Journal of Psychiatric Research 132 (2021) 1–6

Contents lists available at ScienceDirect

Journal of Psychiatric Research

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

Can machine learning be useful as a screening tool for depression in primary care?

Erito Marques de Souza Filho[a,b,*], Helena Cramer Veiga Rey[c], Rose Mary Frajtag[c], Daniela Matos Arrowsmith Cook[d], Lucas Nunes Dalbonio de Carvalho[a], Antonio Luiz Pinho Ribeiro[e], Jorge Amaral[b]

[a] *Universidade Federal Rural do Rio de Janeiro, Rio de Janeiro, Brazil*
[b] *Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brazil*
[c] *Instituto Nacional de Cardiologia, Rio de Janeiro, Brazil*
[d] *ProCardiaco, Rio de Janeiro, Brazil*
[e] *Universidade Federal de Minas Gerais, Minas Gerais, Brazil*

ARTICLE INFO

Keywords:
Depression
Diagnosis
Machine learning

ABSTRACT

Depression is a widespread disease with a high economic burden and a complex pathophysiology disease that is still not wholly clarified, not to mention it usually is associated as a risk factor for absenteeism at work and suicide. Just 50% of patients with depression are diagnosed in primary care, and only 15% receive treatment. Stigmatization, the coexistence of somatic symptoms, and the need to remember signs in the past two weeks can contribute to explaining this situation. In this context, tools that can serve as diagnostic screening are of great value, as they can reduce the number of undiagnosed patients. Besides, Artificial Intelligence (AI) has enabled several fruitful applications in medicine, particularly in psychiatry. This study aims to evaluate the performance of Machine Learning (ML) algorithms in the detection of depressive patients from the clinical, laboratory, and sociodemographic data obtained from the Brazilian National Network for Research on Cardiovascular Diseases from June 2016 to July 2018. The results obtained are promising. In one of them, Random Forests, the accuracy, sensibility, and area under the receiver operating characteristic curve were, respectively, 0.89, 0.90, and 0.87.

1. Introduction

Depression is a ubiquitous disease, with an estimated prevalence of more than 264 million (Global Burden of Disease Study, 2018). The economic burden is equally high. In Canada, for example, it was estimated at 12 billion a year (Tanner et al., 2019). It is a complex pathophysiology disease that is still not completely clarified, which appears twice as often in women, and usually associated as a risk factor for absenteeism at work and suicide (Kessler et al., 2003; Park and Zarate, 2019). According to the Diagnostic and Statistical Manual of Mental Disorders, the diagnosis of depression consists of presence of 5 out of 9 symptoms for two weeks. Also, one of these symptoms must be depressed mood and loss of interest or pleasure in activities of daily life (DSM-V, 2013). The treatment of depression involves the use of different methods, such as the use of antidepressants and psychotherapy; nevertheless, many individuals do not receive adequate treatment, simply because they remain undiagnosed (Razavi et al., 2020). A meta-analysis has revealed that, in primary care, only 50% of patients with depression are diagnosed, and only 15% obtain treatment (Mitchell et al., 2009). Some factors can contribute to explaining this situation. Stigmatization is one of them and plays an important role (Gaum et al., 2019). Also, somatic symptoms, such as fatigue, sleeping problems, headache, and backache, usually accompany depressive episodes, which can confound the diagnosis (Mboya et al., 2020). On the other hand, the patient must remember the frequency of symptoms in the past two weeks.

In this context, tools that can help in diagnostic screening are of great value, as they can contribute to reducing the number of undiagnosed patients. Besides, Artificial Intelligence (AI) has enabled several successful applications in medicine, particularly in psychiatry. In essence, AI is the combination of sophisticated mathematical models and computation, which results in the development of sophisticated algorithms capable of emulating human intelligence (Souza Filho et al.,

* Corresponding author. Universidade Federal Rural do Rio de Janeiro, Rio de Janeiro, Brazil.
E-mail address: mederitomarques@gmail.com (E.M. Souza Filho).

https://doi.org/10.1016/j.jpsychires.2020.09.025
Received 24 March 2020; Received in revised form 7 August 2020; Accepted 25 September 2020
Available online 30 September 2020

