Depression Diagnosis using Text-based AI Methods – A Systematic Review

Martín Di Felice*, Parag Chatterjee and María F. Pollo-Cattaneo

Universidad Tecnológica Nacional Facultad Regional Buenos Aires, Buenos Aires, Argentina

Abstract

Recent years have seen increasing use of artificial intelligence in the domain of healthcare and mental health is no exception. This study is focused on the particular aspect of depression, which affects a significant percentage of the population and is an important concern globally. This systematic review analyzes different methods based on artificial intelligence to diagnose depression, highlighting the global trends of this domain like the huge share of natural language processing algorithms and neural networks on one hand, and illustrating the key issues and future lines of research in applying artificial intelligence in the domain of mental health, on the other hand.

Keywords

Machine Learning, Pattern Recognition, Mental Health, Depression

1. Introduction

Artificial Intelligence (AI) is one of the branches of the computer sciences that is in charge of solving complex, nonlinear problems that usually need human interaction. AI seeks to emulate human behavior in order to automate tasks in such a way they can be solved with a similar efficiency but faster [1].

In order to be able to emulate that human behavior, AI algorithms use big sets of data, called datasets [2]. While bigger, more complete, and more heterogeneous are these datasets, these algorithms may infer better the relationship between the data so they can generate rules that before the appearance of new data they can predict how they are going to behave [3, 4].

Due to the increase in computational power and the availability of more data, AI has increased its involvement in different fields during the last years [5]. One of the domains where it has even more involvement is healthcare [6]. It has increased its participation in the area of mental health but at a lower rate [7]. Although the ethical aspects of its use are still in debate [8], the benefits that come from its application seem quite promising [6, 9], including the speed of diagnosis and the fact of eliminating the expert subjectivity and replacing it with a science-based objective method.

ICAIW 2022: Workshops at the 5th International Conference on Applied Informatics 2022, October 27–29, 2022, Arequipa, Peru

*Corresponding author

mdifelice@frba.utn.edu.ar (M. Di Felice); parag@frba.utn.edu.ar (P. Chatterjee); flo.pollo@gmail.com (M. F. Pollo-Cattaneo)

0000-0003-1388-3220 (M. Di Felice); 0000-0001-6760-4704 (P. Chatterjee); 0000-0003-4197-3880 (M. F. Pollo-Cattaneo)

Although there exist objective and parameterized techniques for the diagnosis of mental health issues, and, in particular, for depression [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20], the use of AI methods not only offers the process of making the diagnosis, transforming data corresponding to symptoms into an output corresponding to a disease, but also helps to find those symptoms by transforming colloquial expressions into objective symptoms, and discover relationships between different types of symptoms as well.

Depression is a mental illness that is characterized by producing mood disorders over long periods of time, which can be for several weeks or more [21, 22]. It affects a significant percentage of the population [23], people of any age and it can be grouped into two large groups: major depressive disorder and persistent depressive disorder. There exist other forms that are also common but they happen with less frequency, such as postpartum depression, premenstrual dysphoric disorder, seasonal affective disorder, and psychotic disorder.

The usage of AI methods to diagnose depression using the analysis of data generated by patients have a high degree of effectiveness according to the present research.

This work intends to obtain a state of the art about the usage of AI methods to diagnose depression using text datasets. In order to define the state of the art, a Systematic Mapping Study (SMS) is made. The SMS is a standardized research process whose goal is to gather existent evidence about a particular topic [24] searching studies about it and summarizing them in order to obtain a conclusion.

2. Goals

The goal of the present study is to identify which research lines are open on the usage of AI techniques for depression diagnosis using text datasets, so eventually develop a new method that allows effectively diagnosing with the purpose of helping with the disease treatment.

An SMS is made following the process proposed by Petersen et al.[24] in order to determine the state of the art on this subject. This process establishes as the first step the definition of the questions that will lead the investigation. The research questions are the following:

- **RQ1:** Which AI method(s) are used to solve the problem?
- **RQ2:** What kind of learning is used to adjust the solution?
- **RQ3:** Which results are obtained after applying each method?
- **RQ4:** How are the results validated by each method?
- **RQ5:** What are the future open research lines?

In order to answer each one of these questions set as research goals for the present study, a systematic review related to depression diagnosis using text-based AI methods was performed. The following sources were used:

- IEEE Xplore[1]
- PubMed[2]
- Scopus[3]

[1]https://ieeexplore.ieee.org/Xplore/home.jsp
[2]https://pubmed.ncbi.nlm.nih.gov/
[3]https://www.scopus.com/home.uri

3. Results

A search and synthesis tool has been developed[4] in order to automate and standardize the search on these sources. The tool connects to each one of those sources using their public API and performs the search using the following terms:
`("artificial intelligence" OR "machine learning" OR "deep learning") AND ("depression diagnos*" OR "depression detection" OR "depression estimation")`

Then, a template was built and used as the foundation for the extraction formulary where each study appears along with their metadata: name, authors, published date, and link (wherever applicable). In order to perform a specific selection of the studies to be included in this review, the SMS establishes the definition of inclusion and exclusion criteria. The inclusion criteria are the following:

- Magazine articles, conference articles, or book chapters.
- Published year equal to or greater than 2012.
- Studies must use AI to solve the problem.
- Studies must diagnose depression.

While the exclusion criteria are:

- Studies must not perform prognosis nor predictions.
- Studies must not use non-text-based datasets.
- Studies must not be written in any other language than English.

The search was performed using the tool, meeting all the inclusion criteria and exclusion criteria, The initial search found a total of 192 articles and after applying the exclusion criteria, 45 articles were obtained. Then, for each article, the present work has been designed, attempting to answer (Table 1) the research questions defined above.

