

Using Machine Learning to Diagnose Depression: A Review of Research Trends and Algorithms

Mr. Geoffrey Mwamba Nyabuto*, Dr. Alice Nambiro

Kibabii University, Information Technology, P.O. Box 1699—50200, Bungoma County

*Corresponding Author

DOI: <https://doi.org/10.51244/IJRSI.2024.1102008>

Received: 11 January 2024; Revised 01 February 2024; Accepted: 06 February 2024; Published: 28 February 2024

ABSTRACT

Depression is one of the most common mental health illnesses that affects millions of people worldwide. In the recent past, there have been numerous studies on depression and mental health at large. Artificial Intelligence (AI) technologies have increasingly been adopted to solve complex human-related problems including the diagnosis and treatment of different diseases. Researchers have been trying to establish if and how Machine Learning (ML), a branch of AI, can be used to diagnose depression to help solve this complex problem of mental health.

In this paper, we seek to review and establish unique research trends in using ML algorithms to diagnose depression as well as establish the most recommended ML algorithm that has shown high accuracy and hence mostly been recommended by many research articles.

The study consisted of 2 phases where in the first phase, 3 major publishers and journals were considered to establish research trends. The phase considered articles published between 2015 to 2023. In the second phase, a total of 20 journal articles with open access and having been published between 2020 and 2023 were considered to establish the most recommended ML algorithm for solving the problem of diagnosing depression among people.

Results from phase 1 of the analysis indicated a sharp increase in research trends in this area. In phase 2, different research articles showed varied accuracy with Deep Learning (DL) showing a high level of accuracy. 4 out of the 7 studies that recommended the use of DL as the preferred algorithm showed an accuracy level of over 96%. The same trends were observed in related statistical measures i.e., F1-Score and Area Under the ROC Curve (AUC) where DL outperformed the other ML algorithms. In some studies, it achieved a high score of 0.995 in AUC and 0.98 in F1-Score. DL was mostly used when a large amount of data was considered i.e., survey and social media data. However, other algorithms like Support Vector Machine (SVM), XGBoost, and Logistic Regression also performed well.

Generally, ML and DL proved to be an efficient tool that can be used to diagnose depression by using data from different sources i.e., social media, Electronic Medical Records (EMR), surveys, interviews, and wearables among others. Though DL proved to be an efficient tool to predict depression, in other studies, other ML algorithms had a relatively good performance and were recommended. This raises the need to compare all the algorithms on a case-by-case basis before deciding which one to use.

Keywords: Depression, Machine learning, deep learning, random forest, mental health, neural networks,

supervised ML, suicide, artificial intelligence, diagnose depression

INTRODUCTION

Depression is a complex mental health condition that affects millions of people worldwide. The World Health Organization (WHO) report released in 2022 estimates the number of people suffering from one or another form of mental health to be nearly 1 billion (WHO, 2022). Depression is one of the most common forms of mental health, The WHO released another report that showed that approximately 3.8% of the general population is suffering from depression (WHO, 2023). Suicide, which is commonly associated with depression accounts for more than 700,000 deaths globally with youths aged between 15 and 24 years being the most affected and approximately 77% of them being from low and middle-income countries (WHO, 2023).

Technology has been employed to solve challenges that were very difficult to solve while others were a reserve of human beings (Joksimovic et al., 2023). Artificial Intelligence (AI) developments have shown great hope in solving such challenges. In the healthcare industry, technology has been employed to manage patient information and offer decision support. Using ML, it is now possible to predict patients who are likely to miss an appointment (Joseph et al., 2022). Having shown positive outcomes, it is possible to try incorporating the same to solve other related problems. ML learns from a large pool of datasets using various algorithms and makes self-adjustments to give optimal output. In healthcare, ML has mainly been used to diagnose different diseases and has shown great positive outcomes (Poudel, 2022).

Recent attempts to solve this complex problem of depression show attempts to incorporate and use technology to help diagnose these cases. ML models have been developed and shown positive outcomes in predicting patients likely to be suffering from mental health (Alharahsheh & Abdullah, 2021). Some studies have evaluated different algorithms to establish the performance of different algorithms in solving this problem. For example, a study by Maranthe et al. evaluated selected supervised machine learning algorithms i.e., RF, SVM, Ada boosting and Voting-Ensemble models and found that RF had the highest accuracy level of 85%. Supervised ML has been used to diagnose depression among HIV HIV-infected populations in Canada showing that 53% of the sampled population were diagnosed with depression. In contrast, only 10% had the same documented in their medical records (Marathe et al., 2022).

This study aims to establish research trends in this area, establish the performance of different algorithms in solving this challenge and propose the most recommended algorithm.