2019). A systematic review, for example, evaluated the use of Machine Learning (ML), an AI subset, in trauma-related disorders, such as Acute Stress Disorder and Posttraumatic Stress Disorder. The clinical and etiological heterogeneity of these diseases favors the use of these tools and brings challenges inherent to its use in clinical practice for the benefit of patients (Ramos-Lima et al., 2020). In another study, an ML algorithm was used to investigate whether transferred entropy, which represents the information flow extracted from electroencephalography in resting-state, could be a predictor of electroconvulsive therapy response in patients who suffer from disorder schizoaffective or schizophrenia. The authors concluded that patients with higher effective connectivity in frontal areas might have a better answer to electroconvulsive therapy (Min et al., 2019). This study aims to evaluate the performance of ML algorithms in the detection of depressive patients from the clinical, laboratory, and sociodemographic data obtained from the National Network for Research on Cardiovascular Diseases from June 2016 to July 2018. The idea is to use this tool to make a screening of patients, which can contribute to reducing the number of undiagnosed cases.

2. Methods

We analyzed the data of the Brazilian National Cardiovascular Disease Research Network during the period from June of 2016 to July of 2018. It was a randomized clinical trial 1:1 using clusters of fixed size. The study was open for intervention and blinded for evaluation designed to analyze the impact of the implantation of a telecardiology system in reducing referrals from primary care patients to cardiologists comparing units that received this tool with others without such technological incorporation. The study protocol was approved and monitored by Instituto Nacional de Cardiologia in Brazil. All patients signed informed written consent.

2.1. Randomization and sample size

A computer program developed from the "R" statistical software to perform the randomization. The blocks of randomization and consequent allocation of units were established according to the chronological entry in the study (R Core Team, 2018). The sample size calculation was structured based on a randomized cluster trial. It required 10-fixed-sized clusters of 50 patients in each group for an absolute reduction in referral to a specialist of 30% considering the power of 80% and alpha of 0.05 with 90 days follow-up.

2.2. Allocation units

The allocation units were 20 primary care units in the city of Rio de Janeiro, in Brazil. The criteria for inclusion and choice of units were to have a Medical Residency Program: all of them with medical preceptorship and operational flows, assuming that there is the same standard medical knowledge and ensuring homogeneity between the groups since the units were subjected to randomization (and not the patients). Therefore, both groups can be considered standardized in training and quality care.

2.3. Inclusion and exclusion criteria

Inclusion criteria were patients seen in primary care units with a referral for electrocardiogram (EKG) (for any indication), over 18 years old, and who agreed to sign the Informed Consent Form. We excluded people with no cognitive ability or literacy to understand the questionnaires and also patients who did not provide redundant contact for follow-up because they were identified as more challenging to follow.

2.4. Experimental and control group

Standard 12-lead EKGs were performed by the local primary care professional, using digital electrocardiographs by Tecnologia Eletrônica Brasileira model ECGPC (São Paulo, Brazil). Exams were performed in rest, with registration for 10 s, at a rate of 25 mm/s. Specific software was developed in-house, capable of capturing an EKG tracing for immediate upload and the patient's self-declared clinical history, to the TNMG analysis center via the internet. The clinical information, EKGs tracings, and reports were stored in a customized database. EKGs were interpreted by a team of trained cardiologists using standardized criteria to generate an EKG report, which was done as free text. The experimental group had EKG performed by a member of the research team in a particular device for the study. The examination was carried out using a system called Sigdiagnosis, developed by Universidade Federal de Minas Gerais (Marino et al., 2016). A specialist by remote access analyzed the EKG. Through this system, it was possible to request a teleconsultation with a cardiologist. The electrocardiographic exams were delivered to the respective unit with a report accomplished by a cardiologist within a maximum period of 24 h. In the control group, the EKG was performed by a local technician regularly on devices from the unit itself that does not issue a report. The trace of EKG without analysis was delivered immediately to the patient following the standard flow of the primary unit care.

