ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor (RJIF): 8.4
IJAR 2023; 9(11): 61-66
www.allresearchjournal.com
Received: 22-07-2023
Accepted: 29-08-2023

Madhvi Patwa
Assistant Professor,
Department of Mathematics,
Rizvi College of Engineering,
Bandra, Mumbai,
Maharashtra, India

Dipti Pimputkar
Assistant Professor,
Department of Mathematics,
Rizvi College of Engineering,
Bandra, Mumbai,
Maharashtra, India

Corresponding Author:
Madhvi Patwa
Assistant Professor,
Department of Mathematics,
Rizvi College of Engineering,
Bandra, Mumbai,
Maharashtra, India

Demystification of mathematical modeling of anxiety disorder based on demographic factors in context of machine learning techniques

Madhvi Patwa and Dipti Pimputkar

Abstract
The mental risk poses a high threat to the individuals, especially overseas demographic, including expatriates in comparison to the general Arab demographic. This paper focuses on a comprehensive analysis of mental health problems such as depression, stress, anxiety, isolation, and other unfortunate conditions. The dataset is developed from a web-based survey. The detailed exploratory data analysis is conducted on the dataset collected from Tamil Nadoo to study an individual's mental health and indicative help-seeking pointers based on their responses to specific pre-defined questions in a multicultural society. The proposed model validates the claims mathematically and uses different machine learning classifiers to identify individuals who are either currently or previously diagnosed with depression or demonstrate unintentional "save our souls" (SOS) behaviors for an early prediction to prevent risks of danger in life going forward. The accuracy is measured by comparing with the classifiers using several visualization tools. This analysis provides the claims and authentic sources for further research in the multicultural public medical sector and decision-making rules by the government.

Keywords: Artificial intelligence, cloud, machine learning, blockchain, mental illness, neural network models, machine learning, data mining

Introduction
Anxiety disorders are among the most prevalent and debilitating mental health conditions worldwide, affecting millions of individuals and imposing a substantial burden on both the affected individuals and society at large. The complex and multifaceted nature of anxiety disorders has prompted researchers to explore innovative approaches for better understanding, diagnosing, and treating these conditions. One such approach that has gained significant attention in recent years is the application of mathematical modeling, coupled with machine learning techniques, to elucidate the intricate relationship between anxiety disorders and demographic factors. Mathematical modeling, a methodology long employed in various scientific disciplines, has emerged as a powerful tool for analyzing the multifarious components of anxiety disorders. These models provide a structured framework for comprehending the intricate interplay of factors contributing to the onset and exacerbation of anxiety, including genetic predisposition, environmental influences, and individual-specific characteristics. Furthermore, the integration of machine learning techniques within this context offers an unprecedented level of precision in extracting insights from large-scale datasets, allowing researchers to uncover hidden patterns, predict risk factors, and tailor personalized interventions. This endeavor, the demystification of mathematical modeling of anxiety disorders based on demographic factors, represents a pivotal juncture in the quest to enhance our understanding of these conditions. The goal is to elucidate how demographic variables such as age, gender, socioeconomic status, and geographic location intersect with individualized experiences, genetic predispositions, and environmental triggers to influence anxiety disorders. By discerning these intricate relationships, we can foster more effective preventive measures, develop targeted treatment strategies, and alleviate the individual and societal burden of anxiety disorders. In this exploration, we will delve into the core concepts, methodologies, and emerging trends in the field of mathematical modeling applied to anxiety disorders.

