**IAES International Journal of Artificial Intelligence (IJ-AI)**
Vol. 11, No. 3, September 2022, pp. 1108~1118
ISSN: 2252-8938, DOI: 10.11591/ijai.v11.i3.pp1108-1118


# Depression prediction using machine learning: a review

**Hanis Diyana Abdul Rahimapandi[1], Ruhaila Maskat[2], Ramli Musa[3], Norizah Ardi[4]**
[1]TESS Innovation Sdn Bhd, Petaling Jaya, Malaysia
[2]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Malaysia
[3]Department of Psychiatry, Kulliyyah of Medicine, International Islamic University Malaysia, Kuantan, Malaysia
[4]Academy of Language Studies, Universiti Teknologi MARA, Shah Alam, Malaysia

**Article Info**

***Article history:***

Received Aug 19, 2021
Revised Mar 11, 2022
Accepted Apr 9, 2022

***Keywords:***

Depression
Literature review
Machine learning
Prediction

**ABSTRACT**

Predicting depression can mitigate tragedies. Numerous works have been proposed so far using machine learning algorithms. This paper reviews publications from online electronic databases from 2016 to 2020 that use machine learning techniques to predict depression. The aim of this study is to identify important variables used in depression prediction, recent depression screening tools adopted, and the latest machine learning algorithms used. This understanding provides researchers with the fundamental components essential to predict depression. Fifteen articles were found relevant. We based our review on the systematic mapping study (SMS) method. Three research questions were answered through this review. We discovered that sixteen variables were deemed important by the literature. Not all of the reviewed literature utilizes depression screening tools in the prediction process. Nevertheless, from the five screening tools discovered, the most frequently used were hospital anxiety and depression scale (HADS) and hamilton depression rating scale (HDRS) for general population, while for literature targeting older population geriatric depression scale (GDS) was often employed. A total of twenty-two machine learning algorithms were identified employed to predict depression and random forest was found to be the most reliable algorithm across the publications.



***Corresponding Author:***

Ruhaila Maskat
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA
Shah Alam, Malaysia
Email: ruhaila@fskm.uitm.edu.my

## 1. INTRODUCTION

As the coronavirus (COVID-19) pandemic spread across the globe, it is causing a significant degree of fear and concern in the public. In terms of public mental health, elevated depression rates are the most significant psychological effect to date. Younger adults had higher mental health rates, while adults enduring serious health issues had more mental health problems [1]. The analysis showed that mental health problems decreased by 5% with every year's rise in age [2]. Children from lower socio-economic classes who were exposed to experiences of mental health problems early in their lives, be it due to both or either parent, were more likely to become mentally ill later in life. Mood disorders and suicide-related findings have soared over the past decade [3], [4]. According to the Institute for Public Health, mental health disorders among adults have increasingly become worrying from 10.7% in 1996 to 29.2% in 2015 [5].

Depression, the most common type of mental illness, is a psychological condition that happens to anyone at various ages due to specific reasons such as loss of self-esteem and social environment. The symptoms faced by depressed individuals may have a severe effect on their capability to deal with any condition in everyday life, which significantly varies from the usual mood variations. Depression affects not only physical but also psychological well-being [6]. It is associated with diabetes, hypertension, and back pain [7]. Besides that, a mental disease is often a burden in the form of tension, marriage breakdown, or homelessness for families, friends, caregivers, and other relationships [8]. Therefore, an initiative and commitment to prevention and treatment for depression are necessary.

Depression is one of the leading mental illnesses that is least diagnosed, considering the incidence and seriousness. The diagnosis and evaluation of signs of depression rely almost exclusively on data provided by patients, family members, friends, or caregivers [9]. This type of article, however, is inaccurate because it relies on the reporter's total integrity. Depression-related self-perceived shame is widespread in societies worldwide and is associated with unwillingness to seek professional assistance [10]. Patients are also hesitant to express their depressive feelings with physicians, so a discussion of depression often relies heavily on a general practitioner's willingness to engage with the patient. The prevalence of depression in Malaysia is considerably higher than in the United States and most other Western countries [11]. Depression is a severe mental illness and a significant public health issue that has a massive effect on society. In the worst case, depression can lead to suicide. Even though it is a severe psychological issue, fewer than half of people with this emotional problem have received mental health services [6]. It may be attributed to various reasons, including lack of knowledge of the disease. Additionally, researchers discovered that embarrassment and self-stigmatization tend to pose as more significant factors for not obtaining medical attention than others' actual prejudice and adverse reactions [12].

The capability to predict depression using machine learning algorithms before conditions worsen is essential. Therefore, in this paper, we conducted a systematic review of literature from 2016 to 2021 (time of writing) to help researchers better understand this area. This review aims to firstly, identify variables relevant to the prediction of depression using machine learning techniques, secondly, identify the latest and most frequent screening types used in detecting depression and finally, popular state-of-the-art techniques in machine learning to predict depression based on chosen metrics and values of performance.

Using machine learning techniques for the prediction of medical conditions is not new. Recent publications show applications in hepatitis [13], autism [14] and cancer [15]. Nevertheless, it is not without weaknesses. The primary weakness of any prediction pipeline involving machine learning techniques is the substantial dependence on correctly annotated data. If a dataset size is small, manually annotating each data point is feasible, however, in this big data era manual annotation of data has become impractical. Since machine learning techniques are trained on these annotations, a dataset with low-quality labels can result in unreliable predictions. Another weakness is the risk of overfitting. In the pursuit of achieving higher prediction performance, these techniques can develop a tendency to induce a model fitted to specific unique data points which do not represent a large portion of the population. Thus, rendering the models useless.