2.5. Database

The patients underwent clinical evaluation documented in the clinical research form (CRF) and registered in a database developed by the Instituto Nacional de Cardiologia, being followed up by telephone contact and review of medical records. All data collected were included a posteriori by two blinded and independent researchers in an electronic CRF. Study data were stored and managed using Research Electronic Data Capture (REDCap) hosted at Instituto Nacional de Cardiologia (Harris et al., 2009). REDCap is a secure, web-based software platform designed to support data capture for research studies, providing: (a) an intuitive interface for validated data capture; (b) auditing trails for tracking data manipulation and exporting procedures; (c) automatic export procedures for seamless data downloads to standard statistical packages; and (d) methods for data integration and interoperability with external resources (Harris et al., 2019). A second blind and independent researcher adjudicated the data, check with the patient's medical records.

2.6. Features

Our study used only clinical-laboratory and sociodemographic data obtained during the patient's follow-up. Table 1 shows the considered information from 971 patients (881 non-depressive and 90 with depression). The analysis excluded 29 patients due to missing data. The attributes used as inputs of the algorithms were gender, age, educational level, household income, smoking, alcoholism, illicit drug-using, physical activity, dyslipidemia, hypertension, and diabetes. The categorical variables were transformed using a one-hot-encoding strategy, and the attributes were normalized. Therefore, a matrix with 34 columns and 971 rows was obtained. The last column corresponds to the label indicating whether the patient is not depressed (0) or is depressed (1). The diagnosis of depression was according to the DSM V. All data were anonymized, as suggested in the General Data Protection Regulation (GDPR, 2016).

2.7. ML algorithms

ML models used to perform the classification were Logistic Regression (LR), K-Nearest-Neighbors (KNN), Classification and Regression Tree (CART), AdaBoost (AB), Gradient Boosting (GB), Extreme Gradient

Table 1
List of attributes.

Categorical patients characteristic	Categories	Meaning
Gender	0	Female
	1	Male
Educational Level	1	Illiterate
	2	Complete elementary school
	3	Incomplete elementary school
	4	Complete high school
	5	Incomplete high school
	6	Complete higher school
	7	Incomplete higher school
Household income	1	Less than or equal to 1 minimum wage
	2	Greater than 1 and less than or equal to 2 minimum wages
	3	Greater than 2 and less than or equal to 5 minimum wages
	4	Greater than 5 and less than or equal to 10 minimum wages
	5	Greater than 10 minimum wages
Smoking	1	Current smoker
	2	Ex-smoker
	3	No
Alcoholism	1	Yes
	2	No
Illicit drug using	1	Yes
	2	No
Physical activity	1	Light
	2	Moderate
	3	Intense
	4	Yes[a]
	5	No
Dyslipidemia	1	Yes
	2	No
Hypertension	1	Yes
	2	No
Diabetes	1	Yes
	2	No

[a] Unknown frequency.

Boosting (XGB), Random Forests (RF) and Support Vector Machine (SVM) (Verhulst, 1845; Fix and Hodges, 1951; Breiman, 2001; Friedman, 2002; Ho, 1995; Chen and Guestrin, 2016; Cortes and Vapnik, 1995).

2.8. Cross-validation

K-fold cross-validation is a useful technique to obtain a robust estimate of the generalizability capability of an ML model. First of all, the database is divided into k equally sized parts (folds). After that, k - 1 folds are used for training of the ML models, and the remaining part is employed as a validation set. The process is repeated until it is ensured that all parts integrate the validation set just once (James et al., 2015; Stone, 1974). Therefore, all patients in the database are used for training the models and appear in the test set only once. Our study used 10-fold cross-validation to analyze the generalizability and the mean of the area under the curve receiver operating characteristic (AUROC) as a performance metric. Sanderson et al. used this technique in their ML study, which showed promising results in predicting death by suicide using administrative health care system data (Sanderson et al., 2020).