Table 1
Extraction formulary

Study	**RQ1**	**RQ2**	**RQ3**	**RQ4**	**RQ5**
Rao et al. [25]	NLP, NN	SL	0.63 (P), 0.57 (R), 0.6 (F1)	Self-Informed	Extend model
McGinnis et al. [26]	DT, KNN, LR. SVM	SL	DT: 0.69 (AUC) KNN: 0.73 (AUC) SVM: 0.76 (AUC) LR: 0.8 (AUC)	Experts	-
Cong et al. [27]	NLP, NN	SL	0.69 (P), 0.53 (R), 0.6 (F1)	Self-Informed	-

[4]https://github.com/mdifelice/hbs

Gerych et al. [28]	NN, SVM	UL	0.92 (AUC), 0.91 (F1)	Question-naires	Increase dataset, extend to other diseases, and extend to general population
Deshpande and Rao [29]	NB, NLP, SVM	SL	NB: 0.84 (P), 0.83 (R), 0.83 (F1) SVM: 0.8 (P), 0.79 (R), 0.8 (F1)	Keywords	Improve validation
Wang et al. [30]	NLP, NN	SL	0.97 (ACC), 0.97 (F1), 0.99 (P), 0.95 (R)	Experts	Increase dataset
Malviya et al. [31]	LR, NB, NLP, NN, RF, SVM, XGB	SL	NN: 0.98 (ACC), 0.98 (F1) SVM: 0.92 (ACC), 0.92 (F1)	Keywords	Increase dataset
Hassan et al. [32]	DT, KNN, LR, NB, SVM	SL	KNN: 0.79 (ACC), 0.6 (R), 0.72 (P), 0.65 (F1) LR: 0.77 (ACC), 0.5 (R), 0.39 (P), 0.44 (F1) SVM: 0.77 (ACC), 0.5 (R), 0.39 (P), 0.44 (F1) NB: 0.77 (ACC), 0.5 (R), 0.39 (P), 0.44 (F1)	Question-naires	Tool creation
Kumar et al. [33]	DT, KNN, LR, NB, SVM	SL	KNN+LR+SVM: 0.9 (ACC) DT+NB+SVM: 0.88 (ACC)	Self-Informed	-
Uddin et al. [34]	NLP, NN	SL	0.86 (ACC)	Experts	-
Victor et al. [35]	DT, KNN, NB, NLP, RF, SVM	SL	0.9 (ACC)	Experts	Increase dataset
Chiong et al. [36]	AB, BP, DT, GB, LR, NLP, NN, RF, SVM	SL	AB+BP+GB+RF: 0.98 (ACC) DT+LR+NN+SVM: 0.96 (ACC)	Keywords	Improve validation
Arun et al. [37]	XGB	SL	0.98 (ACC)	Experts Question-naires	-
Raihan et al. [38]	AB, NN, RF	SL	AB: 0.98 (ACC), NN: 0.83 (ACC), RF: 0.72 (ACC)	Experts	Increase dataset

Govindasamy and Palanichamy [39]	DT, NB, NLP	SL	0.97 (ACC)	Sentiment Analysis	Determine depression level
Al Asad e t al. [40]	NB, NLP, SVM	SL	0.74 (ACC), 1 (P), 0.6 (R)	Question-naires	Include other languages
Tadesse et al. [41]	AB, LR, RF, NLP, NN, SVM	SL	AB: 0.79 (ACC), 0.81 (F1), 0.72 (P), 0.93 (R) LR: 0.89 (ACC), 0.89 (F1), 0.89 (P), 0.92 (R) NN: 0.91 (ACC), 0.93 (F1), 0.9 (P), 0.92 (R) RF: 0.85 (ACC), 0.85 (F1), 0.83 (P), 0.87 (R) SVM: 0.9 (ACC), 0.91 (F1), 0.89 (P), 0.93 (R)	Keywords	Study relation-ship with per-sonality
Shah et al. [42]	NLP, NN	SL	0.81 (F1)	Self-Informed	Improve perfor-mance
Bhat et al. [43]	NLP, NN	SL	0.98 (ACC)	Sentiment Analysis	-
Santana et al. [44]	GA, KNN, RF, SVM	SL	KNN: 0.95 (F1), 0.96 (P), 0.95 (R) RF: 0.86 (F1), 0.85 (P), 0.84 (R) SVM: 0.93 (F1), 0.93 (P), 0.93 (R)	Question-naires	-
Opuku Asare et al. [45]	DT, KNN, LR, RF, SVM, XGB	SL	DT: 0.47 (ACC), 0.19 (P), 0.74 (R) KNN: 0.96 (ACC), 0.86 (P), 0.92 (R) LR: 0.59 (ACC), 0.20 (P), 0.58 (R) RF: 0.98 (ACC), 0.93 (P), 0.94 (R) SVM: 0.86 (ACC), 0.52 (P), 0.81 (R) XGB: 0.98 (ACC), 0.93 (P), 0.96 (R)	Question-naires	Increase dataset
Hemmatirad et al. [46]	NLP, SVM	SL	0.96 (F1), 0.97 (P), 0.96 (R), 0.95 (ACC)	Keywords Self-Informed	Add more mod-els
Zogan et al. [47]	NLP, NN	SL	0.9 (ACC), 0.9 (P), 0.89 (R), 0.89 (F1)	Keywords	Improve val-idation and increase dataset