Our contribution via this study is a systematic review covering key aspects in predicting depression. Significant variables in previous works are identified, depression screening tools used are investigated and popular machine learning algorithms based on classical as well as new measurements of performance are highlighted.

The paper is outlined as follows: in section 2, the systematic literature review methodology is explained. Our proposed methodology and research questions are detailed in section 3. Then, the results of our review are presented in section 4. Finally, in section 5 we conclude this paper.

## 2. SYSTEMATIC MAPPING STUDY (SMS) METHOD

Systematic mapping study (SMS) method organizes published research and their results into structured categories by systematically perusing its primary contents, methodology and results with the aim of mitigating bias and concluding using statistical meta-analysis supported by evidence [16]. Although originally introduced for medical research, SMS method has been adapted for computing. Figure 1 shows the primary three phases of the SMS method used in our study. Each phase produces an outcome which in turn triggers the next phase.

The SMS method begins with the formulation of research questions so that the coverage of existing literature can be framed. Once the scope of the review has been determined, a search of the literature is conducted involving the definition of information sources from various academic online databases, digital libraries, and search engines. Exploration of these sources is performed using search terms constructed to encompass the earlier formulated research questions using Boolean operators. From all the papers extracted, screening based on keywords, abstract, introduction and conclusion sections are carried out to identify only relevant papers that can provide answers to the previous questions.

## 3. RESEARCH METHODOLOGY

In this section, we describe the steps on how we applied the SMS method to systematically review existing literature from 2016 till 2020 (time of writing). In each following subsection, we describe in detail the input, activity and output involved in each step. Finally, we illustrate the summarized evolution of paper filtration process to obtain the final relevant papers for review. These steps are defined research questions, literature search and screening papers.

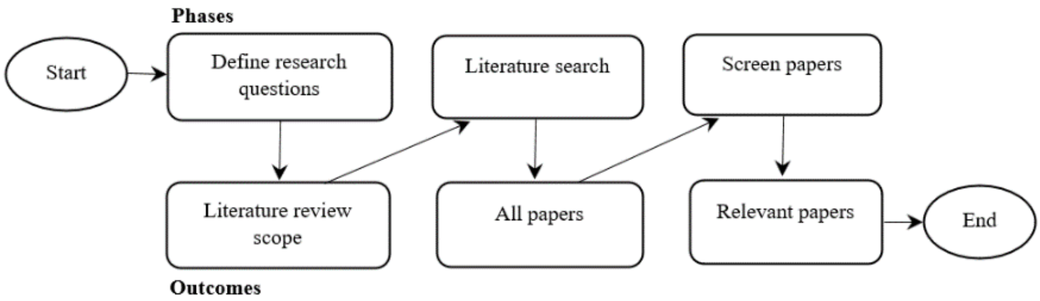


Figure 1. Phases of the SMS method

### 3.1. Define research questions

At this phase, research questions were formed to seek literature within the scope of predicting depression using machine learning methods. The first question is concerned with what variables were used by recent proposals for the prediction process. This answer allows researchers to identify relevant variables. A good selection of variables helps to produce good prediction performance. The second question is which depression screening tools were adopted. This question provides an understanding of a particular screening tool that has been continuously used by researchers and how many of the proposals are not utilizing any screening tools. From the answer to this question, researchers can decide the necessity of adopting specific screening tools into their work. The final question is what machine learning techniques were proposed by existing research? This question helps direct researchers to state-of-the-art machine learning techniques applied to depression prediction. Table 1 lists the constructed research questions and the motivations behind them.

Table 1. Research questions and motivation

| Research questions | Motivation |
|---|---|
| RQ1: What variables were used by recent proposals in predicting depression? | The answer to this question allows researchers to identify variables relevant to the prediction of depression. |
| RQ2: Which depression screening tools were adopted? | The answer to this question identifies the latest and most frequent screening types used in detecting depression. |
| RQ3: What machine learning techniques were proposed by existing research? | The answer to this question provides researchers with popular state-of-the-art techniques in machine learning to predict depression based on chosen metrics and values of performance. |

### 3.2. Literature search

A through search was conducted on four prominent electronic databases utilizing the following keywords: “depression prediction”, “mental health prediction”, and “anxiety, depression, and stress prediction”. The keywords were combined using Boolean AND expression and OR expression. The databases searched were: IEEE Xplore (http://ieeexplore.ieee.org), ACM Digital Library (http://www.portal.acm.org/dl.cfm), Elsevier ScienceDirect (http://www.sciencedirect.com), and Google Scholar (http://scholar.google.com).

### 3.3. Screening papers

The papers were examined based on their relevance to our constructed research questions. We analyzed the title, abstracts, and keywords to ascertain they lie within our focus of interest. Then, the papers were classified into two categories based on the following inclusion (*I*) and exclusion (*E*) criteria:

*I1: Paper should directly relate to depression prediction using machine learning techniques.*

*I2: Papers should provide answers to the research questions.*
*I3: Papers should contain at least one of the search keywords.*
*E1: Posters, panels, abstracts, presentations, and article summaries.*
*E2: Duplicates.*
*E3: Papers without full text.*

The initial collection of papers from all electronic databases yielded 73 papers. Since there exists an overlap due to the search on Google Scholar, duplicates were removed with a remaining of 50 papers. Next, 32 irrelevant papers were excluded after the title and abstract of each paper were perused. The resulting 18 papers were then fully read through and resulted in 3 found irrelevant whereas the rest of the 15 papers were included in this review. Figure 2 shows the screening process.