2.9. Synthetic Minority Oversampling Technique (SMOTE)

The database used in this study is unbalanced. It means that there is an imbalance between the number of patients diagnosed with or without depression. This technique can contribute to the poor performance of ML models and, therefore, to reduce the imbalance, it was necessary to employ the Synthetic Minority Oversampling Technique (SMOTE). The idea is to generate synthetic examples to over-sampling the minority class. The process of obtaining the new synthetic samples considers the neighborhood relations between the elements of this group (k minority class nearest neighbors). New data is produced by interpolation among several minority class instances that are within a defined neighborhood without any change in the majority class (Chawla et al., 2002; Fernández et al., 2018). After performing this technique, 107 new samples (minority class) were generated, and the new dataset composed of 1078 samples. SMOTE was used successfully by Rahman et al. to increase the proportion of autism spectrum disorder (ASD) cases fivefold in a study that evaluated some ML models and their performance in predicting ASD early in life (Rahman et al., 2020). The code was implemented in the Python 3 programming language (van Rossum, 1993; Pedregosa et al., 2011).

3. Results

From Tables 1 and 2, we can see that 64% of the patients are men, 35% completed elementary school, 41% receive less than one minimum wage, 57% are non-smokers, 37% are alcoholic, 4% are users of illicit drugs, 28% are hypertensive, and 21% are diabetic. The average age is 57.67 (±14.47). Table 3 shows that four ML models had an average area under the receiver operator characteristic curve (AUROC) greater than equal to 0.70. RF and KNN have obtained average AUROC equal to 0.87 and 0.81, respectively. AdaBoost had the worst performance (average AUROC = 0.58). The computational times spent by each model during training were less than 2 s. From data variability in average AUROC, it is essential to note that the smallest standard deviations were observed in the KNN, RF, and XGB models (0.07, 0.08, and 0.11, respectively). At the same time, the highest values were found in LR (0.19), AB (0.16), and CART (0.14). The sensitivity (recall) and accuracy were greater than 0.8 on all models, except for AdaBoost. RF had a recall of 0.9 and an accuracy of 0.89. RF, KNN, and XGB had F1-measure greater than or equal to 0.85.

4. Discussion

The results showed that RF achieved an excellent performance. The sensitivity (recall) and precision were 0.90 and 0.88, respectively. Thus, it is possible to use this model as a useful decision-making support tool. It can point out possible patients with depression and contribute to reducing the number of undiagnosed cases. In this context, ML tools seem to have enormous potential for application in the field of mental health illnesses. Another successful example employed an ensemble ML to predict adult-onset internalizing disorders, namely, generalized anxiety disorder, panic disorder, social phobia, depression, and mania). The AUROC of super learner ensembles ranged from 0.76 (depression) to 0.83 (mania) (Rosellini et al., 2020). The determination of the best models took into consideration not only the AUROC value but also the standard deviation perceived. As a result, RF emerged as the best model, followed by XGB. The performance of the RF was also observed in a study in which it was used to foretell future mental healthcare consumption in patients with non-affective psychosis. AUROC was 0.71 (Kwakernaak et al., 2020). On the other hand, XGB had the best accuracy (79%) in predicting Korean adolescents of high-risk suicide (Jung et al., 2019). On the other hand, when comparing the performance of the RF with other models of ML used in psychiatry, we can realize that the results obtained in the scope of this work indicate that the algorithm performed well.

Nevertheless, it is important mentioning that the performance of a model may vary depending on different problems. SVM (our second worst model) was the best in a study in which it was used to verify if an outcome of escitalopram treatment can be foretold from electroencephalographic data on patients who had completed eight weeks of

Table 2
Overview of patient characteristics and frequency.

Categorical patients characteristic	Categories	N	%	–	–		Scale
Gender	0	345	0.36				Binary
	1	626	0.64				
Educational Level	1	37	0.04				Categorical
	2	343	0.35				
	3	178	0.18				
	4	103	0.10				
	5	228	0.24				
	6	45	0.05				
	7	37	0.04				
Household income	1	399	0.41				Categorical
	2	357	0.37				
	3	201	0.21				
	4	13	0.01				
	5	1	0.00[a]				
Smoking	1	135	0.14				Categorical
	2	281	0.29				
	3	555	0.57				
Alcoholism	1	356	0.37				Categorical
	2	615	0.63				
Illicit drug using	1	37	0.04				Categorical
	2	934	0.96				
Physical activity	1	174	0.18				Categorical
	2	107	0.11				
	3	19	0.01				
	4	37	0.03				
	5	634	0.65				
Dyslipidemia	1	212	0.22				Categorical
	2	759	0.78				
Hypertension	1	702	0.72				Categorical
	2	269	0.28				
Diabetes	1	202	0.21				Categorical
	2	769	0.79				
Numerical patients characteristic	**Mean**		**SD**	**Median**	**IQR**	**Range**	**Scale**
Age	57.67		14.47	59.00	20.00	75.00	Years