Machine Learning

ML has been widely adopted to solve complex problems in several domains. In healthcare, ML has been applied in several areas including patient engagement and adherence applications, diagnosis, and treatment applications, administrative applications, robotics, natural language processing, and expert systems (Davenport & Kalakota, 2019). One of the common challenges has been different ML algorithms giving different outputs based on the same dataset hence the need for a holistic review to determine the best algorithm to give an optimal output for a given disease (Poudel, 2022). Three sets of algorithms are available for use in ML, they include supervised, semi-supervised and unsupervised ML algorithms. ML has 2 steps, the first one being the learning or training phase where the algorithm learns the hidden trends in datasets, and the second one being the testing phase to validate if the algorithms can correctly predict some outcomes.

Supervised ML

Supervised ML algorithms are used to predict outcomes i.e., to classify specific outcomes. It uses labelled

data to train algorithms to correctly classify data or an outcome. This class of ML algorithm has been applied to solve different real-world problems including classifying spam mail, and disease diagnosis among others. Supervised learning uses a training dataset to teach the model to achieve some desired output. The dataset has both inputs and desired output. The algorithms measure their performance and re-adjust to achieve the best output.

With Supervised algorithms, the data is split into two sets, training, and test data. The training data set has two key components i.e., input variables and labelled output. Some examples of supervised ML algorithms include RF, Baiyes Naïve Models, Decision Tree (DT), linear and logistic regression, and SVM (Kotsiantis, 2007).

In disease diagnosis, supervised ML has extensively been used to diagnose different diseases including rare diseases that are complex for humans to diagnose (Schaefer et al., 2020). Different algorithms in this group have been applied and they have yielded different results in the prediction of depression.

Semi-Supervised ML

This technique involves using both labelled and un-labelled datasets for training purposes. In many cases, it uses large volumes of un-labelled data as well as smaller sets of labelled data (Fatima & Pasha, 2017). In depression diagnosis, both social media, smartphone, and PHQ-9 clinical can be considered. Given that social media and smartphone usage data is largely unlabeled whereas the PHQ-9 data is classified, semi-supervised ML algorithms can be used to diagnose depression in such cases (Aleem et al., 2022).

Unsupervised ML

In unsupervised ML algorithms, objects are classified using characteristics where those that exhibit similar behaviour are classified together. Some examples of unsupervised ML algorithms include the frequent pattern growth algorithm, the ECLAT algorithm, the Apriori algorithm, clustering using k-means, and principal components analysis (Naeem et al., 2023).

In this category, the algorithm can self-learn and develop trends with unlabeled data. The algorithm uses two approaches, clustering, and association to group similar items together. In clustering, the data is broken into groups based on its similarities or differences. On the other hand, the association is a rule-based approach to mine and reveals interesting relationships in data. Though there is a prevalence of using supervised ML to solve the problem of diagnosing depression, unsupervised ML has also been used to solve this challenge (Kung et al., 2022).

Diagnosing Depression Using ML

ML has extensively been used to not only diagnose common diseases but has also shown the ability to diagnose rare diseases (Schaefer et al., 2020). DL and ML algorithms have been used to diagnose depression. Supervised ML algorithms have been given prevalence. ML is an efficient tool that can be used to identify and predict depression and anxiety-associated factors among school-going children (Qasrawi et al., 2021). Different studies have found that different algorithms have been able to yield different results. There is no single algorithm that best suits the problem of depression than other algorithms hence the testing of all the algorithms before choosing the best for diagnosing depression.

RF, one of the ML algorithms has shown good performance in terms of prediction with an accuracy of as high as 99% (Moon et al., 2021). Similarly, other supervised ML algorithms have shown equal success in diagnosing depression. Sayani Ghosala and Amita Jainb did a study on the same using social media data and were able to identify XGBoost as the best algorithm to identify depressive cases with an accuracy of

71% (Ghosala & Jainb, 2023). Another research done by Shaunak et al. using social media data from Reddit identified SVM as the best-performing algorithm in diagnosing depression as compared to logistic regression and XGBoost (Inamdar et al., 2023).

DL has been used to diagnose depression and it has shown greater success and accuracy as compared to ML. This could be due to the number of data points it uses to predict. With social media generating millions of records daily, DL comes in as a very good algorithm that can be used to diagnose depression from this data. A comparison of different DL and ML algorithms i.e., Random Forest, Logistic regression, B Naïve Bayes, and Bert algorithms shows that DL algorithms (RoBERTa) perform better at 98% accuracy as compared to other algorithms (Bokolo & Liu, 2023). Similar findings were upheld by another research that used the USA Veteran dataset to predict depressive cases (Qu et al., 2023).

Challenges of Using ML to Diagnose Depression

While studies have shown great potential of using ML to solve depression related challenges, there exist several challenges that should also be addressed to make its incorporation easier and better. One of the biggest problems is the un-explainability and opaqueness of ML algorithms since the end users do not get to understand how and what considerations the algorithm uses to conclude that one is at risk of depression (Tornero-Costa et al., 2023). There is lack of a clear criterion that can be followed to arrive at a conclusion since many of the parameters being used to know if one is depressed or not are subjective and cannot easily be quantified (Yan et al., 2022). This makes it hard to have a standardized diagnostic approach when building the ML model.