We will also investigate the various demographic factors that influence the manifestation and progression of anxiety and how machine learning techniques can facilitate the development of predictive models that can assist clinicians in diagnosis and treatment decisions. Our quest for demystification will not only empower healthcare professionals with valuable insights but will also provide hope to those affected by anxiety disorders, offering the promise of more personalized and effective interventions. Anxiety disorders, a significant category of mental health conditions, affect a substantial portion of the global population, including the Indian population. Anxiety disorders in the Indian population present unique challenges and characteristics influenced by a combination of cultural, societal, and individual factors. Efforts are being made in India to address these challenges by improving mental health services, raising awareness, and reducing stigma. Nevertheless, there is still a long road ahead to ensure that individuals with anxiety disorders in India receive the care and support they need. Understanding the unique cultural and societal factors at play is essential for tailoring effective interventions and support systems for the Indian population dealing with anxiety disorders. Efforts are being made in India to address these challenges by improving mental health services, raising awareness, and reducing stigma. Nevertheless, there is still a long road ahead to ensure that individuals with anxiety disorders in India receive the care and support they need. Understanding the unique cultural and societal factors at play is essential for tailoring effective interventions and support systems for the Indian population dealing with anxiety disorders.

Relevant studies

The influence of artificial intelligence in mental health has a vital role in developing and designing the systems applying machine learning and deep learning to find the facts related to social intelligence and human-computer interaction (HCI) covering the theme. This research brings out claims and justifications by referring to the more profound insights into the current landscape of machine learning applications for mental health from the two domains of HCI and computing science. Several machine learning and computational pipelines have been developed to standardize predicting a mental state and cognitive behaviour in advance. These pipelines have stimulated the behavioural data (i.e., reaction time, decision, mood swings, etc.) and the empirical results, and offer solid behavioural predictions. Furthermore, computational models can be trained on the data to extract latent features, and be used to predict the different behavioural actions in the environment. Thus, these computational models are gaining traction in metal state prediction. A wide range of literature is available to understand the correlation between cognition and emotion basic principles of psychopathology and emotions and the neural enactment of mental processes. Considering other domains, computational models are frequently used in the fields of biology and health care to identify reasonable outcomes of processes and are increasingly applied to social and cultural psychology to develop long-term demographic-level computational models. The different models are taken as agent-based and considered complex social influences, putting their implications for societal and cultural evolution and generating the testable hypothesis which readily challenges easy, logical analysis. Computational modelling and natural processing language have been used to predict and monitor mental disorders such as depression, anxiety, etc. Several kinds of research have been done in sensor, audio, video, structures, and multimodal system use, with different mental health behaviour having been explored, such as depression, suicide, stress, mood, bipolar, post-traumatic stress disorder (PTSD), anxiety, substance abuse, schizophrenia, and other mental health conditions. The author explored association rule mining to identify depression and anxiety using the Internet-based cognitive behavioural therapy (ICBT) program in patients. The linguistic features are considered for finding better correlations and detecting outcomes in a specific context of inpatient mental health. The paper has stated the different classes of artificially intelligent systems for identifying the fairness of systems for research on people with disabilities. A complete roadmap to explore the opportunities for researchers to create systems exploiting the power of artificially intelligent systems has been described. Nemesure *et al*. (2018) [19] developed a prediction model to diagnose anxiety in psychiatric assessment using an electronic health record (EHR) dataset containing undergraduate students; hybrid machine learning and deep learning models are considered in the pipeline and explored biometric and demographic data for predicting psychiatric illness using various non-psychiatric input features. The model was tested, while the results were validated on different information retrieval metrics to determine which features are essential for the prediction of diseases. Sharma and Verbeke (2022) [20] investigated the variable importance hierarchy of biomarkers for anxiety disorders. The various univariate and multivariate models are experimented with to find correlations among the four anxiety disorders—generalized anxiety disorder (GAD), agoraphobia (AP), social anxiety disorder (SAD), and panic disorder (PD) in the Dutch citizens' dataset. The key elements of the previous research studies are as under:

Research Methodology

In this research, a comprehensive analysis of mental health conditions and unintentional SOS behavior will be conducted for different individuals in Tamil Nadoo state. The survey will generate this dataset based on multiple-choice questions consisting of discrete and continuous variables. The text-based inputs from the survey fillers will be encoded using the label encoders or one hot encoder defined under the scikit learn package of python. Also, an EDA will be conducted on the dataset to identify the characteristics that will help in early prediction. The analysis will have four data points and four inferences generated, classified into two classes, one from the machine learning models and the second from data analysis, for example using a mathematical model. The first inference will have the machine learning model results with three points, namely supervised, unsupervised, and semi-supervised learning algorithms. The reduced support vector machine (RSVM) and logistic regression will be used to give good results for text-based data. The hierarchical clustering will be used unsupervised due to the abrasiveness of the dataset, which will have high abrasiveness. Also, the K-nearest neighbor (KNN) will be used as it deals with the classification problem, which can be modelled into a clustering problem. Capsule neural networks will be used for supervised learning methods, while different integral, customized and activation functions will be used to optimize the accuracy. The results obtained from the different state-of-the-art machine learning classifiers will be analyzed, with the best possible results to be generated from the inference. The second inference generated from the mathematical model by looking at the secondary differential question or order of two linear equations derived will be studied from the data points, with the features being taken from the dataset. These mathematical results will develop the prediction model to fit the machine learning model. Several information retrieval metrics counting for the values of precision, recall, F1-score, accuracy, and R^2 error [42] will identify the subject as diseased or not. The above-stated models will be validated to claim whether the two inferences are synchronized.

Data collection

Data is collected through a survey. Data collection is done so that there is no inherent bias in the dataset. The collected data is then cleaned for any blank or NA values. Exploratory data analysis is done to understand the dataset. A relationship study between various features is done in a reverse "ablation" study using boxplots. Various outputs are mentioned in file EDA, while the baseline inference is drawn from this step, which will later work as our validator. The data is split into k=3 random groups for analysis purposes. Therefore, a three-fold cross-validation is done, in which the model is trained and tested three times with three different datasets (which is a subset of the primary data).

Data Preparation

The data has been prepared by conducting a survey considering the local and international demographics in Tamil Nadoo. The dataset consists of 1,000 different people based on 27 different features. Once the data is collected, fundamental exploratory data analysis (E.D.A.) shall be performed. The analysis is conducted to find the relevant patterns in the dataset using several supervised, semi-supervised, and unsupervised algorithms. Based on the analysis, a mathematical model is also drawn, which will again be re-run to indicate individuals suffering from depression based on the analysis done by the proposed model.

Experiment Analysis

The analysis of early prediction is done on the survey dataset. Data is collected from 1,000 individuals using a survey format targeting those in their 20s to 40s. The features and characteristic of data used for the study include residential status (object), gender (object), age (numeric), academic status (object), proficiency of mother tongue (numeric), proficiency in Arabic (numeric), relationship status (binary), religious belief (binary), frequency-little interest or pleasure in doing things (object), frequency-feeling down, depressed, or hopeless (object), frequency-trouble falling or staying asleep, or oversleeping (object), frequency-feeling tired or having little energy (object), frequency-poor appetite or overeating (object), frequency-feeling bad about yourself or that you are a failure or have let yourself or your family down (object), frequency-trouble concentrating on things, such as reading the newspaper or watching television (object), frequency-suicidal thoughts (object), disassociation from world (object), loneliness (object), lack of friends (object), social discrimination (object), opportunistic discrimination (object), racial discrimination (object), financial status (binary), frequency-self-harm (numeric), fear-failure (object), fear-family disappointment (object), frequency-harming others (numeric). The analysis is performed by training the model using machine learning classifiers.

