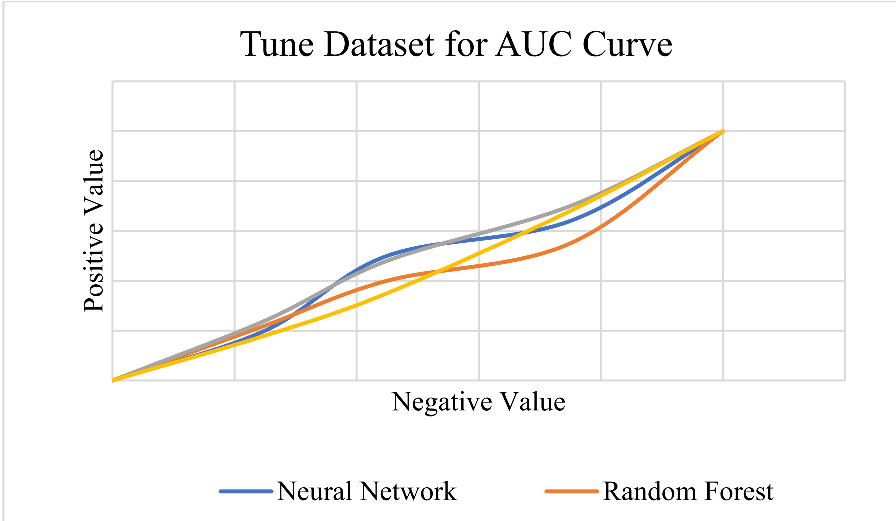

Fig. 4. Tune Dataset for AUC Curve.

Table 6
Tune Dataset Performance of Model.

Method	AUC	Bootstrap interval 95 %
XGBoost	0.723	0.663–0.779
SVM	0.755	0.700–0.800
Random Forest	0.755	0.699–0.805
Logistic Regression	0.749	0.692–0.804
Neural Network	0.714	0.657–0.768

psychological health might also want to focus more on caretaker assessments. Furthermore, this encourages parents to participate in physicians' assessments of their adolescents' and children's psychiatric diagnostic and psychological health. Additional research with comparable goals should use symptomatic evaluations for mental health problems, particularly neurological abnormalities, as their paradigm.

Sensitive findings indicate that a much more aggressive cut-off improved prediction effectiveness marginally but not substantially. This means that the following research can utilize verified cut-offs for their country or the originating investigation, whichever they prefer. This also shows that the more exceptional examples do not constitute a separate serious category. The extensive investigation of various factors related to teenage psychological health is one of the article's advantages. Furthermore, the use of self-report measures suggests that these risk variables can be identified by non-clinicians, implying a low-priced future approach for massive psychological screening. The findings must be interpreted in the context of several obstacles. Firstly, research results may not apply to singletons, but the researcher used a twin group, which may have factors recognized in contrast to single people. Prior research has shown little variation in mental health among twin and single-parent families.

In conclusion, research predictions had an acceptable AUC, although none performed statistically significantly better than the others. Granting supervised machine learning approaches generates a lot of buzz in the scientific community; most research doesn't need to skip regression analysis, particularly if the databases are short and the connections are mostly constant. Furthermore, research findings support the routine screening of people's neuro-developmental syndrome and training challenges for subsequent psychiatric risks. Even though machine learning algorithms appear encouraging for further integrating risks beyond various fields for forecasting psychological problems in adolescence, clinical use would be unwarranted. However, because proper intervention for that and other psychological symptoms significantly reduced adverse consequences and manifestations, there is hope for adolescent psychological problems being prevented with correctly planned change.

5. Conclusion and future work

The significance of several significant cautions related to this research must be acknowledged in light of these findings. The results' generalizability to a larger population of young people, particularly those in various high schools or community contexts, is still questionable. The study's limited relevance is due to its dependence on a small sample of young graduates. The alignment of these self-reports with objective measurements of practical problem-solving ability is another issue raised by using self-rated evaluations for problem-solving capability. Despite showing promise, the predictive model created in this study is not yet ready for clinical application. It should be viewed as a first step in building stronger models that predict psychological health outcomes. To provide a more thorough picture of mental health variables, future studies should think about including parent-rated evaluations. Additionally, the decision to use logistic regression as the modeling method calls for a close examination and researchers should consider the advantages of applying more complex techniques in similar studies. However, based on observed symptoms, this prediction model can be used to spot early indications of mental health issues and launch preemptive therapies. In the future, incorporating digital technologies, such as web platforms and smartphone applications, can significantly improve the precision and accessibility of mental health prediction and assistance. Technology integration into social-emotional learning programs to promote psychosocial development and well-being among individuals could be a viable area for future study and interventions, given the rising prevalence of mental stress. This study offers important new information about the early detection of mental health problems in young people, especially recent high school graduates. Although there are drawbacks, including sample size and modeling technique selection, the study is a crucial first step in improving our comprehension of mental well-being prediction. As digital technologies advance, there is an increasing chance to improve and broaden these prediction models, resulting in more effective interventions and better results for young people's mental health in various contexts.