Another challenge associated with using ML techniques in solving depression is the potential of bias since ML algorithms learn based on the data provided. Studies have consistently shown some levels of biases when applying these ML techniques to solving the problem of depression (Dang et al., 2023). If the data is biased, chances are high that the model developed will be biased.

METHODOLOGY

This research employed a systematic literature review also called desktop research where already published research in referred journals was searched, located, evaluated, and used (Brereton et al., 2007). Readily available resources on the Internet were given priority with those published after 2015 being considered. This research used several keywords to search for relevant journals that have already been published. Some of the keywords considered were “Using machine learning to diagnose depression”, “AI and depression” and “depression and machine learning”.

This research was done in 2 phases with the first phase trying to establish the general research trends in the application of ML to solve the problem of depression and the second phase analyzing specific selected articles to determine the different ML models/algorithms that have widely been used to solve this problem. In the first phase, the research considered 3 key sources to get relevant information i.e., Science Direct, Springer and IEEE. These sources were considered as one can easily extract meta-data information about relevant journals without necessarily reading through the whole article. In the second phase, a general Google search was done, and extra inclusion and exclusion criteria were employed as described in phase 2 below.

Phase 1

Inclusion

- i. Using keywords, only research articles related to the use of machine learning to detect/diagnose

depression were considered.

- ii. Journal articles published in the 3 pre-selected journals were considered for phase 1. They included Science Direct, Springer and IEEE.
- iii. Only journal articles written in English were considered.

Exclusion.

- i. All studies done before 2015 and after 2023 were not considered.

Phase 2

In phase 2, a sample of studies in diagnosing depression using machine learning and, in this case, a simple Google search using the keyword “diagnosing depression using machine learning” and journal articles that appeared first were considered. Below is a detailed explanation of the inclusion and exclusion criteria used.

Inclusion

- a. All journal articles relevant to solving the challenge of diagnosing depression using machine learning were considered.
- b. Only journal articles published between 2020 and 2023 were considered.
- c. Considered articles included at least an experimental and analytical aspect in using machine learning to diagnose depression.
- d. Only journal articles with full or open access were considered.
- e. Only journal articles written in English were considered.

Exclusion

- a. Review journals were not considered.
- b. Duplicated articles were dropped.
- c. Any articles that do not substantially discuss aspects of using machine learning to diagnose depression were also excluded.
- d. Articles that do not recommend at least one ML algorithm were also dropped.

The study used two phases with each of the phases having a different aim. In phase 1, the study focused mainly on bringing out research trends around the field of machine learning and its application to solve the problem of depression in terms of screening and identifying depression cases among the general or specified population. This is necessary because it points out how important that topic is based on the number of research articles published each year. Any other important information like the country where this research was done was also considered if it was easily available in the downloaded manifest file.

In the second phase, an in-depth analysis was carried out to determine the specific algorithms that are preferred in solving this problem as well as data sources used among others as discussed in the next section. This phase reviewed the most recent articles published between 2020 and 2023 with open access.

STUDY FINDINGS AND ANALYSIS

Phase 1

Using the keyword “Diagnosing depression using machine learning” findings in the three sources i.e., Science Direct, Springer and IEEE were downloaded and analyzed per year of publication and results were visualized as shown in the table below,