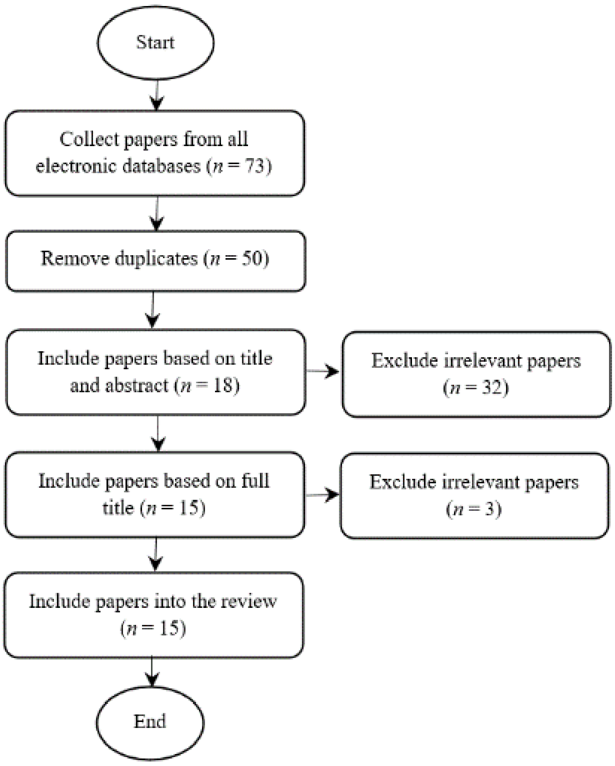


Figure 2. Paper screening process

## 4. RESULTS AND DISCUSSION

The 15 relevant papers included in this review is listed in Table 2 by year, source, the scope of prediction and number of citations. The list suggests that studies on depression prediction were actively conducted in 2020 (31%) and 2016 (25%). The former is most likely due to the COVID-19 pandemic whereas for the latter no prominent event could be linked. In relation to the number of citations, sources based on computing and technology received a large number of citations since they lead to the introduction of new techniques, whereas medical-centered sources are lesser cited, owing to their more general application of these new techniques. IEEE, a widely known online database, recorded the highest number of cited sources (ICHI and KDE).

### 4.1. RQ1: what variables were used by recent proposals in predicting depression?

To predict depression, the researchers use several types of datasets. Some of them predict depression using demographics and clinical attributes, some use social media to collect information by using text analytics, hence, benefits from textual features instead of attributes. The various common variables in depression prediction found in 6 of the relevant papers are presented in this section. Table 3 shows the demographics and clinical variables that had been used in past research. Based on previous studies, it is found that the most used variable is age and marital status, followed by gender, educational status, and socio-economic status. For clinical variables, diabetes has been used twice in previous studies, while others have only been used once and most of them in P3.

Table 2. List of relevant literatures

| Paper ID | Year | Reference | Source | Scope of prediction | Number of citations |
|---|---|---|---|---|---|
| P1 | 2016 | [17] | Biomedical Signal Processing and Control | Depression | 18 |
| P2 | 2016 | [18] | International Journal of Computer Applications | Depression | 21 |
| P3 | 2017 | [19] | Healthcare Technology Letters | Anxiety and depression | 34 |
| P4 | 2017 | [20] | Proceedings-2017 IEEE International Conference on Healthcare Informatics, ICHI 2017 | Depression | 7 |
| P5 | 2017 | [21] | Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence | Depression | 109 |
| P6 | 2018 | [22] | CEUR Workshop Proceedings | Depression and anorexia | 16 |
| P7 | 2019 | [23] | Informatics in Medicine Unlocked | Anxiety and depression | 32 |
| P8 | 2019 | [24] | Journal of Medical Internet research | Depression | 41 |
| P9 | 2019 | [25] | International Conference on Human Centered Computing | Depression | NA |
| P10 | 2019 | [26] | International Conference on Advances in Engineering Science Management and Technology (ICAESMT)-2019 | Anxiety, depression, and stress | 16 |
| P11 | 2020 | [27] | IEEE Transactions on Knowledge and Data Engineering | Depression | 85 |
| P12 | 2020 | [28] | Procedia Computer Science | Depression | 20 |
| P13 | 2020 | [29] | Doctoral dissertation, École de technologie supérieure-Superior Technology School | Depression | 1 |
| P14 | 2020 | [30] | Healthcare | Depression | 1 |
| P15 | 2020 | [31] | Journal of Affective Disorders | Anxiety, depression, and stress | 2 |