There are several significant limitations to the study and its predictive model, mostly brought on by data-related restrictions and the complexity of mental health. The first drawback is the study's use of historical data, which could introduce biases and exclude changing social norms or new mental health issues. Prediction accuracy may also be impacted by problems with data quality, such as recall bias and social desirability bias in self-reported data. Furthermore, the findings' generalizability may be constrained because they might not immediately apply to all demographic groups or varied cultural situations. It is vital to take into account the database's constraints. The availability of data, including the completeness and detail of some essential aspects like genetic predisposition or medication history, may be constrained, making it more difficult for the model to consider all pertinent factors. Additionally, the database may not sufficiently reflect recent changes in mental health patterns and therapies, limiting its timeliness if it largely comprises older data. The combination of these restrictions may ultimately impact the accuracy of the suggested model. Predictions that do not adequately reflect the genuine state of mental well-being may result from biases and generalization problems. Models that fall short of fully capturing the complexity of mental health can result from incomplete data due to missing or insufficient data on crucial variables. Researchers should prioritize eliminating data biases, enhancing data quality, and considering additional external data sources to increase accuracy. The model's ability to predict mental well-being will depend on developing advanced methods for handling imbalanced data, careful feature selection, and continuing model validation and adaptation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number 223202.

References

[1] S.-J. Blakemore, Adolescence and mental health, Lancet 393 (10185) (2019) 2030–2031.
[2] W. Bor, A.J. Dean, J. Najman, R. Hayatbakhsh, Are child and adolescent mental health problems increasing in the 21st century? A systematic review, Australian & New Zealand J. Psychiatry 48 (7) (2014) 606–616.
[3] M. Srividya, S. Mohanavalli, N. Bhalaji, Behavioral modeling for mental health using machine learning algorithms, J Med Syst 42 (5) (2018) 88, https://doi.org/10.1007/s10916-018-0934-5.
[4] R. Chambers, J. Belcher, Predicting mental health problems in general practitioners, Occup. Med. 44 (4) (1994) 212–216.
[5] J. Chung, J. Teo, Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges, Applied Computational Intelligence and Soft Computing 2022 (2022).
[6] G. Schulte-Körne, "Mental Health Problems in a School Setting in Children and Adolescents", Deutsches Ärzteblatt, International (2016) 9.
[7] C.S. Rees, A.J. Smith, P.B. O'Sullivan, G.E. Kendall, L.M. Straker, Back and neck pain are related to mental health problems in adolescence, BMC Public Health 11 (1) (Dec. 2011) 382, https://doi.org/10.1186/1471-2458-11-382.
[8] O. Troitskaya and A. Zakharov, "Machine Learning in Mental Health: Recognizing the Symptoms of Depressive and Anxiety Disorders," PsyArXiv, preprint, Nov. 2021. doi: 10.31234/osf.io/edpnj.
[9] I. Derluyn, C. Mels, E. Broekaert, Mental Health Problems in Separated Refugee Adolescents, J. Adolesc. Health 44 (3) (Mar. 2009) 291–297, https://doi.org/10.1016/j.jadohealth.2008.07.016.
[10] H.A. Rahman M. Kwicklis M. Ottom A. Amornsriwatanakul K.H. Abdul-Mumin M. Rosenberg I.D. Dinov Prediction Modeling of Mental Well-Being Using Health Behavior Data of College Students Research Square. https:// 2022 doi.org/ 10.21203%2Frs.3.rs-1281305%2Fv1.
[11] Kolenik, Tine. 2022. "Methods in Digital Mental Health: Smartphone-Based Assessment and Intervention for Stress, Anxiety, and Depression." In *Integrating Artificial Intelligence and IoT for Advanced Health Informatics*, edited by Carmela Comito, Agostino Forestiero, and Ester Zumpano, 105–28. Internet of Things. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-91181-2_7.
[12] A. Elhadad, F. Alanazi, A.I. Taloba, A. Abozeid, Fog Computing Service in the Healthcare Monitoring System for Managing the Real-Time Notification, Journal of Healthcare Engineering 2022 (2022).