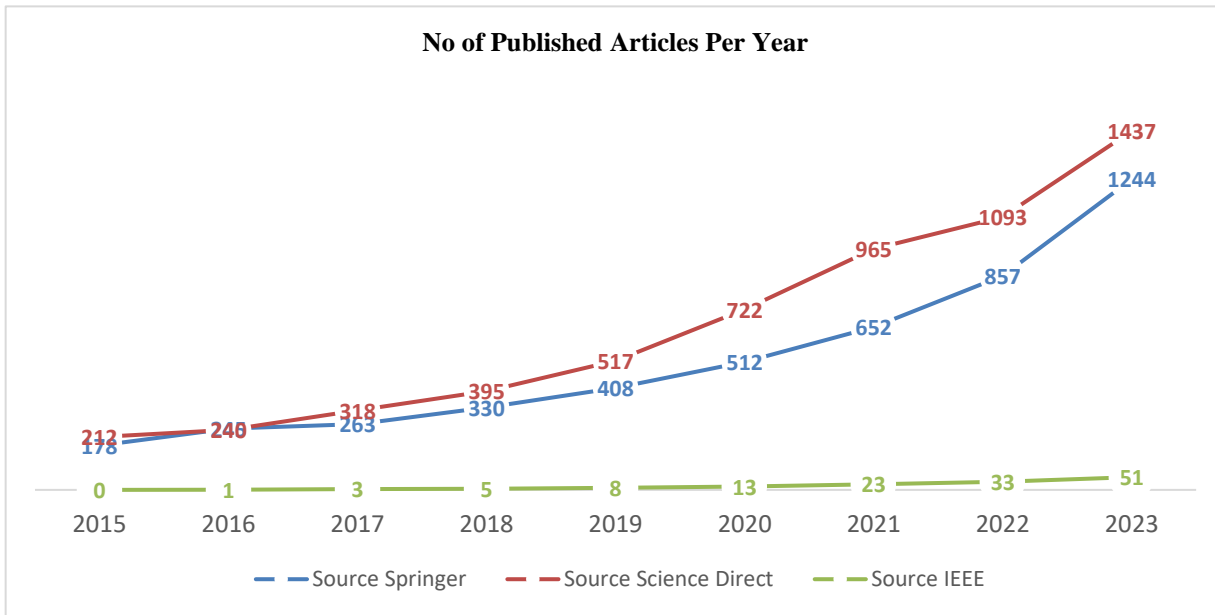


Figure 1: Research Trends in Diagnosing Depression using ML.

As per this analysis, it can be noted that the number of articles published online from 2015 to 2023 on this specific topic has steadily increased over time. For example, in Springer, in 2015, there were only 178 articles published. However, in 2023, the number of articles increased to 1,244, the same trends were noted in all the other sources considered. This clearly shows that there has been increased research interest in this area over time and the same trend is likely to be seen in 2024 and going forward. This has been pushed further by the fact that ML has successfully been applied in other areas as well as in medicine to diagnose and even treat different diseases.

Phase 2

Findings from the first phase clearly showed a growing interest in solving the problem of diagnosing depressive cases using ML and DL algorithms. Further to this, phase two of the research reviewed some of the recently published articles (from 2020 to 2023) in this area, trying to understand the specific approaches given, data sources and key findings. Out of 85 articles found, 20 were found to be relevant to meeting the inclusion and exclusion criteria set there before. Many of these articles were published in Science Direct, Springer, BMC, PLOS One and Frontiers among others as shown in the figure below.

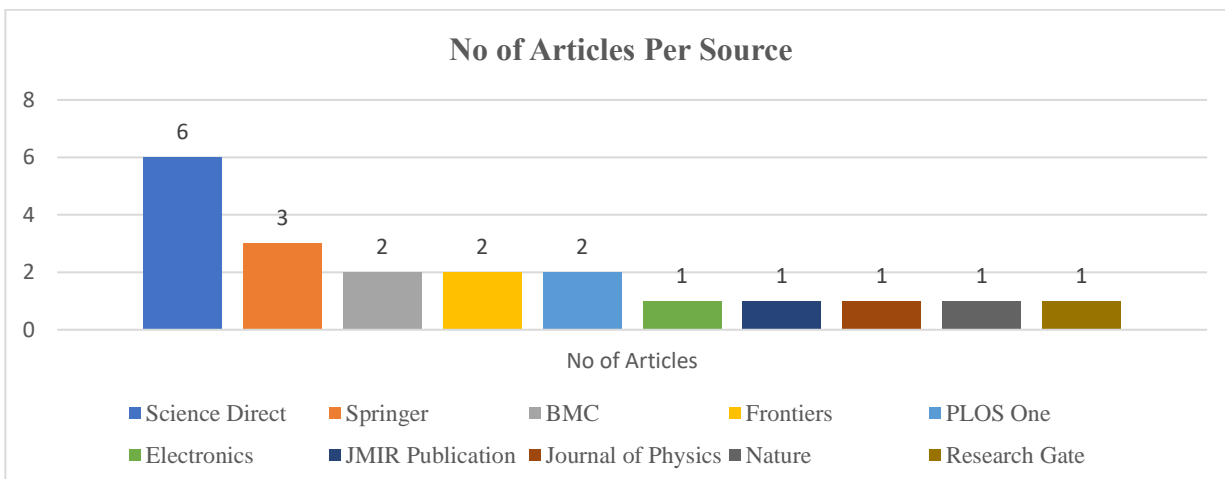


Figure 2: Number of Articles per Source/Journal

In terms of year of publication, many of the articles published were from 2023 as shown in the figure below. Over time, there seems to be a decrease in the number of studies that use interviews to get their data for training the different models. On the other hand, it is clearly shown that the number of articles using social media data to train and validate their models is increasing over time. This could have been triggered by the fact that there is a lot of data being shared by individuals on social media platforms that could potentially be used to know if someone is depressed or not. In terms of accuracy, precision, recall and f1-score for data from different sources, there was mixed performance ranging from 0.7 to 0.99 for the different scores. This ideally means the performance of any given algorithm cannot be attributed to a given source of its data. Many of the articles published were from India (4), the USA (3), Australia (2) and other countries (1).