Haque et al. [48]	DT, NB, RF, XGB	SL	XGB: 0.95 (ACC), 0.85 (P), 0.99 (SPC), 0.48 (R) RF: 0.95 (ACC), 0.99 (P), 1 (SPC), 0.44 (R) DT: 0.95 (ACC), 0.94 (P), 1 (SPC), 0.45 (R) NB: 0.94 (ACC), 0.69 (P), 0.98 (SPC), 0.51 (R)	Experts	-
Chiong et al. [49]	AB, BP, DT, GB, LR, NLP, NN, RF, SVM	SL	DT: 0.82 (ACC), 0.83 (P), 0.84 (R), 0.84 (F1) LR: 0.93 (ACC), 0.93 (P), 0.72 (R), 0.81 (F1) NN: 0.85 (ACC), 0.87 (P), 0.86 (R), 0.86 (F1) SVM: 0.87 (ACC), 0.9 (P), 0.87 (R), 0.88 (F1)	Keywords	Include unsupervised learning
Narziev et al. [50]	RF, SVM	SL	0.96 (ACC)	Questionnaires	Increase dataset
Islam et al. [51]	DT, EL, KNN, NLP, SVM	SL	DT: 0.71 (ACC) KNN: 0.6 (ACC) SVM: 0.71 (ACC) EL: 0.64 (ACC)	Keywords	Increase dataset
Zogan et al. [52]	NLP, NN	SL, UL	0.91 (P), 0.9 (R), 0.91 (F1), 0.9 (ACC)	Keywords	Increase dataset
Xu et al. [53]	ES	SL	0.79 (ACC), 0.81 (P), 0.85 (R), 0.83 (F1)	Questionnaires	Tool creation
Shrestha et al. [54]	DBSCAN, GMM, IF, KM, NLP, SVM	UL	KM: 0.63 (P), 0.61 (R), 0.51 (F1) GMM: 0.64 (P), 0.64 (R), 0.64 (F1) DBSCAN: 0.77 (P), 0.42 (R), 0.27 (F1) IF: 0.62 (P), 0.62 (R), 0.54 (F1) SVM: 0.6 (P), 0.59 (R), 0.59 (F1)	Experts	Increase dataset

Alsagri and Ykhlef [55]	DT, NB, NLP, SVM	SL	DT: 0.78 (ACC), 0.59 (R), 0.62 (F1), 0.78 (P), 0.6 (AUC) NB: 0.8 (ACC), 0.81 (R), 0.72 (F1), 0.65 (P), 0.67 (AUC) SVM: 0.83 (ACC), 0.85 (R), 0.79 (F1), 0.74 (P), 0.78 (AUC)	Keywords	Add more models
Xezonaki et al. [56]	NLP, NN, SVM	SL	0.72 (F1)	Experts Questionnaires	Tool creation
Ramiandrisoa and Mothe [57]	LR, NLP, RF	SL	LR: 0.51 (F1), 0.38 (P), 0.8 (R) LR+RF: 0.51 (F1), 0.38 (P), 0.8 (R) RF: 0.58 (F1), 0.69 (P), 0.51 (R)	Self-Informed	Increase dataset
Burdisso et al. [58]	ES, NLP	SL	0.61 (F1), 0.63 (P), 0.6 (R)	Self-Informed	Extend to other diseases
Stankevich et al. [59]	NLP, RF, SVM	SL	SVM: 0.58 (P), 0.77 (R), 0.66 (F1) RF: 0.63 (P), 0.53 (R), 0.6 (F1)	Questionnaires	Add more models
Choi et al. [60]	GMM, LR	SSL, UL	UL: 3.105 (ANOVA), 2.732 (ANOVA)	Experts Questionnaires	Increase dataset
Khan et al. [61]	NLP, NN	SL	0.96 (ACC), 0.99 (P), 0.95 (F1), 0.93 (R), 0.98 (SPC)	Sentiment Analysis	Tool creation
Zhang et al. [62]	LR, NLP, RF, SVM	SL	RF: 0.78 (ACC), 0.78 (F1), 0.85 (AUC) LR: 0.78 (ACC), 0.79 (F1), 0.86 (AUC) SVM: 0.79 (ACC), 0.79 (F1), 0.86 (AUC)	Experts	-
Ren et al. [63]	NLP, NN	SL	0.91 (ACC), 0.92 (P), 0.96 (R), 0.94 (F1)	Experts	Extend to other diseases
Amanat et al. [64]	NLP, NN	SL	0.98 (P), 0.99 (R), 0.98 (F1)	Experts	Extend to other diseases

Almars [65]	NLP, NN	SL	Negatives: 0.98 (P), 0.84 (R), 0.85 (F1) Positives: 0.78 (P), 0.83 (R), 0.81 (F1)	Experts	Increase dataset
Inkpen at al. [66]	NLP, NN	SL	HAN: 0.33 (average hit rate), 0.66 (average closeness rate) BERT: 0.79 (average difference between overall depression levels) RoBERTa: 0.3 (depression category hit rate)	Self-Informed	Improve performance
Wu et al. [67]	NLP, NN	SL	0.83 (P), 0.71 (R), 0.77 (F1)	Questionnaires	Improve performance, extend to other diseases, and tool creation
Shah et al. [68]	NB, NLP, DT, SVM, SGD, RF	SL	NB: 0.8 (ACC), 0.61 (P), 0.4 (R), 0.48 (F1) DT: 0.8 (ACC), 0.58 (P), 0.53 (R), 0.55 (F1) SVM: 0.77 (ACC) SGD: 0.79 (ACC), 0.55 (P), 0.54 (R), 0.54 (F1) RF: 0.83 (ACC), 0.77 (P), 0.4 (R), 0.53 (F1)	Experts	-
Stankevich et al. [69]	AB, GB, LR, NB, NLP, NN, RF, SVM, XGB	SL	RF: 0.74 (AUC), 0.59 (P), 0.71 (R), 0.65 (F1) AB: 0.72 (AUC), 0.57 (P), 0.68 (R), 0.62 (F1) LR: 0.69 (AUC), 0.51 (P), 0.68 (R), 0.58 (F1) XGB: 0.68 (AUC), 0.48 (P), 0.74 (R), 0.58 (F1) SVM: 0.67 (AUC), 0.44 (P), 0.84 (R), 0.58 (F1)	Questionnaires	Add more models