Exploratory Data Analysis of the Dataset

The four basic python libraries are used, such as NumPy (for running different kinds of running matrix algorithms, panda (for processing and data import and export), seaborn (for advance data visualization) and matplotlib (for basic data visualization) are used for conducting EDA [43]. After applying inputs, the shape of the data is checked. The data contains 1,000 rows and 27 columns. The data has been augmented, trained and tested on different models further. The data has been read and the data type checked, where most of the data types observed in the survey data are of the object type. The data count has been done, and the zero (null values) are discarded. The relationships between the various factors in the dataset are identified for analysis. The box plots have been used to check for the correlation between various datasets. The first box plot in Fig. 3 reveals that most of the density of people who participated in the survey belong to the 20-35 years old demographic. The relationship between age, academic position, and the frequency of suicidal thoughts have also been identified as people with maximum suicidal thoughts being in the 40-45 years old age group, along with graduate studies. The inference drawn from Fig. 4 reveals that the people in middle management of the corporate sector only with bachelor's degrees are highly depressed. A disturbing fact that comes out of the above plot is that people having an undergraduate degree and relatively younger in age tend to be suicidal. Another depressing fact has revealed that the people who have not yet graduated also have negative thoughts of suicide and can be considered depressed. The mapping of suicidal thoughts and their financial status has been done, revealing that the above financial status is more perceived as showing suicidal tendencies.

Experimental Setting

The experiments are conducted in two different ways as follows. The EDA is conducted on the taxonomic hierarchical survey data. Different modules as inbuilt packages in Python language [44] have been used for implementation. The matrix analysis is done using NumPy, the plots and complex graphs are made using matplotlib and seaborn, the data is ingested using pandas, and the libraries sklearn and Tensor Flow are used for machine learning tasks. The hardware requirements taken for experimentation were an Inter i5 8th Gen processor, 16 GB RAM, and CentOS operating system. The model was developed on the longitudinal data, while training of the model was performed to learn compact representation, with encoding of the dynamics attained by observing the longitudinal measures for each subject. This work contains hyper-parametric classifiers from KNN, semi-supervised algorithms, and capsule networks. Fig. 10 presents the accuracy at each iteration by considering 27 features in the dataset.