Table 3. Variables used by recent proposals

| Variable | P2 | P3 | P4 | P7 | P14 | P15 |
|---|---|---|---|---|---|---|
| 1. Age | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 2. Gender | ✓ | ✓ | ✓ | | ✓ | ✓ |
| 3. Residence status | | ✓ | ✓ | | | |
| 4. Educational status | ✓ | | ✓ | ✓ | | ✓ |
| 5. Marital status | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| 6. Income | ✓ | | ✓ | | ✓ | |
| 7. Employment status | | ✓ | | ✓ | ✓ | ✓ |
| 8. Socio-economic status | | ✓ | ✓ | | | |
| 9. Smoking Status | | | ✓ | | ✓ | ✓ |
| 10. Drinking | | | | | ✓ | ✓ |
| 11. Diabetes | | ✓ | | | ✓ | |
| 12. Hearing problem | | ✓ | | | | |
| 13. Visual impairment | | ✓ | | | | |
| 14. Mobility impairment | | ✓ | | | | |
| 15. Insomnia | | ✓ | | | | |
| 16. Stroke | | | | | ✓ | |

### 4.2. RQ2: which depression screening tools were adopted?

Our review discovered 5 screening tools popularly used by past studies in depression prediction: geriatric depression scale (GDS), hospital anxiety and depression scale (HADS), patient health questionnaire (PHQ), hamilton depression rating scale (HDRS) and depression anxiety stress scale 21 (DASS-21). Refer to Table 4. We discovered that proposals predicting depression utilize screening tools when their methodology require the self-construction of a dataset. The motivation driving this construction is mainly because of the absence of an available dataset necessary to accomplish a research's unique objective of filling up a specific gap in the knowledge. For example, the use of GDS is targeted at screening depression in elders. These tools allow patients to assess themselves and ratings are based on this assessment. These self-assessment tools are not meant to replace a psychiatrist's diagnosis but instead function as a signpost to the presence of symptoms or to reinforce an earlier diagnosis that a psychiatrist may be considering. Our result shows that both HADS and HDRS were adopted by more research as compared to PHQ and DASS-21 in relation to the general population. GDS, however, was adopted when the older population is the subject of interest.

Table 4. Screening tools adopted

| Paper ID | Screening tools |
|---|---|
| P1 | GDS |
| P2 | GDS |
| P3 | HADS |
| P4 | None |
| P5 | None |
| P6 | None |
| P7 | HADS |
| P8 | None |
| P9 | PHQ |
| P10 | None |
| P11 | None |
| P12 | DASS-21 |
| P13 | HDRS |
| P14 | None |
| P15 | HDRS |

#### 4.2.1. Geriatric depression scale (GDS)

GDS [32], [33] consists of 30 questions targeted at the older population of 65 years and more who are medically ill. Although other depression screening tools are available, GDS has become the popular tool for this category of people. GDS simply requires a yes or no answer of how an elder feels in the past week. Because of its high sensitivity of 92% and specificity of 89%, GDS is viewed to be a valid and reliable tool. Table 5 shows the severity ratings produced by GDS.

Table 5. GDS severity ratings

| Severity | Depression |
|---|---|
| Normal | 0-4 |
| Mild | 5-8 |
| Moderate | 9-11 |
| Severe | 12-15 |

#### 4.2.2. Hospital anxiety and depression scale (HADS)

HADS [34], [35] measures the severity of not only depression but also anxiety. Since its introduction in 1983, HADS has become a popular screening tool for these two mental conditions. Comprising of 7 questions for anxiety and 7 questions for depression, HADS can be easily completed within a few minutes. The validity of HADS has been proven and is now on the recommendation list of the National Institute for Health and Care Excellence (NICE) to diagnose depression and anxiety. Table 6 displays HADS' severity ratings.

Table 6. HADS severity ratings

| Severity | Depression |
|---|---|
| Mild | 8-10 |
| Moderate | 11-14 |
| Severe | 15-21 |

#### 4.2.3. PHQ

The PHQ [36], [37] is a multipurpose method for screening, tracking, diagnosing, and measuring depression severity. It is a self-administered instrument with two distinct types, the PHQ-2 containing two items and the PHQ-9 containing nine items. PHQ-2 assesses the frequency of depressive episodes and anhedonia for the last two weeks, while PHQ-9 presents a clinical diagnosis of depression and measures the severity of symptoms. Table 7 shows the PHQ severity rating.

Table 7. PHQ severity ratings

| Severity | Depression |
|---|---|
| Mild | 0-5 |
| Moderate | 6-10 |
| Moderately severe | 11-15 |
| Severe | 16-20 |

#### 4.2.4. DASS-21

DASS-21 is a compilation of three scales of self-report by a patient that determines the patient's depression, anxiety, and emotional stress states. The underlying notion is these states tend to be correlated where anxiety and depression were discovered to be comorbid illnesses [38] and depression is a stress-related mental disorder [39]. Each state is measured by answering 7 questions relating to how a patient feels over the past week. DASS was designed to calculate the level of negative emotions to assist both researchers and clinicians to observe a patient's condition over time with the aim of determining the course of treatment. Table 8 shows the DASS-21 severity ratings.

Table 8. DASS-21 severity ratings

| Severity | Depression | Anxiety | Stress |
|---|---|---|---|
| Normal | 0-9 | 0-7 | 0-14 |
| Mild | 10-13 | 8-9 | 15-18 |
| Moderate | 14-20 | 10-14 | 19-25 |
| Severe | 21-27 | 15-19 | 26-33 |
| Extremely Severe | 28+ | 20+ | 34+ |

#### 4.2.5. HDRS

HDRS [40], [41] is specialized in assessing the severity of depression and has also been proven useful before, during, and after therapy to assess a patient's level of depression. It is widely perceived as an effective treatment for hospitalized patients. 21 items are listed in the HDRS form. The scoring basis is on the first 17 items, with 18 to 21 items used to qualify depression further. Table 9 shows the HDRS severity rating.