AB: AdaBoost, ACC: Accuracy, ANOVA: Analysis of Variance, AUC: Area Under the Curve, BP: Bagging Predictors, DBSCAN: Density-Based Spatial Clustering of Applications with Noise, DT: Decision Trees, EL: Ensembled Learning, ES: Expert System, F1: F-Score, GA: Genetic Algorithms, GB: GradientBoost, GMM: Gaussian Mixture Model, IF: Isolation Forest, KM: K-Means, KNN: K-

Nearest Neighbors, LR: Logistic Regression, NB: Naive Bayes, NLP: Natural Language Processing, NN: Neural Networks, P: Precision, R: Recall, RF: Random Forest, SGD: Stochastic Gradient Descent, SL: Supervised Learning, SPC: Specificity, SSL: Semi-Supervised Learning, SVM: Support Vector Machines, UL: Unsupervised Learning, XGB: XGBoost.

3.1. RQ1: Which AI method or methods are used to solve the problem?

Most of the studies use more than one type of algorithm, with Neural Networks (NN) (14.2%), Support Vector Machines (14.2%), and Natural Language Processing (NLP) (20.4%) having the highest share. Rao et al. [25], Cong et al. [27], Wang et al. [30], Uddin et al. [34], Shah et al. [42], Bhat et al. [43], Zogan et al. [47], Zogan et al. [52], Khan et al. [61], Ren et al. [63], Amanat et al. [64], Almars [65], Inkpen et al. [66] and Wu et al. [67] use the combination of NLP and NN algorithms on their methods. Deshpande and Rao [29], Malviya et al. [31], Victor et al. [35], Chiong et al. [36], Govindasamy and Palanichamy [39], Al Asad et al. [40], Tadesse et al. [41], Hemmatirad et al. [46], Chiong et al. [49], Islam et al. [51], Shrestha et al. [54], Xezonaki et al. [56], Ramiandrisoa and Mothe [57], Burdisso et al. [58], Stankevich et al. [59], Zhang et al. [62], Shah et al. [68] and Stankevich et al. [69] on the other hand, use NLP along with other kind of algorithms (including NN but not exclusively). Finally, Gerych et al. [28] and Raihan et al. [38] used NN combined with other kinds of algorithms but not NLP. This way, 34 of the 45 primary studies (a 75.6%) used NN or NLP algorithms.

From the rest of the studies, the most observed combinations are Decision Trees (DT), K-Nearest Neighbors (KNN), Logistic Regression (LR) and Support Vector Machines (SVM) in McGinnis et al. [26], Hassan et al. [32], Kumar et al. [33][31], Victor et al. [35] and Opuku Asare et al. [45]. Figure 1 shows the algorithm type distribution and Figure 2 shows the NLP and NN prevalence.

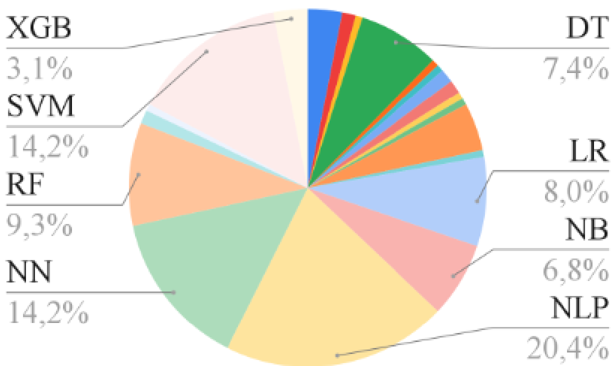

Figure 1: Algorithm Types

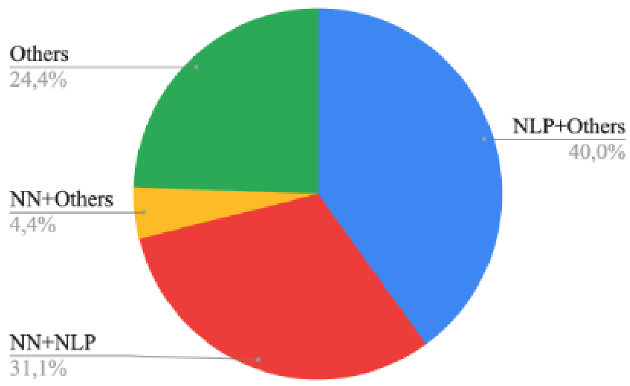

Figure 2: NLP and NN prevalence

3.2. RQ2: What kind of learning is used to adjust the solution?

An important tendency towards the use of supervised learning has been observed (91.1%). Only Gerych et al. [28] and Shresta et al. [54] have chosen to investigate methods based on unsupervised learning. Zogan et al. [52] and Choi et al. [60] use hybrid methods combining supervised and unsupervised learning, and unsupervised and semi-supervised learning respectively.