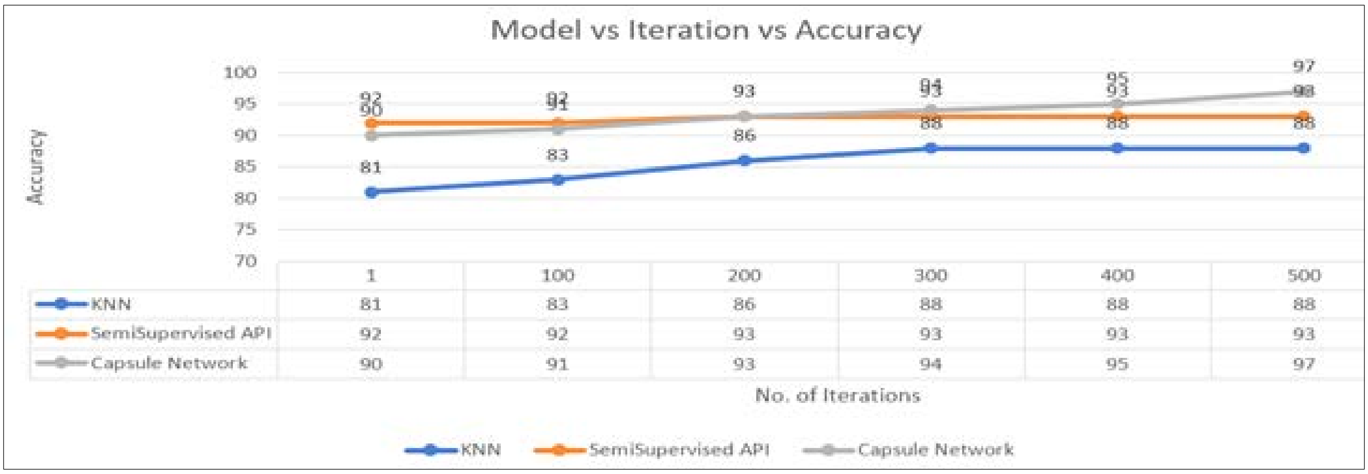

Fig 1: Accuracy at each iteration

Ablation Study

The ablation study is performed on the different datasets by excluding some features. The sets for this study are taken as *P*, *Q*, *R*, and *S*. A total of 26 features are considered and marked in set *P* for experiments. Likewise, set *Q* consists of the listed features in Table 2, excluding "education". Set *R* excludes the self-harm feature, while set *S* excludes two features, namely age + education + self-harm. It has been observed that we have received the result with 81% accuracy, excluding the "age" feature. After that, to highlight the dominant feature, experiments were repeatedly performed by removing the feature "education" it was observed that the accuracy raised to 83%. Then, the feature of self-harm dropped and the accuracy decreased to 79%. However, removing all three mentioned features gives more reduction and achieves a 71% accuracy. Thus, these experiments proved that the most dominant and highlighted feature is "education." The results on accuracy by calculating the average accuracy from the base and modified data are shown in Table 2.

Table 1: Evaluation accuracy as per the ablation study

Ablation study	Average accuracy with base data (%)	Accuracy with modified data (%)
Features (sans) age	92.67	81
Features (sans) education	92.67	83
Features (sans) self-harm	92.67	79
Features (sans) age + education + self-harm	92.67	71

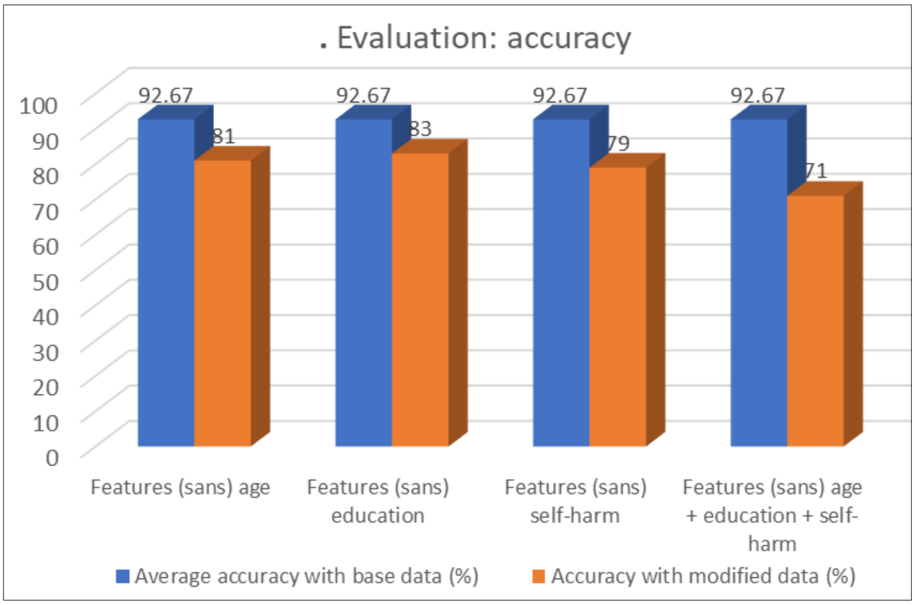

Fig. 2: Evaluation accuracy as per the ablation study

Evaluation Metrics

The proposed approach is validated using various evaluation metrics such as precision, recall, F- measure, and accuracy. The proposed model is trained and tested using a ratio of 80:20. Table 3 demonstrates the results of the proposed work. The experiments are conducted on the survey dataset, achieving an above 80% accuracy in all three models. This accuracy is adequate as we have surveyed related works extensively and observed that the accuracy for users' responses is below 75% in existing research works to the best of our knowledge. Precision measures the number of positive labels predicted in the actual positive class.

Table: 2: Showing the Evaluation metrics of present work.

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F-score (%)
Un-supervised	88.50	82	84	82
Semi-supervised	93.38	96	95	95
Supervised	98.29	99	98	98