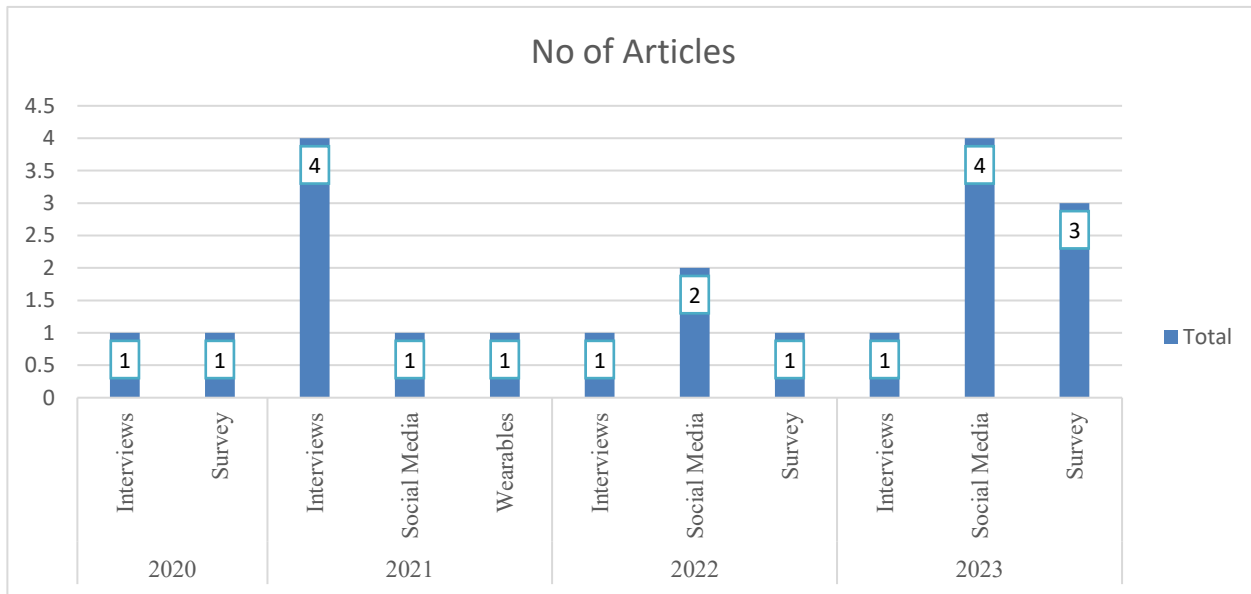


Figure 3: No of articles reviewed per data source and year.

Different DL and ML Algorithms were used by different studies analyzed with Majority showing RF and DL performing better than the other algorithms. Even for those studies where DL performed better, RF came in as the second-best algorithm in terms of accuracy. SVM came in as another algorithm that performed well and, in some cases, it appeared as the second best-performing algorithm.

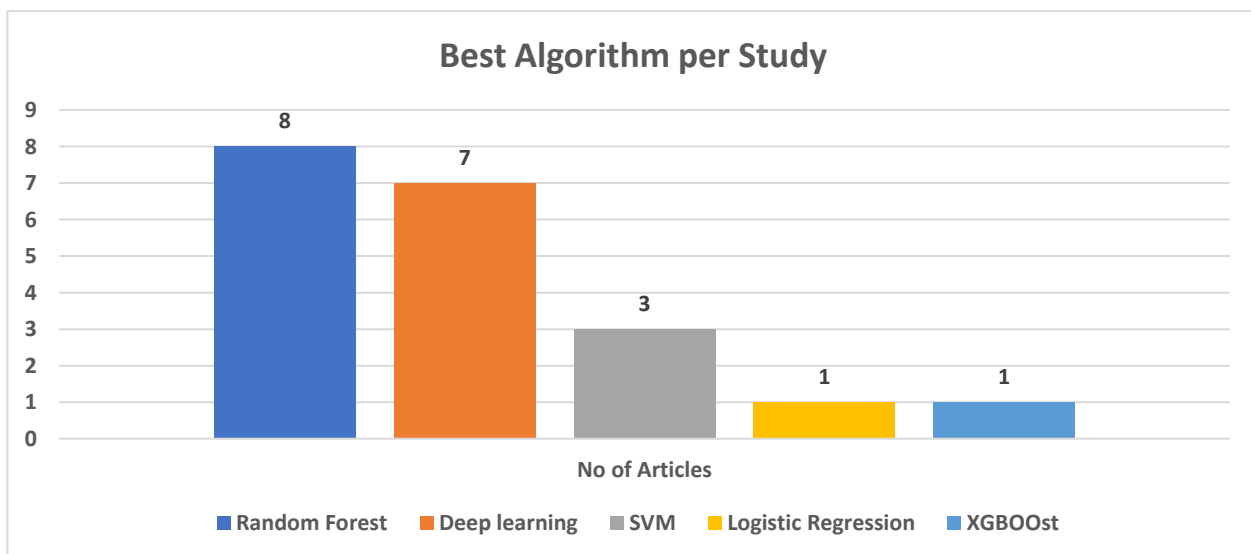


Figure 4: Best Recommended ML and DL Algorithm per Article

Of the 20 studies under review, 5 of them had the best prediction accuracy of between 70% and 79%, 6 studies had the best accuracy score of between 80% and 89% while the rest scored over 90%. Different versions of Neural Networks were used in implementing different DL approaches to diagnosing depression. Studies that employed both DL and ML showed that deep learning performed better than the other ML algorithms. In terms of accuracy, many of the studies reviewed proposed DL (especially Neural Networks) algorithms as 4 out of the 6 studies had an accuracy of over 96%. Random Forest had 1 study achieving an accuracy score of 99%. Studies that employed DL have large datasets from surveys and social media hence validating the choice to use DL as opposed to ML. DL had many studies achieving over 90% score in accuracy, f1-score, and AUC as shown in the table below.