3.3. RQ3: Which results are obtained after applying each method?

Both the used metrics and the obtained results vary from one study to another. Rao et al. [25], McGinnis et al. [26], Deshpande and Rao [29], Malviya et al. [31], Hassan et al. [32], Kumar et al. [33], Victor et al. [35], Chiong et al. [36], Raihan et al. [38], Tadesse et al. [41], Bhat et al. [43], Santana et al. [44], Opuku Asare et al. [45], Haque et al. [48], Islam et al. [51], Shrestha et al. [54], Alsagri and Ykhlef [55], Ramiandrisoa and Mothe [57], Stankevich et al. [59], Khan et al. [61], Zhang et al. [62], Shah et al. [68], and Stankevich et al. [69] use different models and compare them to see which ones work better. Govindasamy and Palanichamy [39] and Xezonaki et al. [56] also illustrate multiple results, but comparing the same model with different datasets; and Chiong et al. [49] use several models on two different datasets.

The most used metrics are accuracy (25.1%), recall (22.9%), precision (22.3%), and F1 (21.3%). Figure 3 shows the most used metrics.

The accuracy rises from 0.47 to 0.98, with an average of 0.86 and a median of 0.9; recall goes from 0.33 to 0.99, with an average of 0.75 and a median of 0.79; precision goes from 0.19 to 1, with an average of 0.76 and a median of 0.83; and finally, F1 goes from 0.27 to 0.98, with an average of 0.82 and a median of 0.85.

3.4. RQ4: How are the results validated by each method?

Primarily, five different validation types were identified. McGinnis et al. [26], Wang et al. [30], Uddin et al. [34], Victor et al. [35], Raihan et al. [38], Haque et al. [48], Shrestha et al. [54],

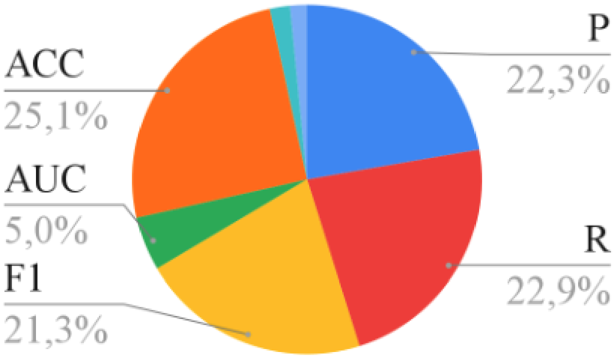

Figure 3: Metrics

Xezonaki et al. [56], Zhang et al. [62], Ren et al. [63], Amanat et al. [64], Almars [65], and Shah et al. [68] use the analysis of experts to determine if a record belongs to a depression patient or not. This is the most used validation type with 30.6% of the cases.

On the other hand, the second most used validation type is the use of questionnaires. Gerych et al. [28], Hassan et al. [32], Al Asad et al. [40], Santana et al. [44], Opoku Asare et al. [45], Narziev et al. [50], Xu et al. [53], Stankevich et al. [59], Wu et al. [67] and Stankevich et al. [69] use this kind of validation (26.5% of the total).

In third place, with the 20.4% of the cases, Deshpande and Rao [29], Malviya et al. [31], Chiong et al. [36], Tadesse et al. [41], Hemmatirad et al. [46], Zogan et al. [47], Chiong et al. [49], Islam et al. [51], Zogan et al. [52] and Alsagri and Ykhlef [55] search for keywords inside the datasets to determine if a patient if depressive or not.

Also, some studies (16.3%) use datasets where the labeling is made by the participants themselves (self-informed). Rao et al. [25], Cong et al. [27], Kumar et al. [33], Shah et al. [42], Ramiandrisoa and Mothe [57], Burdisso et al. [58] and Inkpen et al. [66] use these datasets.

And, in the last instance, with 6.1% of the distribution, Govindasamy and Palanichamy [39], Bhat et al. [43] and Khan et al. [61] use sentiment analysis to label their datasets. Figure 4 shows the validation types.

3.5. RQ5: What are the future open research lines?

Among the studies that mention future work (82.2%), the most mentioned one is increasing the dataset or datasets used for the experiment. Gerych et al. [28], Wang et al. [30], Malviya et al. [31], Victor et al. [35], Raihan et al. [38], Opuku Asare et al. [45], Zogan et al. [47], Narziev et al. [50], Islam et al. [51], Zogan et al. [52], Shrestha et al. [54], Ramiandrisoa and Mothe [57], Choi et al. [60] and Almars [65] mention this possibility.

Gerych et al. [28], Burdisso et al. [58], Ren et al. [63], Amanat et al. [64] and Wu et al. [67] indicate that in the future they would be willing to extend their models to diagnose other diseases. Hassan et al. [32], Xu et al. [53], Xezonaki et al. [56], Khan et al. [61] and Wu et al. [67] propose to create a tool or a practical application of their model. Figure 5 shows all the possibilities.