SD: Standard deviation; IQR: Inter-quartile range.
[a] The value without rounding is 0.00103.

treatment for depression. In this study, SVM achieved high accuracy (82.4%), specificity (79.2%), and sensitivity (85.5%) (Zhdanov et al., 2020). In addition to the best performance achieved by RF, KNN and XGB, it is worth mentioning that this study only made use of data that is very common in the physician's routine and which are usually easily obtained from electronic health records.

The anamnesis and physical examination provide most of the required information. Only two attributes (dyslipidemia and diabetes) require diagnosis through blood tests. This scenario favors the use of the models developed here as well as the small number of attributes needed to run the algorithms. In this context, it is essential to highlight a work by Kuang et al. that used Bayesian networks to assess the predictive capacity of heart rate variability in the diagnosis of depression. The results obtained were 86.4% accuracy, 89.5% sensitivity, and 84.2% specificity (Kuang et al., 2017). Therefore, attributes of different natures can be used to provide the diagnosis of depression using ML models successfully. This situation receives a contribution from the systemic nature of the disease since the existence of any mental stress can modify the central and peripheral nervous systems physiology and biochemistry. As a result, it makes depression a psychological disorder that affects the body as a whole (Noyan, 2015), not to mention the relevant contribution of environmental factors. Niedhammer and cols, for instance, protruded the role of psychosocial work factors in depression (Niedhammer et al., 2020).

It is important to emphasize that tools such as the one developed in this study must be inserted in a context of support to the decision-making process and do not propose to make any professional substitution. Instead, what is proposed is to redesign of the modus operandi of the health work process, as well as to expand medical skills (Souza Filho et al., 2019). Nevertheless, as pointed out by Schwenk, we must keep the focus on the needs of patients and the protection of the sacred covenant between doctors and patients (Schwenk, 2020).

On the other hand, we emphasize that the tool is most useful in carrying out screening patients in primary care. Depression's diagnosis is prerogative of the doctor, and the tool should not be used, in our reading, for this purpose. It does not prevent the existence of false positives/negatives since they are probabilistic mathematical-computational models. Therefore, improving the ML model's performance is an objective to be pursued due to the associated economic burden. A way to achieve this goal is to carry out a continuous improvement process: it is essential to acquire new data collected and processed correctly - which will be used as input for training and testing the models (Souza Filho et al., 2019). It is also essential to assess the external generalization of the model in other contexts, also considering multicenter data.

Also, the results obtained by the model developed here do not allow quantifying the magnitude of the importance of each variable in the classification process. Besides, Engel et al. point out that dimensions of well-being and quality of life can be affected by depression even though in different magnitudes. In this work's scope, only the educational background and income were used, both of which are related to the quality of life and well-being (Engel et al., 2018). On the other hand, Puterman et al. developed some ML models aiming to predict the mortality of a cohort containing 13,611 adults with ages ranging from 52 to 104 years. A total of 57 factors with different natures (economic, behavioral, social, and psychological) were evaluated. They showed that

Table 3
Computational results (10-fold cross validation).

Model	AUROC	SD	Time (s)	Model	Precision	SD	Time (s)
LR	0.61	0.19	0.38	LR	0.72	0.23	0.37
CART	0.65	0.14	0.09	CART	0.84	0.04	0.09
AB	0.58	0.16	0.84	AB	0.78	0.09	0.91
GB	0.74	0.12	1.44	GB	0.82	0.05	1.52
XGB	0.77	0.11	2.12	XGB	0.86	0.02	1.90
RF	0.87	0.08	1.51	RF	0.88	0.04	1.51
KNN	0.81	0.07	0.10	KNN	0.89	0.03	0.13
SVM	0.61	0.18	0.29	SVM	0.68	0.24	0.61
Model	**Recall**	**SD**	**Time (s)**	**Model**	**F1**	**SD**	**Time (s)**
LR	0.80	0.19	0.41	LR	0.74	0.23	0.37
CART	0.81	0.04	0.09	CART	0.82	0.04	0.09
AB	0.78	0.18	0.86	AB	0.73	0.21	0.87
GB	0.82	0.12	1.47	GB	0.79	0.12	1.55
XGB	0.86	0.05	1.85	XGB	0.85	0.04	1.90
RF	0.90	0.03	1.46	RF	0.89	0.03	1.52
KNN	0.83	0.04	0.11	KNN	0.85	0.04	0.12
SVM	0.80	0.19	0.31	SVM	0.73	0.23	0.30