Table 9. HDRS severity ratings

| Severity | Depression |
|---|---|
| Normal | 0-7 |
| Mild | 8-13 |
| Moderate | 14-18 |
| Severe | 19-22 |
| Very severe | 23+ |

### 4.3. RQ3: What machine learning techniques were proposed by existing research?

Table 10 shows a list of the proposed machine learning techniques that were used in past research. For papers that compare the performance of the techniques, the highest scored technique is also listed in the table. Figure 3 summarizes in a tree map the number of papers using the proposed technique. Most papers experimented on random forest (RF), support vector machine (SVM), random tree (RT), naïve Bayes (NB), logistic regression (LR) and decision tree (DT). While this indicates the popularity of a specific machine learning technique among researchers, it is more importantly to know which of these techniques consistently scores the best performance when applied over different datasets. Out of the 15 papers reviewed, 12 papers conducted a comparison of performance. Therefore, from Figure 4, the graph shows RF returning the best performance in 4 instances of the comparison. RF prevails across different performance metrics in terms of achieving the best performance against other machine learning techniques. This is not only true for classical performance metrics e.g., accuracy, precision, and recall, but also newer forms of performance metrics such as early risk detection error (ERDE). It is noteworthy of publications proposing newer machine learning techniques i.e., Sons & Spouses algorithm (SS) superseding RF on traditional measurements of performances specifically accuracy, f-measure, precision, recall and area under the receiver curve. A particularly new performance metric is ERDE formulated specifically for detecting mental illness early.

Nomenclature:

- ADA: AdaBoost
- BA: Bagging
- BN: BayesNet
- CNN: convolutional neural networks
- DT: decision Tree
- GB: gradient Boosting
- KNN: K-nearest neighbor
- LR: logistic regression
- MDL: multimodal depressive dictionary learning

- MLP: multi-layer perceptron
- MSNL: multiple social networking learning
- NB: naïve Bayes
- NN: neural network
- RF: random forest
- RT: random tree
- RSS: random subspace
- SMO: sequential minimal optimization
- SS: Sons & Spouses
- SVM: support vector machine
- WDL: wasserstein dictionary learning

Table 10. Proposed machine learning techniques

| Paper ID | Machine learning techniques used | Best technique | Performance metrics | Best performance |
|---|---|---|---|---|
| P1 | RF, RT, MLP, and SVM | RF | Accuracy | 95.45 |
| | | | Mean absolute error | 0.12 |
| | | | Root mean squared error | 0.22 |
| | | | Relative absolute error | 24.30 |
| | | | Root relative squared error | 44.79 |
| P2 | BN, LR, MLP, SMO, and decision table | BN | Accuracy | 91.67 |
| | | | Precision | 0.92 |
| | | | ROC area | 0.98 |
| | | | Root mean squared error | 0.25 |
| P3 | BN, LR, MLP, NB, RF, RT, DT, random optimization, sequential, random sub-space, and K star | RF | Accuracy | 89 |
| | | | True positive rate | 89 |
| | | | False positive rate | 10.9 |
| | | | Precision/positive prediction value | 89.1 |
| | | | F-measures | 89 |
| | | | Area under the receiver curve | 94.3 |
| P4 | Stacking of LR DT, NBN, NN, SVM | LR (base-level learner) with DT, NBN, NN, SVM (meta-level learner) | Mean area under the receiver curve | 75 |
| | | | Mean accuracy | 86 |
| P5 | NB, MSNL, WDL and MDL | MDL | Precision | 84 |
| | | | Recall | 85 |
| | | | F1-measure | 84 |
| | | | Accuracy | 84 |
| P6 | CNN with TF-IDF information | Not compared | $ERDE_5$ | 10.81 |
| | | | $ERDE_{50}$ | 9.22 |
| | | | F-score | 37 |
| P7 | CatBoost, LR, NB, RF, and SVM | CatBoost | Accuracy | 89 |
| | | | Precision | 84 |
| P8 | DT, RT, and RF | RF | $ERDE_5$ | 18.51 |
| | | | $ERDE_{50}$ | 15.20 |
| | | | F-measure | 20 |
| | | | Precision | 12 |
| | | | Recall | 0 |
| P9 | BN, SVM, SMO, RT, and DT | BN | Accuracy | 77.8 |
| P10 | NB, RF, GB, and Ensemble Vote Classifier | Ensemble Vote Classifier | Accuracy | 85 |
| | | | F-score | 76.9 |
| P11 | CNN | Not compared | $ERDE_{20}$ | 9.46 |
| | | | $ERDE_{20}$ | 7.47 |
| | | | $F_{latency}$ | 0.45 |
| P12 | NB, RF, DT, SVM, and KNN | RF | Accuracy | 79.8 |
| | | | Error rate | 0.20 |
| | | | Precision | 88.1 |
| | | | Recall | 67.8 |
| | | | Specificity | 91.0 |
| | | | F1 score | 76.6 |
| P13 | SVM, RT, and RF | RT | Accuracy | 91.3 |
| | | | Recall | 91.2 |
| | | | Precision | 91.3 |
| P14 | SS, TAN, LR, DT, NN, SVM, ADA, BA, RF, RSS | SS | F measure | 91.8 |
| | | | Accuracy | 93.0 |
| | | | Area under the receiver curve | 76.9 |
| | | | Precision | 93.1 |
| | | | Recall | 90.6 |