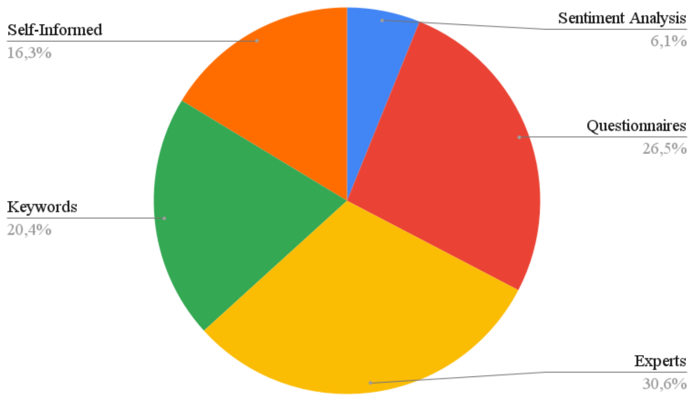

Figure 4: Validation Types

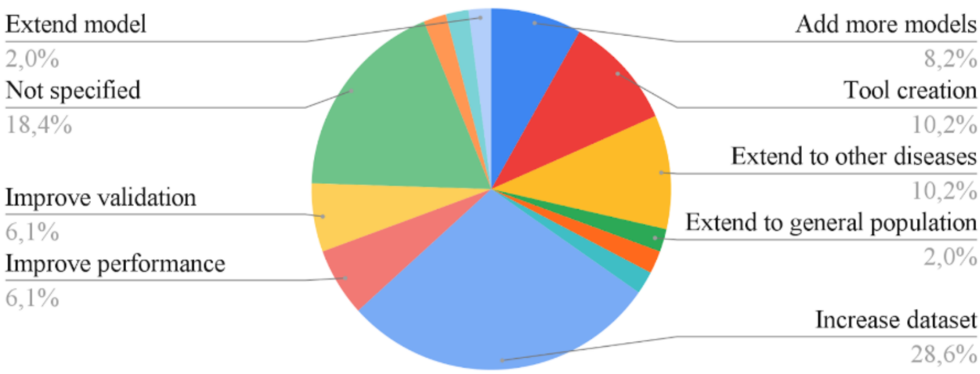

Figure 5: Future work

4. Conclusions

This systematic review was performed to highlight the state of the art in the domain of depression diagnosis through AI using text-based methods. Analyzing the relevant literature, it is concluded that NLP and NN are the most used algorithms, used in conjunction with colloquial text-based datasets, mostly extracted from social networks. In most of the studies, the lack of sufficiently big datasets was stated, illustrating the demand for larger datasets for future work. Also, the use of supervised learning preferred over using unsupervised learning has been noted, whereas, only a small section of the studies has opted for unsupervised learning. As the domain of mental health embraces AI tools for different purposes like diagnosis and predictions of mental health issues, aspects like the generation of significantly large databases are indispensable for better training of the algorithms. Especially in the area of depression, this also opens the possibility of studies in larger domains, providing more reliable and reusable AI models for diagnosis.

5. Acknowledgements

This work was supported and financed by the Cloudgenia group through its technical and operational initiatives.

References

[1] M. A. Boden, Artificial intelligence, Elsevier, 1996.

[2] A. H. Renear, S. Sacchi, K. M. Wickett, Definitions of dataset in the scientific and technical literature, Proceedings of the American Society for Information Science and Technology 47 (2010) 1–4.

[3] A. Ajiboye, R. Abdullah-Arshah, H. Qin, H. Isah-Kebbe, Evaluating the effect of dataset size on predictive model using supervised learning technique, Int. J. Comput. Syst. Softw. Eng 1 (2015) 75–84.

[4] F. Velosa, H. Florez, Edge solution with machine learning and open data to interpret signs for people with visual disability, CEUR Workshop Proceedings (2020) 15–26.

[5] R. L. Villars, C. W. Olofson, M. Eastwood, Big data: What it is and why you should care, White paper, IDC 14 (2011) 1–14.

[6] S. Graham, C. Depp, E. E. Lee, C. Nebeker, X. Tu, H.-C. Kim, D. V. Jeste, Artificial intelligence for mental health and mental illnesses: an overview, Current psychiatry reports 21 (2019) 1–18.

[7] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, Q. Dong, H. Shen, Y. Wang, Artificial intelligence in healthcare: past, present and future, Stroke and vascular neurology 2 (2017).

[8] J. Morley, C. C. Machado, C. Burr, J. Cowls, I. Joshi, M. Taddeo, L. Floridi, The ethics of ai in health care: a mapping review, Social Science & Medicine 260 (2020) 113172.

[9] B. X. Tran, R. S. McIntyre, C. A. Latkin, H. T. Phan, G. T. Vu, H. L. T. Nguyen, K. K. Gwee, C. S. Ho, R. C. Ho, The current research landscape on the artificial intelligence application in the management of depressive disorders: a bibliometric analysis, International journal of environmental research and public health 16 (2019) 2150.

[10] K. Kroenke, T. W. Strine, R. L. Spitzer, J. B. Williams, J. T. Berry, A. H. Mokdad, The phq-8 as a measure of current depression in the general population, Journal of affective disorders 114 (2009) 163–173.

[11] K. Kroenke, R. L. Spitzer, The phq-9: a new depression diagnostic and severity measure, 2002.

[12] American Psychological Association, Beck depression inventory (bdi),, 2022.

[13] A. T. Beck, R. A. Steer, G. Brown, Beck depression inventory–ii, Psychological assessment (1996).

[14] PsycNet, Children's depression inventory, 2022.

[15] A. P. Association, Apa - the structured clinical interview for dsm-5, 2022.

[16] S. Shiffman, A. A. Stone, M. R. Hufford, Ecological momentary assessment, Annu. Rev. Clin. Psychol. 4 (2008) 1–32.

[17] J. Johnson, Resilience appraisals scale, 2022.