Algorithm	Accuracy	F1-Score	Area Under ROC Curve
Deep Learning	4	3	2
Random Forest	2	0	0
SVM	1	0	0
XGBoost	0	0	0
Logistic Regression	1	0	0

Table 1: No of DL and ML Algorithms Achieving over 90% Score.

Of the 20 studies reviewed, 4 only used accuracy as a measure of the performance of the different algorithms under consideration. Five studies used both the F-measure (f1-score) and the area under the ROC curve (AUC) to measure precision, recall, sensitivity, and specificity. Nine studies used F1-score only whereas the remaining used only AUC. From the analysis, DL scored the highest in ROC with some studies having a ROC score of as high as 0.995 followed by 0.92 among others. SVM and RF had a relatively good ROC score of between 0.83 and 0.82 respectively. In terms of F1-score, DL scored well with the first 6 studies scoring the highest F1-score all being DL algorithms. Their F1-score ranged from a low of 0.78 to a high of 0.98 with 3 of the studies having an F1-score of over 0.96.

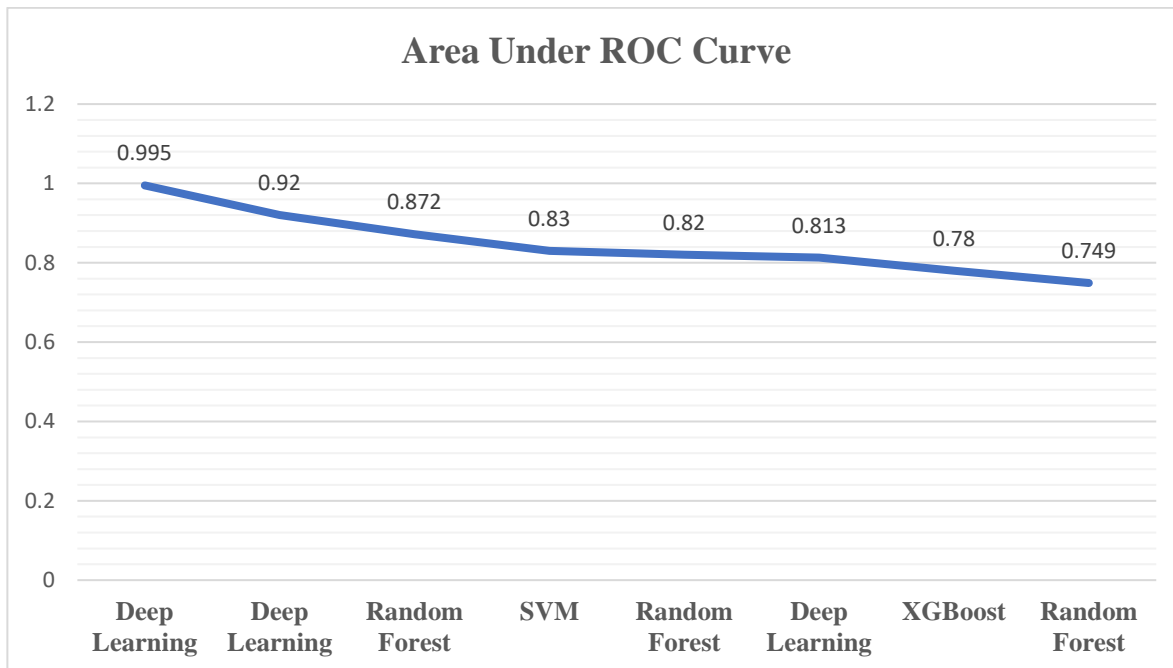


Figure 5: AUC Performance for DL and ML Algorithms

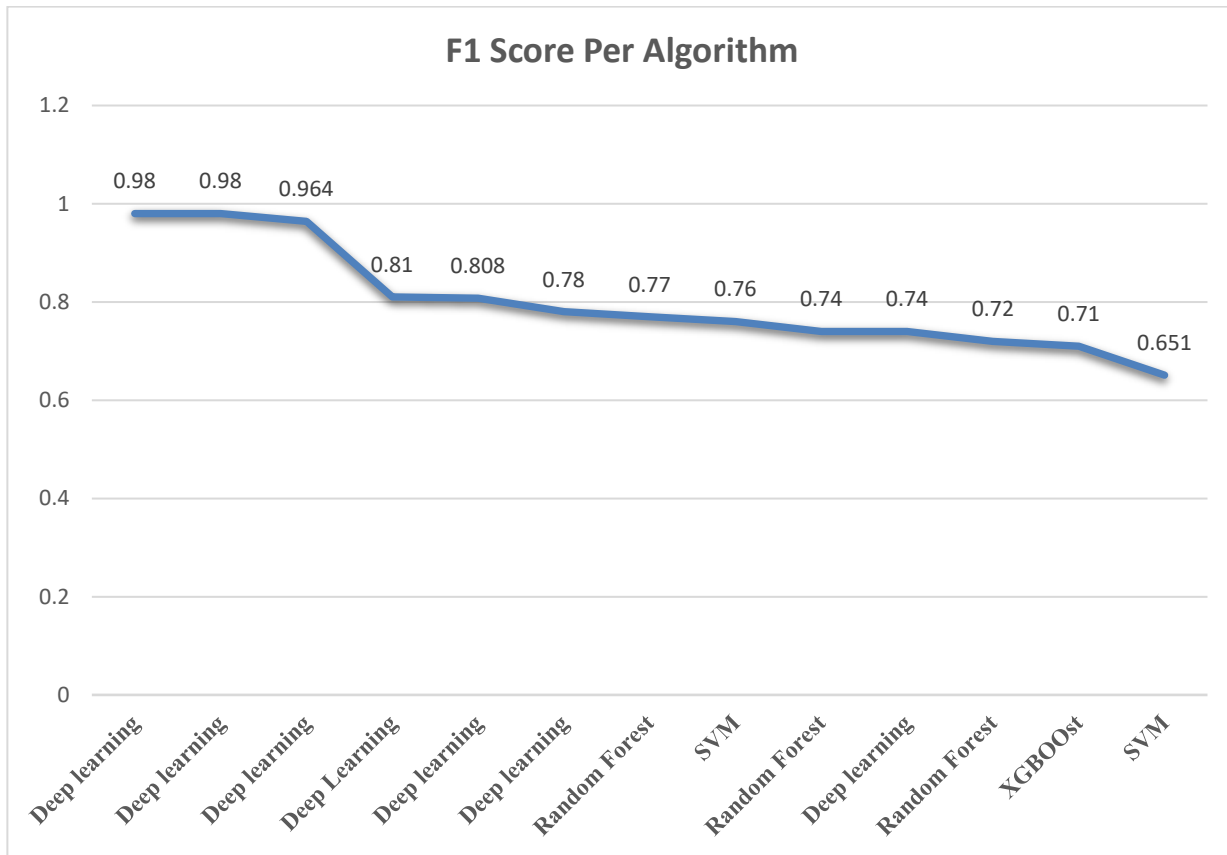


Figure 6: F1-Score for Different DL and ML algorithms

CONCLUSION

This systematic review supports the potential use of ML and DL approaches and different sources of datasets (EMR, social media, surveys, interviews) to diagnose depression. ML and DL have shown great potential in diagnosing depression among different populations. There has been steadily growing interest in researching in this area and better outcomes have been noted. The first phase of analysis clearly showed this.