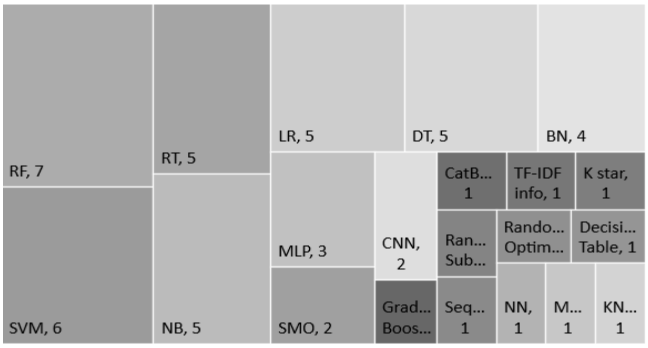


Figure 3. The number of papers using the proposed technique

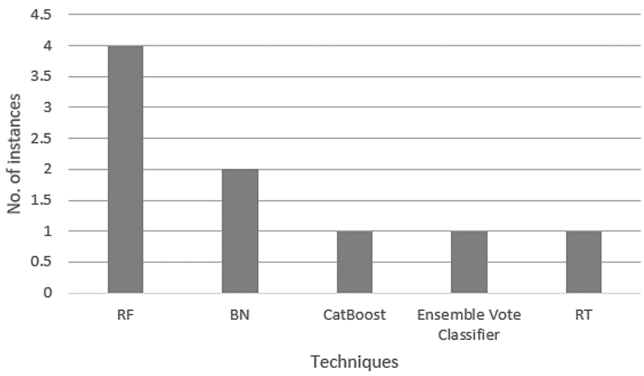


Figure 4. Techniques with consistently high performance

## 5. CONCLUSION

In this timely paper, we have reviewed depression prediction literature from 2016 to 2020 that used machine learning techniques. We employed the SMS method, and the result is a total of 15 works were found relevant to the research questions constructed. The research questions focus on three important aspects of predicting depression using machine learning; they are the variables used in the literature to predict, the screening tools adopted, the machine learning techniques experimented, the metrics employed to measure each techniques' performance and the highest values achieved by the top-performing techniques. Our review has led us to conclude that information on age, marital status, gender, educational status, and socio-economic status are repeatedly used across the proposals. In addition, most of the works which made use of depression screening tools relied on self-reporting types. Furthermore, Random Forest was not only the most popular machine learning algorithm among researchers but also returns the best performance in a majority of the time inclusive of newer performance metrics e.g., ERDE. It is expected that this survey will enlighten researchers on the latest machine learning techniques, performance measurements and variables used in predicting depression.

## ACKNOWLEDGEMENTS

We are extremely grateful to the reviewers who took the time to provide constructive feedback and useful suggestions for the improvement of this article. The Malaysian Government is funding this research through the Fundamental Research Grant Scheme (FRGS) at Universiti Teknologi MARA (UiTM) Shah Alam, Malaysia (FRGS/1/2019/SS05/UITM/02/5).

## REFERENCES

[1] H. Dai, S. X. Zhang, K. H. Looi, R. Su, and J. Li, "Perception of health conditions and test availability as predictors of adults' mental health during the COVID-19 pandemic: a survey study of adults in Malaysia," *International Journal of Environmental Research and Public Health*, vol. 17, no. 15, Jul. 2020, doi: 10.3390/ijerph17155498.

[2] N. Sahril, N. A. Ahmad, I. B. Idris, R. Sooryanarayana, and M. A. Abd Razak, "Factors associated with mental health problems among Malaysian children: a large population-based study," *Children*, vol. 8, no. 2, Feb. 2021, doi: 10.3390/children8020119.

[3] J. Bilsen, “Suicide and youth: risk factors,” *Frontiers in Psychiatry*, vol. 9, pp. 1–5, Oct. 2018, doi: 10.3389/fpsyt.2018.00540.

[4] L. Brådvik, “Suicide risk and mental disorders,” *International Journal of Environmental Research and Public Health*, vol. 15, no. 9, Sep. 2018, doi: 10.3390/ijerph15092028.

[5] A. Bakar *et al.*, “National health and morbidity survey 2015-volume ii : non-communicable diseases, risk factors & other health problems,” 2015.

[6] K. Katchapakirin, K. Wongpatikaseree, P. Yomaboot, and Y. Kaewpitakkun, “Facebook social media for depression detection in the Thai community,” in *2018 15th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Jul. 2018, pp. 1–6., doi: 10.1109/JCSSE.2018.8457362.

[7] J. F. Greden and R. Garcia-tosi, *Mental health in the workplace: strategies and tools to optimize outcomes*. Cham: Springer International Publishing, 2019., doi: 10.1007/978-3-030-04266-0.

[8] T. Kongsuk, S. Supanya, K. Kenbubpha, S. Phimtra, S. Sukhawaha, and J. Leejongpermpoon, “Services for depression and suicide in Thailand,” *WHO South-East Asia Journal of Public Health*, vol. 6, no. 1, 2017, doi: 10.4103/2224-3151.206162.