[18] R. L. Spitzer, K. Kroenke, J. B. Williams, B. Löwe, A brief measure for assessing generalized anxiety disorder: the gad-7, Archives of internal medicine 166 (2006) 1092–1097.
[19] L. S. Radloff, The use of the center for epidemiologic studies depression scale in adolescents and young adults, Journal of youth and adolescence 20 (1991) 149–166.
[20] J. B. Awotunde, S. A. Ajagbe, H. Florez, Internet of things with wearable devices and artificial intelligence for elderly uninterrupted healthcare monitoring systems, in: International Conference on Applied Informatics, Springer, 2022, pp. 278–291.
[21] F. Edition, et al., Diagnostic and statistical manual of mental disorders, Am Psychiatric Assoc 21 (2013) 591–643.
[22] E. O. Ogunseye, C. A. Adenusi, A. C. Nwanakwaugwu, S. A. Ajagbe, S. O. Akinola, Predictive analysis of mental health conditions using adaboost algorithm, ParadigmPlus 3 (2022) 11–26.
[23] G. Y. Lim, W. W. Tam, Y. Lu, C. S. Ho, M. W. Zhang, R. C. Ho, Prevalence of depression in the community from 30 countries between 1994 and 2014, Scientific reports 8 (2018) 1–10.
[24] K. Petersen, R. Feldt, S. Mujtaba, M. Mattsson, Systematic mapping studies in software engineering, in: 12th International Conference on Evaluation and Assessment in Software Engineering (EASE) 12, 2008, pp. 1–10.
[25] G. Rao, Y. Zhang, L. Zhang, Q. Cong, Z. Feng, Mgl-cnn: a hierarchical posts representations model for identifying depressed individuals in online forums, IEEE Access 8 (2020) 32395–32403.
[26] R. S. McGinnis, E. W. McGinnis, J. Hruschak, N. L. Lopez-Duran, K. Fitzgerald, K. L. Rosenblum, M. Muzik, Rapid anxiety and depression diagnosis in young children enabled by wearable sensors and machine learning, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2018, pp. 3983–3986.
[27] Q. Cong, Z. Feng, F. Li, Y. Xiang, G. Rao, C. Tao, Xa-bilstm: A deep learning approach for depression detection in imbalanced data, in: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2018, pp. 1624–1627.
[28] W. Gerych, E. Agu, E. Rundensteiner, Classifying depression in imbalanced datasets using an autoencoder-based anomaly detection approach, in: 2019 IEEE 13th International Conference on Semantic Computing (ICSC), IEEE, 2019, pp. 124–127.
[29] M. Deshpande, V. Rao, Depression detection using emotion artificial intelligence, in: 2017 international conference on intelligent sustainable systems (iciss), IEEE, 2017, pp. 858–862.
[30] Y. Wang, Z. Wang, C. Li, Y. Zhang, H. Wang, A multimodal feature fusion-based method for individual depression detection on sina weibo, in: 2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC), IEEE, 2020, pp. 1–8.
[31] K. Malviya, B. Roy, S. Saritha, A transformers approach to detect depression in social media, in: 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), IEEE, 2021, pp. 718–723.
[32] M. M. Hassan, M. A. R. Khan, K. K. Islam, M. M. Hassan, M. F. Rabbi, Depression detection system with statistical analysis and data mining approaches, in: 2021 International Conference on Science & Contemporary Technologies (ICSCT), IEEE, 2021, pp. 1–6.
[33] P. Kumar, R. Chauhan, T. Stephan, A. Shankar, S. Thakur, A machine learning implementation for mental health care. application: Smart watch for depression detection, in:

2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), IEEE, 2021, pp. 568–574.
[34] A. H. Uddin, D. Bapery, A. S. M. Arif, Depression analysis from social media data in bangla language using long short term memory (lstm) recurrent neural network technique, in: 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), IEEE, 2019, pp. 1–4.
[35] D. B. Victor, J. Kawsher, M. S. Labib, S. Latif, Machine learning techniques for depression analysis on social media-case study on bengali community, in: 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, 2020, pp. 1118–1126.
[36] R. Chiong, G. S. Budhi, S. Dhakal, Combining sentiment lexicons and content-based features for depression detection, IEEE Intelligent Systems 36 (2021) 99–105.
[37] V. Arun, V. Prajwal, M. Krishna, B. Arunkumar, S. Padma, V. Shyam, A boosted machine learning approach for detection of depression, in: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2018, pp. 41–47.
[38] M. Raihan, A. K. Bairagi, S. Rahman, A machine learning based study to predict depression with monitoring actigraph watch data, in: 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), IEEE, 2021, pp. 1–5.
[39] K. A. Govindasamy, N. Palanichamy, Depression detection using machine learning techniques on twitter data, in: 2021 5th international conference on intelligent computing and control systems (ICICCS), IEEE, 2021, pp. 960–966.
[40] N. Al Asad, M. A. M. Pranto, S. Afreen, M. M. Islam, Depression detection by analyzing social media posts of user, in: 2019 IEEE International Conference on Signal Processing, Information, Communication & Systems (SPICSCON), IEEE, 2019, pp. 13–17.
[41] M. M. Tadesse, H. Lin, B. Xu, L. Yang, Detection of depression-related posts in reddit social media forum, IEEE Access 7 (2019) 44883–44893.
[42] F. M. Shah, F. Ahmed, S. K. S. Joy, S. Ahmed, S. Sadek, R. Shil, M. H. Kabir, Early depression detection from social network using deep learning techniques, in: 2020 IEEE Region 10 Symposium (TENSYMP), IEEE, 2020, pp. 823–826.
[43] P. Bhat, A. Anuse, R. Kute, R. Bhadade, P. Purnaye, Mental health analyzer for depression detection based on textual analysis, Journal of Advances in Information Technology Vol 13 (2022).
[44] R. Santana, B. Santos, T. Lima, M. Teodoro, S. Pinto, L. Zárate, C. Nobre, Genetic algorithms for feature selection in the children and adolescents depression context, in: 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), IEEE, 2019, pp. 1470–1475.
[45] K. O. Asare, Y. Terhorst, J. Vega, E. Peltonen, E. Lagerspetz, D. Ferreira, et al., Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study, JMIR mHealth and uHealth 9 (2021) e26540.
[46] K. Hemmatirad, H. Bagherzadeh, E. Fazl-Ersi, A. Vahedian, Detection of mental illness risk on social media through multi-level svms, in: 2020 8th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), IEEE, 2020, pp. 116–120.
[47] H. Zogan, I. Razzak, X. Wang, S. Jameel, G. Xu, Explainable depression detection with