Model	Accuracy	SD	Time (s)
LR	0.80	0.19	0.41
CART	0.80	0.05	0.07
AB	0.78	0.18	0.88
GB	0.82	0.12	1.50
XGB	0.86	0.05	1.90
RF	0.89	0.03	1.76
KNN	0.83	0.04	0.13
SVM	0.80	0.19	0.38

Legend: LR: Logistic Regression; CART: Classification and Regression Tree; AB: Adaptive Boosting; GB: Gradient Boosting; XGB: Extreme Gradient Boosting; RF: Random Forests; KNN: K-Nearest-Neighbors; SVM: Support Vector Machine; SD: Standard Deviation; F1: F1-measure; AUROC: Area Under the Receiver Operating Characteristics.

in addition to traditional risk factors, such as physical inactivity, smoking, and alcohol consumption, other variables also played an important role in mortality, such as recent financial difficulties, history of unemployment, childhood adversities, and affective negativity (Puterman et al., 2020). Thus, we believe that ML models can, many times, bring a broader view under specific issues. However, it is essential to have data representative of the phenomenon to be studied. In this context, in the scope of the work developed here, related variables could also serve as inputs for ML models, bring improvements in the results obtained and, consequently, increase the performance of the model and bring some valuable insight into the decision-making process.

4.1. Limitations

For limitations, it is crucial to put the results obtained herein perspective. The computational execution times spent by ML models favor its use; however, the need to obtain information such as educational level and household income brings with it some challenges. These are variables that the values can vary a lot depending on the region or country considered. The World Population Review showed that Brazil, for instance, has a median household income of $ 7.522, which corresponds to 30%, 17%, and 24% of the median household income in Finland, United States of America, and France respectively (Word Population Review, 2020). Besides, not all patients feel comfortable providing personal information, such as income, and may even lie about their values. This behavior is influenced by both non-economic and economic aspects (Cappelen et al., 2013). In contrast, it is important to note that the information used in ML models was acquired retrospectively at 20 primary care units. All of these units are inserted in the same geographic region (Rio de Janeiro), which can be a limiting factor regarding the generalization capacity for other regions with a different profile. Another point is that we did not include other variables in the analysis (for example, related to mental elements). This inclusion can be quite intriguing and increase the performance of the models. Thus, we understand that future work in this direction can bring additional clarifications on the subject. In addition, we also believe that future research ought to be conducted to assess a possible excess of psychiatric consultations and the rate of recovery from undetected cases.

5. Conclusions

Our findings underlined insight that some ML models can be useful in detecting depressed patients from sociodemographic, clinical, and laboratory data common in clinical practice. RF, KNN and XGB were the algorithms that had the best performances. From a treatment perspective, it represents a new tool for screening, which can assist in reducing the number of undiagnosed cases of this disease as well as facilitate as early treatment onset.

Contributors

Conceptualization of study: EMSF, DMAC, ALPR, and HCVR. Data acquisition and preparation: EMSF, RMF, HCVR, DMAC, ALPR. Statistical analysis: EMSF, JLMA, LNDC. Interpretation of data: EMSF, JLMA, HCVR, RMF, DMAC, ALPR. Drafting the manuscript: EMSF, JLMA, and HCVR. Revising the manuscript for important intellectual content: EMSF, JLMA, HCVR, LNDC. All authors have approved the final manuscript.

Role of funding source

The authors would like to acknowledge the support from Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Finance Code 001.

Declaration of competing interest

The authors declare they have no conflict of interest.