[9] Y. Yang, C. Fairbairn, and J. F. Cohn, “Detecting depression severity from vocal prosody,” *IEEE Transactions on Affective Computing*, vol. 4, no. 2, pp. 142–150, Apr. 2013, doi: 10.1109/T-AFFC.2012.38.

[10] K. M. Griffiths, B. Carron-Arthur, A. Parsons, and R. Reid, “Effectiveness of programs for reducing the stigma associated with mental disorders. a meta-analysis of randomized controlled trials,” *World Psychiatry*, vol. 13, no. 2, pp. 161–175, Jun. 2014, doi: 10.1002/wps.20129.

[11] S. H. Yeoh, C. L. Tam, C. P. Wong, and G. Bonn, “Examining depressive symptoms and their predictors in malaysia: stress, locus of control, and occupation,” *Frontiers in Psychology*, vol. 8, pp. 1–10, Aug. 2017, doi: 10.3389/fpsyg.2017.01411.

[12] G. Schomerus, H. Matschinger, and M. C. Angermeyer, “The stigma of psychiatric treatment and help-seeking intentions for depression,” *European Archives of Psychiatry and Clinical Neuroscience*, vol. 259, no. 5, pp. 298–306, Aug. 2009, doi: 10.1007/s00406-009-0870-y.

[13] J. E. Aurelia, Z. Rustam, I. Wirasati, S. Hartini, and G. S. Saragih, “Hepatitis classification using support vector machines and random forest,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 2, pp. 446–451, Jun. 2021, doi: 10.11591/ijai.v10.i2.pp446-451.

[14] N. A. Mashudi, N. Ahmad, and N. M. Noor, “Classification of adult autistic spectrum disorder using machine learning approach,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 3, pp. 743–751, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp743-751.

[15] C. K. Chin, D. A. binti Awang Mat, and A. Y. Saleh, “Hybrid of convolutional neural network algorithm and autoregressive integrated moving average model for skin cancer classification among Malaysian,” *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 3, pp. 707–716, Sep. 2021, doi: 10.11591/ijai.v10.i3.pp707-716.

[16] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, “Systematic mapping studies in software engineering,” in *International Journal of Software Engineering and Knowledge Engineering*, Jun. 2008, vol. 17, no. 1, pp. 33–55., doi: 10.14236/ewic/EASE2008.8.

[17] I.-M. Spyrou, C. Frantzidis, C. Bratsas, I. Antoniou, and P. D. Bamidis, “Geriatric depression symptoms coexisting with cognitive decline: a comparison of classification methodologies,” *Biomedical Signal Processing and Control*, vol. 25, pp. 118–129, Mar. 2016, doi: 10.1016/j.bspc.2015.10.006.

[18] I. Bhakta and S. Arkaprabha, “Prediction of depression among senior citizens using machine learning classifiers,” *International Journal of Computer Applications*, vol. 144, no. 7, pp. 11–16, 2016

[19] 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, Dec. 2017, doi: 10.1049/htl.2016.0096.

[20] E. S. Lee, “Exploring the performance of stacking classifier to predict depression among the elderly,” in *2017 IEEE International Conference on Healthcare Informatics (ICHI)*, Aug. 2017, pp. 13–20., doi: 10.1109/ICHI.2017.95.

[21] G. Shen *et al.*, “Depression detection via harvesting social media: a multimodal dictionary learning solution,” in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Aug. 2017, pp. 3838–3844., doi: 10.24963/ijcai.2017/536.

[22] Y. T. Wang, H. H. Huang, and H. H. Chen, “A neural network approach to early risk detection of depression and anorexia on social media text,” in *CEUR Workshop Proceedings*, 2018, vol. 2125.

[23] A. Sau and I. Bhakta, “Screening of anxiety and depression among seafarers using machine learning technology,” *Informatics in Medicine Unlocked*, vol. 16, 2019, doi: 10.1016/j.imu.2019.100228.

[24] 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, Jun. 2019, doi: 10.2196/12554.

[25] Z. Yang, H. Li, L. Li, K. Zhang, C. Xiong, and Y. Liu, “Speech-based automatic recognition technology for major depression disorder,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11956, 2019, pp. 546–553., doi: 10.1007/978-3-030-37429-7_55.

[26] A. Kumar, A. Sharma, and A. Arora, “Anxious depression prediction in real-time social data,” *SSRN Electronic Journal*, pp. 1–7, 2019, doi: 10.2139/ssrn.3383359.

[27] M. Trotzek, S. Koitka, and C. M. Friedrich, “Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 3, pp. 588–601, Mar. 2020, doi: 10.1109/TKDE.2018.2885515.

[28] 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.

[29] B. Abdallah, “Computer based technique to detect depression in Alzheimer patients,” Université du Québec, 2020.

[30] F. J. Costello, C. Kim, C. M. Kang, and K. C. Lee, “Identifying high-risk factors of depression in middle-aged persons with a novel sons and spouses bayesian network model,” *Healthcare*, vol. 8, no. 4, Dec. 2020, doi: 10.3390/healthcare8040562.

[31] N. Xiong *et al.*, “Demographic and psychosocial variables could predict the occurrence of major depressive disorder, but not the severity of depression in patients with first-episode major depressive disorder in China,” *Journal of Affective Disorders*, vol. 274, pp. 103–111, Sep. 2020, doi: 10.1016/j.jad.2020.05.065.