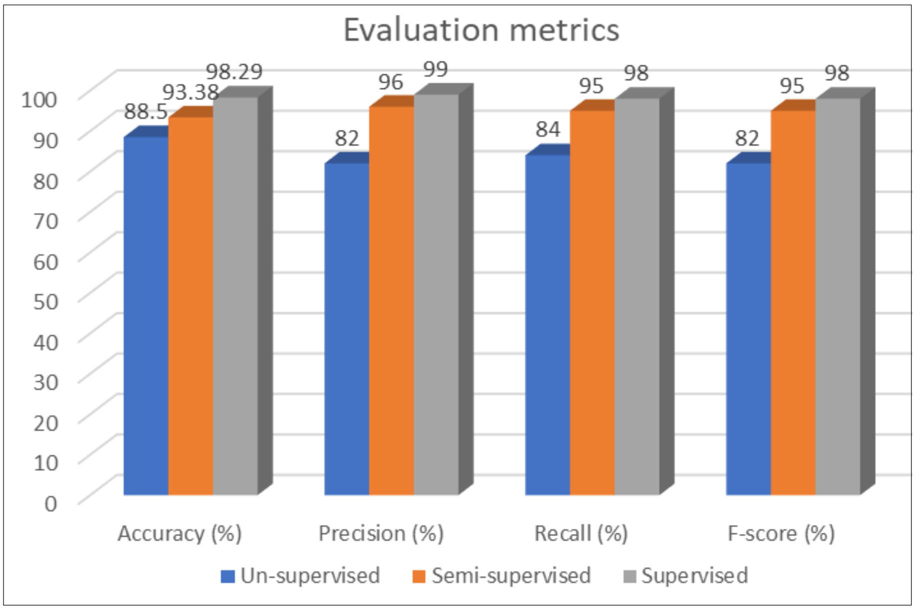

Fig 3: Evaluation: accuracy as per the ablation study.

A receiver operating characteristic (ROC) curve is also calculated with the different classifiers used in the experimentation, providing the best results with supervised learning.

Conclusion and Future Direction

In this study, mental illness such as depression has been studied by experiment using machine learning and deep learning classifiers. Machine learning models have been implemented using packages like Tensor Flow, Keras, etc. In addition to the above, the help of various medical documentation, such as PHQ-9, Social Connectedness Scale, etc., is taken to accurately find a correlation between various variables identified from the dataset. The significant outcomes are based on the standard data collection protocols taken from the survey with questions including measurement scales, medical history, etc. The validations are done using the mathematical model, which proved that the inferences are aligned with that of the mathematical model because the underlying mathematical philosophy of our software platform coincides with the mathematical model. Also, a clinical setup can be used for training machine learning algorithms, which can help health professionals predict the correct response to the data in identifying anxiety disorders. Health professionals can interact with the clinical decision support system. The research conducted in this study helped diagnose the early prediction of depression by using computational techniques. The results are validated using a mathematical model, and any claims will be justified by measuring the accuracy of the results using machine learning classifiers. The dataset consists of records from both international and local individuals, which will be used to examine their mental health conditions, including expatriates in a multicultural environment. The paper well-stated the correlation between various essential features using analysis and conducting the ablation study on the dataset. The graphs and figures are used as visualization tools to justify the effectiveness of the proposed model. In the future, the study can be further investigated by the GAN model on other hospital datasets taken from Tamil Nadoo. Also, a cloud-based data parser and an accurate time assessment of a person's mental health can be introduced to get early medical assistance.

References

1. Balakrishnan A, Kadiyala R, Dhiman G, Ashok G, Kautish S, Yadav K. A personalized eccentric cyber-physical system architecture for smart healthcare. Security and Communication Networks; c2021. p. 1747077. https://doi.org/10.1155/2021/1747077
2. Beiter R, Nash R, McCrady M, Rhoades D, Linscomb M, Clarahan M, *et al*. The prevalence and correlates of depression, anxiety, and stress in a sample of college students. Journal of Affective Disorders. 2015;173:90-96.
3. Dardas LA, Bailey DE, Simmons LA. Adolescent depression in the Arab region: a systematic literature review. Issues in Mental Health Nursing. 2016;37(8):569-585.
4. Dardas LA, Silva S, Noonan D, Simmons LA. Studying depression among Arab adolescents: methodological considerations, challenges, and lessons learned from Jordan. Stigma and Health. 2018;3(4):296-304.
5. Ebert JF, Huibers L, Christensen B, Christensen MB. Paper- or web-based questionnaire invitations as a method for data collection: cross-sectional comparative study of differences in response rate, completeness of data, and financial cost. Journal of Medical Internet Research. 2018;20(1):e8353. https://doi.org/10.2196/jmir.8353