References

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32.
Cappelen, A.W., Sorensen, E.O., Tungodden, B., 2013. When do we lie? J. Econ. Behav. Organ. 93, 258–265.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over–sampling technique. J. Artif. Intell. Res. 16, 321–357.
Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.
Dsm-V, 2013. Diagnostic and Statistical Manual of Mental D. isorders, fifth ed. American Psychiatric Association, Washington, DC.
Engel, L., Chen, G., Richardson, J., Mihalopoulos, C., 2018. The impact of depression on health-related quality of life and wellbeing: identifying important dimensions and assessing their inclusion in multi-attribute utility instruments. Qual. Life Res. 27 (11), 2873–2884. https://doi.org/10.1007/s11136-018-1936-y.
Fernández, A., García, S., Herrera, F., Chawla, N.V., 2018. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J. Artif. Intell. Res. 61, 863–905.
Fix, E., Hodges, J.L., 1951. Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties. Technical Report 4. USAF School of Aviation Medicine, Randolph Field, Texas.
Friedman, J.H., 2002. Stochastic gradient boosting. Comput. Stat. Data Anal. 38 (4), 367–378.
Gaum, P.M., Brey, F., Kraus, T., Lang, J., 2019 Nov 5. Does stigmatization moderate the association between intention and implementation of learned prevention-strategies at work after a depressive episode? - a cross-sectional pilot study. J. Occup. Med. Toxicol. 14, 26.
Global Burden of Disease Study GBD, 2018. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 392, 1789–1858.
GDPR, 2016. Regulation (eu) 2016/679 of the European Parliament and of the Council of 27 April 2016:on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/