multi-aspect features using a hybrid deep learning model on social media, World Wide Web 25 (2022) 281–304.
[48] U. M. Haque, E. Kabir, R. Khanam, Detection of child depression using machine learning methods, PLoS one 16 (2021) e0261131.
[49] R. Chiong, G. S. Budhi, S. Dhakal, F. Chiong, A textual-based featuring approach for depression detection using machine learning classifiers and social media texts, Computers in Biology and Medicine 135 (2021) 104499.
[50] N. Narziev, H. Goh, K. Toshnazarov, S. A. Lee, K.-M. Chung, Y. Noh, Stdd: Short-term depression detection with passive sensing, Sensors 20 (2020) 1396.
[51] M. Islam, M. A. Kabir, A. Ahmed, A. R. M. Kamal, H. Wang, A. Ulhaq, et al., Depression detection from social network data using machine learning techniques, Health information science and systems 6 (2018) 1–12.
[52] H. Zogan, I. Razzak, S. Jameel, G. Xu, Depressionnet: learning multi-modalities with user post summarization for depression detection on social media, in: proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval, 2021, pp. 133–142.
[53] X. Xu, P. Chikersal, J. M. Dutcher, Y. S. Sefidgar, W. Seo, M. J. Tumminia, D. K. Villalba, S. Cohen, K. G. Creswell, J. D. Creswell, et al., Leveraging collaborative-filtering for personalized behavior modeling: A case study of depression detection among college students, Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 5 (2021) 1–27.
[54] A. Shrestha, E. Serra, F. Spezzano, Multi-modal social and psycho-linguistic embedding via recurrent neural networks to identify depressed users in online forums, Network Modeling Analysis in Health Informatics and Bioinformatics 9 (2020) 1–11.
[55] H. S. AlSagri, M. Ykhlef, Machine learning-based approach for depression detection in twitter using content and activity features, IEICE Transactions on Information and Systems 103 (2020) 1825–1832.
[56] D. Xezonaki, G. Paraskevopoulos, A. Potamianos, S. Narayanan, Affective conditioning on hierarchical networks applied to depression detection from transcribed clinical interviews, arXiv preprint arXiv:2006.08336 (2020).
[57] F. Ramiandrisoa, J. Mothe, Early detection of depression and anorexia from social media: A machine learning approach, in: Circle 2020, volume 2621, 2020, pp. 1–12.
[58] S. G. Burdisso, M. Errecalde, M. Montes-y Gómez, A text classification framework for simple and effective early depression detection over social media streams, Expert Systems with Applications 133 (2019) 182–197.
[59] M. Stankevich, I. Smirnov, N. Kiselnikova, A. Ushakova, Depression detection from social media profiles, in: International Conference on Data Analytics and Management in Data Intensive Domains, Springer, 2019, pp. 181–194.
[60] B. Choi, G. Shim, B. Jeong, S. Jo, Data-driven analysis using multiple self-report questionnaires to identify college students at high risk of depressive disorder, Scientific reports 10 (2020) 1–13.
[61] M. R. Khan, S. Z. Rizvi, A. Yasin, M. Ali, Depression analysis of social media activists using the gated architecture bi-lstm, in: 2021 International Conference on Cyber Warfare and Security (ICCWS), IEEE, 2021, pp. 76–81.

[62] Y. Zhang, H. Lyu, Y. Liu, X. Zhang, Y. Wang, J. Luo, et al., Monitoring depression trends on twitter during the covid-19 pandemic: observational study, JMIR infodemiology 1 (2021) e26769.

[63] L. Ren, H. Lin, B. Xu, S. Zhang, L. Yang, S. Sun, et al., Depression detection on reddit with an emotion-based attention network: algorithm development and validation, JMIR Medical Informatics 9 (2021) e28754.

[64] A. Amanat, M. Rizwan, A. R. Javed, M. Abdelhaq, R. Alsaqour, S. Pandya, M. Uddin, Deep learning for depression detection from textual data, Electronics 11 (2022) 676.

[65] A. M. Almars, Attention-based bi-lstm model for arabic depression classification, CMC-COMPUTERS MATERIALS & CONTINUA 71 (2022) 3091–3106.

[66] D. Inkpen, R. Skaik, P. Buddhitha, D. Angelov, M. T. Fredenburgh, uottawa at erisk 2021: Automatic filling of the beck's depression inventory questionnaire using deep learning., in: CLEF (Working Notes), 2021, pp. 966–980.

[67] M. Y. Wu, C.-Y. Shen, E. T. Wang, A. L. Chen, A deep architecture for depression detection using posting, behavior, and living environment data, Journal of Intelligent Information Systems 54 (2020) 225–244.

[68] E. Shah, M. K. Ahsan, I. Mazahir, Machine learning based methodology for depressive sentiment analysis, in: International Conference on Intelligent Technologies and Applications, Springer, 2020, pp. 93–99.

[69] M. Stankevich, I. Smirnov, N. Kiselnikova, A. Ushakova, Depression detection from social media profiles, in: International Conference on Data Analytics and Management in Data Intensive Domains, Springer, 2019, pp. 181–194.