[13] A. Blasco-Belled, C. Alsinet, The Architecture of Psychological Well-Being: A Network Analysis Study of the Ryff Psychological Well-Being Scale, Scand. J. Psychol. (2022), https://doi.org/10.1111/sjop.12795.
[14] McDougall, Tim. "Mental health problems in childhood and adolescence." Nursing Standard (through 2013) 26, no. 14 (2011): 48.
[15] U. de la Barrera, K. Schoeps, J.-A. Gil-Gómez, I. Montoya-Castilla, Predicting Adolescent Adjustment and Well-Being: The Interplay between Socio-Emotional and Personal Factors, Int. J. Environ. Res. Public Health 16 (23) (2019) 4650.
[16] Franzen, Jessica, Françoise Jermann, Paolo Ghisletta, Serge Rudaz, Guido Bondolfi, and Nguyen Toan Tran. "Psychological Distress and Well-Being among Students of Health Disciplines: The Importance of Academic Satisfaction." *International Journal of Environmental Research and Public Health* 18 (4): 2151, 2021, doi: 10.3390/ijerph18042151.
[17] A.I. Taloba, M. Abd El-Aziz, H.d.M. Rasha, Alshanbari, and Abdal-Aziz H. El-Bagoury, "Estimation and Prediction of Hospitalization and Medical Care Costs Using Regression in Machine Learning." Journal of Healthcare, Engineering (2022).
[18] M. Shamsi, T. Iakovleva, E. Olsen, R.P. Bagozzi, Employees' Work-Related Well-Being during COVID-19 Pandemic: An Integrated Perspective of Technology Acceptance Model and JD-R Theory, Int. J. Environ. Res. Public Health 18 (22) (2021) 11888, https://doi.org/10.3390/ijerph182211888.
[19] Taloba, Ahmed I., Ahmed Elhadad, Alanazi Rayan, Rasha M. Abd El-Aziz, Mostafa Salem, Ahmad A. Alzahrani, Fahd S. Alharithi, and Choonkil Park. "A blockchain-based hybrid platform for multimedia data processing in IoT-Healthcare." Alexandria Engineering Journal (2022).
[20] M H Marghny, Rasha Abd M El-aziz and Ahmed I Taloba. Article: Differential Search Algorithm-based Parametric Optimization of Fuzzy Generalized Eigenvalue Proximal Support Vector Machine. International Journal of Computer Applications 108(19):38-46, 2014, doi: 10.5120/19023-0540.
[21] M. Pandey, D. Parmar, and S. Mishra, "Mental Health Prediction for Juvenile Using Machine Learning Techniques," p. 10, 2019.
[22] A.E. Tate, R.C. McCabe, H. Larsson, S. Lundström, P. Lichtenstein, R. Kuja-Halkola, Predicting mental health problems in adolescence using machine learning techniques, PLoS One 15 (4) (Apr. 2020) e0230389.
[23] U.M. Butt, S. Letchmunan, M. Ali, F.H. Hassan, A. Baqir, H.H.R. Sherazi, Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications, Journal of Healthcare Engineering vol (2021) 2021.
[24] T. Verbeek, et al., Postpartum depression predicts offspring mental health problems in adolescence independently of parental lifetime psychopathology, J. Affect. Disord. 136 (3) (2012) 948–954.
[25] J. Zheng, Z. Yu, A Novel Machine Learning-Based Systolic Blood Pressure Predicting Model, J. Nanomater. 2021 (2021).
[26] J. Worland "Predicting Mental Health", The invulnerable child 1987 185.
[27] K. von Simson, I. Brekke, and I. Hardoy, "The Impact of Mental Health Problems in Adolescence on Educational Attainment," Scandinavian Journal of Educational Research, pp. 1–15, Jan. 2021, doi: 10.1080/00313831.2020.1869077.
[28] S. F. I. Rizvi, "PARENTAL PSYCHOLOGICAL ABUSE AND MENTAL HEALTH PROBLEMS IN ADOLESCENTS," Pak J Med Sci, vol. 30, no. 2, Dec. 1969, doi: 10.12669/pjms.302.4593.
[29] V. Lankinen, S. Fröjd, M. Marttunen, R. Kaltiala-Heino, Perceived rather than actual overweight is associated with mental health problems in adolescence, Nord. J. Psychiatry 72 (2) (Feb. 2018) 89–96, https://doi.org/10.1080/08039488.2017.1389987.
[30] A. El-Aziz, M. Rasha, A.I. Taloba, F.H.A. Alghamdi, Quantum computing optimization technique for IoT platform using modified deep residual approach, Alex. Eng. J. 61 (12) (2022) 12497–12509.
[31] A.I. Taloba, M.R. Riad, H.A. Taysir, Soliman, Developing an efficient spectral clustering algorithm on large scale graphs in spark, in: In 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), 2017, pp. 292–298.
[32] K.C. Ravikumar, N. Pandi Chiranjeevi, M. Devarajan, C. Kaur, A.I. Taloba, Challenges in internet of things towards the security using deep learning techniques, Measurement: Sensors 24 (2022), 100473.