[32] S. A. Greenberg, “The geriatric depression scale (GDS) validation of a geriatric depression screening scale: a preliminary report,” *Best Practices in Nursing Care to Older Adults*, no. 4, 2019

[33] J. Wancata, R. Alexandrowicz, B. Marquart, M. Weiss, and F. Friedrich, “The criterion validity of the geriatric depression scale: a systematic review,” *Acta Psychiatrica Scandinavica*, vol. 114, no. 6, pp. 398–410, Dec. 2006, doi: 10.1111/j.1600-0447.2006.00888.x.

[34] A. S. Zigmond and R. P. Snaith, “The hospital anxiety and depression scale,” *Acta Psychiatrica Scandinavica*, vol. 67, no. 6, pp. 361–370, Jun. 1983, doi: 10.1111/j.1600-0447.1983.tb09716.x.

[35] I. Bjelland, A. A. Dahl, T. T. Haug, and D. Neckelmann, "The validity of the hospital anxiety and depression scale," *Journal of Psychosomatic Research*, vol. 52, no. 2, pp. 69–77, Feb. 2002, doi: 10.1016/S0022-3999(01)00296-3.
[36] J. Ford, F. Thomas, R. Byng, and R. McCabe, "Use of the patient health questionnaire (PHQ-9) in practice: interactions between patients and physicians," *Qualitative Health Research*, vol. 30, no. 13, pp. 2146–2159, Nov. 2020, doi: 10.1177/1049732320924625.
[37] L. P. Richardson *et al.*, "Evaluation of the patient health questionnaire-9 item for detecting major depression among adolescents," *Pediatrics*, vol. 126, no. 6, pp. 1117–1123, Dec. 2010, doi: 10.1542/peds.2010-0852.
[38] N. H. Kalin, "The critical relationship between anxiety and depression," *American Journal of Psychiatry*, vol. 177, no. 5, pp. 365–367, May 2020, doi: 10.1176/appi.ajp.2020.20030305.
[39] P. Koutsimani, A. Montgomery, and K. Georganta, "The relationship between burnout, depression, and anxiety: a systematic review and meta-analysis," *Frontiers in Psychology*, vol. 10, Mar. 2019, doi: 10.3389/fpsyg.2019.00284.
[40] S. Obeid, C. Abi Elias Hallit, C. Haddad, Z. Hany, and S. Hallit, "Validation of the hamilton depression rating scale (HDRS) and sociodemographic factors associated with Lebanese depressed patients," *L'Encéphale*, vol. 44, no. 5, pp. 397–402, Nov. 2018, doi: 10.1016/j.encep.2017.10.010.
[41] M. Zimmerman, J. H. Martinez, D. Young, I. Chelminski, and K. Dalrymple, "Severity classification on the Hamilton depression rating scale," *Journal of Affective Disorders*, vol. 150, no. 2, pp. 384–388, Sep. 2013, doi: 10.1016/j.jad.2013.04.028.


## BIOGRAPHIES OF AUTHORS

**Hanis Diyana Abdul Rahimapandi** completed her M.Sc. in Data Science from the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, in 2021. In 2018, she received a B.Sc. in Management Mathematics from the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Arau. She is currently employed with TESS Innovation Sdn Bhd in Selangor, Malaysia, as a Data Analyst. Machine learning and data mining are two of the topics of her research. She can be contacted at email: hd.hanisdiyana@gmail.com.

**Dr. Ruhaila Maskat** is a senior lecturer at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA Shah Alam, Malaysia. In 2016, she was awarded a Ph.D. in Computer Science from the University of Manchester, United Kingdom. Her research interest then was in Pay-As-You-Go dataspaces which later evolved to Data Science where she is now an EMC Dell Data Associate as well as holding four other professional certifications from RapidMiner in the areas of machine learning and data engineering. Recently, she was awarded with the Kaggle BIPOC grant. Her current research grant with the Malaysian government involves conducting analytics on social media text to detect mental illness. She can be contacted at: ruhaila@fskm.uitm.edu.my.

**Prof. Ramli Musa** is a consultant psychiatrist at Department of Psychiatry, International Islamic University Malaysia (IIUM). He received a few accolades including Outstanding Research Quality Award (Health and Allied Sciences) for 2 consecutive years (2012 and 2013), University Best Researcher Award; Best Fundamental Research Grant Scheme (FRGS) (Social Science) 2015, winner in the 14th International Conference and Exposition on Inventions by Institutions of Higher Learning (PECIPTA 2015), Commercial Potential Award; winner for The Best Project FRGS 2/2011 by Ministry of Higher Education, Malaysia Technology Expo 2018. Being passionate in research, he established a database of all the translated and validated questionnaires in Malaysian language and a website called the Mental Health Information and Research (MaHIR). He can be contacted at: drramli@iium.edu.my.

**Dr. Norizah Ardi** is an Associate Professor at Academy of Language Studies, Universiti Teknologi MARA Shah Alam since July 2017. She has Ph.D. in Malay Language Studies and an expertise in Malay Linguistics (Sociolinguistics, Applied Linguistics, Translation Studies). She was teaching Malay Language course at Universiti Teknologi MARA since 1994. Currently, she involved with several research project regarding language and computational linguistics. She is member of Malaysian Translation Association and Member of Malaysian Linguistic Association Malaysia. She has published over 40 papers in national and international journals and conferences. She can be contacted at email: norizah@uitm.edu.my.