46/EC (General Data Protection Regulation). Official Journal of the European Union L 119.
Harris, P.A., Taylor, R., Thielke, R., et al., 2009 Apr. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J. Biomed. Inf. 42 (2), 377–381.
Harris, P.A., Taylor, R., Minor, B.L., et al., 2019 Jul. The REDCap consortium: building an international community of software platform partners. J. Biomed. Inf. 95, 103208.
Ho, T.K., 1995. Random decision forests. In: Proceedings of the 3rd International Conference on Document Analysis and Recognition; 1995 Aug 14-16; Montreal. IEEE Computer Society, Washington, DC, pp. 278–282.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2015. An Introduction to Statistical Learning, with Applications in R, 6th Printing. Springer, New York.
Jung, J.S., Park, S.J., Kim, E.Y., et al., 2019 Jun 6. Prediction models for high risk of suicide in Korean adolescents using machine learning techniques. PloS One 14 (6), e0217639.
Kessler, R.C., Berglund, P., Demler, O., et al., 2003. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). J. Am. Med. Assoc. 289, 3095–3105.
Kuang, D., Yang, R., Chen, X., et al., 2017 Dec. Depression recognition according to heart rate variability using Bayesian Networks. J. Psychiatr. Res. 95, 282–287.
Kwakernaak, S., van Mens, K., Group Investigators, Cahn, W., Janssen, R., 2020 Mar 5. Using machine learning to predict mental healthcare consumption in non-affective psychosis. pii: S0920-9964 Schizophr. Res. (20). https://doi.org/10.1016/j.schres.2020.01.008, 30023-2, ([Epub ahead of print]).
Marino, B.C.A., Ribeiro, A.L.P., Alkmim, M.B., et al., 2016 Jul 1. Coordinated regional care of myocardial infarction in a rural area in Brazil: Minas Telecardio Project 2. Eur Heart J Qual Care Clin Outcomes 2 (3), 215–224.
Mboya, I.B., John, B., Kibopile, E.S., et al., 2020 Jan 29. Factors associated with mental distress among undergraduate students in northern Tanzania. BMC Psychiatr. 20 (1), 28.
Min, B., Kim, M., Lee, J., Byun, J.I., Chu, K., Jung, K.Y., Lee, S.K., Kwon, J.S., 2019 Dec 26. Prediction of individual responses to electroconvulsive therapy in patients with schizophrenia: machine learning analysis of resting-state electroencephalography. pii: S0920-9964 Schizophr. Res. (19). https://doi.org/10.1016/j.schres.2019.12.012, 30582-1, ([Epub ahead of print]).
Mitchell, A., Vaze, A., Rao, S., 2009. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 374 (9690), 609–619.
Niedhammer, I., Coindre, K., Memmi, S., et al., 2020 Jan 22. Working conditions and depression in the French national working population: results from the SUMER study. J. Psychiatr. Res. 123, 178–186.
Noyan, M.A., 2015 Sep. Depression is a systemic disease. Turk Kardiyol. Dernegi Arsivi 43 (6), 503–504.
Park, L.T., Zarate Jr., C.A., 2019. Depression in the primary care setting. N. Engl. J. Med. 380, 6.
Pedregosa, F., Varoquaux, G., Gramfort, A., et al., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Puterman, E., Weiss, J., Hives, B.A., et al., 2020. Predicting mortality from 57 economic, behavioral, social, and psychological factors. Proc. Natl. Acad. Sci. Unit. States Am. https://doi.org/10.1073/pnas.1918455117, 201918455.
R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available online at. https://www.R-project.org/.
Rahman, R., Kodesh, A., Levine, S.Z., Sandin, S., Reichenberg, A., Schlessinger, A., 2020. Identification of newborns at risk for autism using electronic medical records and machine learning. Eur. Psychiatr. 63 (1), 1–7 e22.
Ramos-Lima, L.F., Waikamp, V., Antonelli-Salgado, T., Passos, I.C., Freitas, L.H.M., 2020. The use of machine learning techniques in trauma-related disorders: a systematic review. J. Psychiatr. Res. 121, 159–172.
Razavi, R., Gharipour, A., Gharipour, M., 2020 Jan 24. Depression screening using mobile phone usage metadata: a machine learning approach. J. Am. Med. Inf. Assoc. https://doi.org/10.1093/jamia/ocz221 pii: ocz221, ([Epub ahead of print]).
Rosellini, A.J., Liu, S., Anderson, G.N., et al., 2020 Feb. Developing algorithms to predict adult onset internalizing disorders: an ensemble learning approach. J. Psychiatr. Res. 121, 189–196.
Sanderson, M., Bulloch, A.B.M., Wang, J., et al., 2020 Mar 1. Predicting death by suicide using administrative health care system data: can recurrent neural network, one-dimensional convolutional neural network, and gradient boosted trees models improve prediction performance? J. Affect. Disord. 264, 107–114.
Schwenk, T.L., 2020 Feb 24. What does it mean to Be a physician? J. Am. Med. Assoc. https://doi.org/10.1001/jama.2020.0146 ([Epub ahead of print]).
Souza Filho, E.M., Fernandes, F.A.F., Soares, C.L.A., et al., 2019 Jan. Artificial intelligence in cardiology: concepts, tools and challenges - "the horse is the one who runs, you must Be the jockey". Arq. Bras. Cardiol. 114 (1), 194. https://doi.org/10.36660/abc.20180431. S0066-782X2019005022109.
Stone, M., 1974. Cross-validatory choice and assessment of statistical predictions. J. Roy. Stat. Soc. B 36 (2), 111–133.
Tanner, J.A., Hensel, J., Davies, P.E., et al., 2019 Dec 13. Economic burden of depression and associated resource use in manitoba, Canada. Can. J. Psychiatr. https://doi.org/10.1177/0706743719895342, 706743719895342, ([Epub ahead of print]).
van Rossum, G., 1993. An introduction to Python for UNIX/C programmers. In: Proc. Of the NLUUG Najaarsconferentie. (Dutch UNIX Users Group).
Verhulst, P., 1845. Recherches mathématiques sur la loi d'accroissement de la population. [Mathematical Researches into the Law of Population Growth Increase]. Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles 18, 4–54.
Word Population Review, 2020-02-17. Median Income by Country Population. Retrieved 2020-02-24, from. http://worldpopulationreview.com/countries/median-income-by-country/.
Zhdanov, A., Atluri, S., Wong, W., et al., 2020 Jan 3. Use of machine learning for predicting escitalopram treatment outcome from electroencephalography recordings in adult patients with depression. JAMANetw Open 3 (1), e1918377.