In the second phase, the analysis showed several insights as highlighted in the discussions above. One of the findings is the use of different sources of data to feed into ML and DL algorithms. Some of the most used sources were social media, interviews, surveys, and wearables among others. The choice of social media platforms as a source of data used to train ML algorithms is necessitated by the fact that there are numerous records of data being generated from them, be it Twitter, Facebook, Instagram, or TikTok among others that can potentially be used to detect depressive cases. Surveys were also noted to be another reliable source of data for building, training, and validating ML models. Surveys are usually done after a given period and they have enormous sources of data essential in ML and DL. Wearables are also a good source of data for detecting depression.

DL was the most recommended algorithm for solving depression as it performed well in accuracy, F1-score and AUC score hence was able to accurately predict depression cases with high precision and recall levels. For accuracy, other algorithms like RF, SVM, XGBoost and LR have shown varied successes with RF being the second most recommended algorithm. It was more interesting to learn that all the articles reviewed focused on supervised ML, unlike unsupervised or semi-supervised ML algorithms. All the articles reviewed were from developed countries and maybe there could be more pre-disposing factors in developing and under-developed countries which may tilt the performance of these algorithms.

RECOMMENDATIONS

Since different populations have different pre-disposing factors to depression, the study recommends future population-specific studies be done to find out which algorithm can best be used to predict depression in each population. The high penetration of social media and the availability of vast amounts of data on these sites come in handy, supporting the use of this data in training and predicting depression among the general population and specific populations.