6. Eldar E, Rutledge RB, Dolan RJ, Niv Y. Mood as representation of momentum. Trends in Cognitive Sciences. 2016;20(1):15-24.
7. Fletcher JA, Doebeli M. A simple and general explanation for the evolution of altruism. Proceedings of the Royal Society B: Biological Sciences. 2009;276(1654):13-19.
8. Forstmann BU, Wagenmakers EJ. An Introduction to Model-Based Cognitive Neuroscience. New York, NY: Springer; c2015.
9. Grahek I, Musslick S, Shenhav A. A computational perspective on the roles of affect in cognitive control. International Journal of Psychophysiology. 2020;151:25-34
10. Jones KH, Jones PA, Middleton RM, Ford DV, Tuite-Dalton K, Lockhart-Jones H, *et al*. Physical disability, anxiety and depression in people with MS: an internet-based survey via the UK MS Register. PloS One. 2014;9(8):e104604. https://doi.org/10.1371/journal.pone.0104604
11. Lee MD, Criss AH, Devezer B, Donkin C, Etz A, Leite FP, *et al*. Robust modeling in cognitive science. Computational Brain & Behavior. 2019;2(3):141-153.
12. Nadakinamani RG, Reyana A, Kautish S, Vibith AS, Gupta Y, Abdelwahab SF, *et al*. Clinical data analysis for prediction of cardiovascular disease using machine learning techniques. Computational Intelligence and Neuroscience; c2022. p. 2973324. https://doi.org/10.1155/2022/2973324
13. Rayhan RU, Zheng Y, Uddin E, Timbol C, Adewuyi O, Baraniuk JN. Administer and collect medical questionnaires with Google documents: a simple, safe, and free system. Applied Medical Informatics. 2013;33(3):12-21.
14. Shah SM, Al Dhaheri F, Albanna A, Al Jaberi N, Al Eissaee S, Alshehhi NA, *et al*. Self-esteem and other risk factors for depressive symptoms among adolescents in United Arab Emirates. PloS One. 2020;15(1):e0227483. https://doi.org/10.1371/journal.pone.0227483
15. Singh SK, Azzaoui AE, Kim TW, Pan Y, Park JH. DeepBlockScheme: a deep learning-based blockchain driven scheme for secure smart city. Human-centric Computing and Information Sciences. 2021;11:12. https://doi.org/10.22967/HCIS.2021.11.01
16. Slewa-Younan S, McKenzie M, Thomson R, Smith M, Mohammad Y, Mond J. Improving the mental wellbeing of Arabic speaking refugees: an evaluation of a mental health promotion program. BMC Psychiatry. 2020;20314. https://doi.org/10.1186/s12888-020-02732-8
17. Wang W, Xu H, Alazab M, Gadekallu TR, Han Z, Su C. Blockchain-based reliable and efficient certificate less signature for IIoT devices. IEEE Transactions on Industrial Informatics. 2022;18(10):7059-7067.
18. Wong JG, Cheung EP, Chan KK, Ma KK, Tang SW. Web-based survey of depression, anxiety and stress in first-year tertiary education students in Hong Kong. Australian & New Zealand Journal of Psychiatry. 2006;40(9):777-782.
19. Reiter MJ, Nemesure A, Madu E, Reagan L, Plank A. Frequency and distribution of incidental findings deemed appropriate for S modifier designation on low-dose CT in a lung cancer screening program. Lung Cancer. 2018 Jun 1;120:1-6.
20. Gupta GK, Sharma DK. A review of over fitting solutions in smart depression detection models. In2022 9th International Conference on Computing for Sustainable Global Development (India com) IEEE; c2022 Mar 23. p. 145-151.