REFERENCES

1. Aleem, S., Huda, N. u., Amin, R., Khalid, S., Alshamrani, S. S., & Alshehri, A. (2022). Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions. *Electronics*, 1-20. <https://doi.org/https://doi.org/10.3390/electronics11071111>
2. Alharahsheh, Y. E., & Abdullah, M. A. (2021). Predicting Individuals Mental Health Status in Kenya using Machine Learning Methods. *ResearchGate*. <https://doi.org/http://dx.doi.org/10.1109/ICICS52457.2021.9464608>
3. Bokolo, B. G., & Liu, Q. (2023). Deep Learning-Based Depression Detection from Social Media: Comparative Evaluation of ML and Transformer Techniques. *Electronics*.
4. Brereton, P., Kitchenham, B. A., Budgen, D., Turner, M., & Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, 571-583.
5. Dang, V. N., Cascarano, A., Mulder, R. H., Cecil, C., A.Zuluaga, M., Hernández-González, J., & Lekadir, K. (2023). Fairness and bias correction in machine learning for depression prediction: results from four different study populations. *arxiv*, 1-11. <https://doi.org/http://dx.doi.org/10.48550/arXiv.2211.05321>
6. Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 94-98.
7. Fatima, M., & Pasha, M. (2017). Survey of Machine Learning Algorithms for Disease Diagnostic. *Journal of Intelligent Learning Systems and Applications*, 1-16.
8. Ghosala, S., & Jainb, A. (2023). Depression and Suicide Risk Detection on Social Media using fastText Embedding and XGBoost Classifier. *International Conference on Machine Learning and Data Engineering*.
9. Inamdar, S., Chapekar, R., Gite, S., & Pradhan, B. (2023). Machine Learning Driven Mental Stress Detection on Reddit Posts Using Natural Language Processing. *Human-Centric Intelligent Systems*.
10. Joksimovic, S., Ifenthaler, D., Marrone, R., Laat, M. D., & Siemens, G. (2023). Opportunities of artificial intelligence for supporting complex problem-solving: Findings from a scoping review. *Computers and Education: Artificial Intelligence*.
11. Joseph, J., Senith, S., Kirubaraj, A. A., & Ramson, S. R. (2022). Machine Learning for Prediction of Clinical Appointment No-Shows. *International Journal of Mathematical, Engineering and Management Sciences*, 558-574. <https://doi.org/https://doi.org/10.33889/IJMEMS.2022.7.4.036>
12. Kotsiantis, S. B. (2007). Supervised Machine Learning: A Review of Classification Techniques. *Informatica*, 249-268.
13. Kung, B., Chiang, M., Perera, G., Pritchard, M., & Stewart, R. (2022). Unsupervised Machine Learning to Identify Depressive Subtypes. *Healthcare Informatics Research*, 256-266.
14. Marathe, G., Moodie1, E. E., Brouillette, M. -J., Cox1, J., Cooper, C., Delaunay, C. L., Conway, B., Hull, M., Martel-Laferrière, V., Vachon, M. -L., Walmsley, S., Wong, A., & Klein, M. B. (2022, August 12). Predicting the presence of depressive symptoms in the HIV-HCV co-infected population in Canada using supervised machine learning. *Springer*. <https://doi.org/https://doi.org/10.1186/s12874-022-01700-y>

15. Moon, N. N., Mariam, A., Sharmin, S., Islam, M. M., Nur, F. N., & Debnath, N. (2021). Machine learning approach to predict the depression in job sectors in Bangladesh. *Elsevier*.
16. Naeem, S., Ali, A., Anam, S., & Ahmed, M. (2023). An Unsupervised Machine Learning Algorithms: Comprehensive Review. *International Journal of Computing and Digital Systems*, 911-921.
17. Poudel, S. (2022). A Study of Disease Diagnosis Using Machine Learning. *Medical Science Forum*.
18. Qasrawi, R., Polo, S. V., Al-Halawah, D. A., Hallaq, S., & Abdeen, Z. (2021). Schoolchildren Depression and Anxiety Prediction Using Machine Learning Algorithms. *JMIR Publication*.
19. Qu, Z., Wang, Y., Guo, D., He, G., Sui, C., Duan, Y., Zhang, X., Lan, L., Meng, H., Wang, Y., & Liu, X. (2023). Identifying depression in the United States veterans using deep learning algorithms, NHANES 2005–2018. *BMC Psychiatry*.
20. Schaefer, J., Lehne, M., Schepers, J., Prasser, F., & Thun, S. (2020). The use of machine learning in rare diseases: a scoping review. *Orphanet Journal of Rare Diseases*.
21. Tornero-Costa, R., Martinez-Millana, A., Azzopardi-Muscat, N., Lazeri, L., Traver, V., & Novillo-Ortiz, D. (2023). Methodological and Quality Flaws in the Use of Artificial Intelligence in Mental Health Research: Systematic Review. *JMIR*. <https://doi.org/https://doi.org/10.2196/42045>
22. WHO. (2022). *World mental health report*. Retrieved November 28, 2023, from WHO. <https://iris.who.int/bitstream/handle/10665/356115/9789240050860-eng.pdf?sequence=1>
23. WHO. (2023, March 31). *Depressive disorder (depression)*. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/depression>
24. WHO. (2023, August 28). *Suicide*. World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/suicide>
25. Yan, W.-J., Ruan, Q.-N., & Jiang, K. (2022). Challenges for Artificial Intelligence in Recognizing. *diagnostics*, 13(2), 1-10. <https://doi.org/https://doi.org/10.3390/diagnostics13010002>