



Review of Machine Learning Algorithms for Diagnosing Mental Illness

Gyeongcheol Cho¹, Jinyeong Yim², Younyoung Choi³ ✉, Jungmin Ko⁴, and Seoung-Hwan Lee⁵

¹Department of Psychology, McGill University, Montreal, Quebec, Canada

²Georgia Institute of Technology, North Avenue, Atlanta, USA

³Department of Adolescent Psychology, Hanyang Cyber University, Seoul, Republic of Korea

⁴Department of Mathematics Education, Sungkyunkwan University, Seoul, Republic of Korea

⁵Department of Psychiatry, Inje University Ilsan Paik Hospital, Goyang, Republic of Korea

Objective Enhanced technology in computer and internet has driven scale and quality of data to be improved in various areas including healthcare sectors. Machine Learning (ML) has played a pivotal role in efficiently analyzing those big data, but a general misunderstanding of ML algorithms still exists in applying them (e.g., ML techniques can settle a problem of small sample size, or deep learning is the ML algorithm). This paper reviewed the research of diagnosing mental illness using ML algorithm and suggests how ML techniques can be employed and worked in practice.

Methods Researches about mental illness diagnostic using ML techniques were carefully reviewed. Five traditional ML algorithms-Support Vector Machines (SVM), Gradient Boosting Machine (GBM), Random Forest, Naïve Bayes, and K-Nearest Neighborhood (KNN)-frequently used for mental health area researches were systematically organized and summarized.

Results Based on literature review, it turned out that Support Vector Machines (SVM), Gradient Boosting Machine (GBM), Random Forest, Naïve Bayes, and K-Nearest Neighborhood (KNN) were frequently employed in mental health area, but many researchers did not clarify the reason for using their ML algorithm though every ML algorithm has its own advantages. In addition, there were several studies to apply ML algorithms without fully understanding the data characteristics.

Conclusion Researchers using ML algorithms should be aware of the properties of their ML algorithms and the limitation of the results they obtained under restricted data conditions. This paper provides useful information of the properties and limitation of each ML algorithm in the practice of mental health.

Psychiatry Investig 2019;16(4):262-269

Key Words Machine learning, Mental illness, Big data.

INTRODUCTION

The exponential expansion of data has emerged in every field from the rapid advancement of computer and internet technology such as a considerable number of students' data from OECD in educational area and customer-related data from Walmart and Facebook in industrial side.¹⁻⁷ Such a huge growth of data also has occurred in the health domain as well. Various personal health information on patients and popula-

tion started to be digitized in the electronic health record (EHR) system in many countries. In 2012, the EHR produced around 500 petabytes of data, and its size is expected to be 25,000 petabytes by 2020.^{5,8} Moreover, the growth is not restricted to its quantity. Developments of medical technology enable measuring various forms of human biology such as a gene, cerebral blood flow, and EEG, at a relatively lower cost but with higher accuracy than before.⁹⁻¹¹ This big quality data has a high potential to advance healthcare sector by deepening our understanding of human disease mechanisms but requires a different approach from traditional one for their efficient analysis.

Machine Learning (ML) has been acknowledged as an appropriate method for the analysis of big data. ML, conceptually suggested by Allan Turing and coined by Arthur Samuel in 1950s,^{12,13} has been in widespread use in various field including medical area since 1990s, due to incessant efforts of ML researchers along with expeditious developments of com-

Received: September 16, 2018 Revised: November 14, 2018

Accepted: December 21, 2018

✉ Correspondence: Younyoung Choi, PhD

Department of Adolescent Psychology, Hanyang Cyber University, 220 Wangsimni-ro, Seongdong-gu, Seoul 04763, Republic of Korea

Tel: +82-2-2290-0367, Fax: +82-2-2290-0600, E-mail: 1120008@hycu.ac.kr

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

puting power. ML techniques have merits in handling big data in that scalable ML algorithm to process large scale of data have been already studied and devised by ML researchers even before the importance of big data emerged.¹⁴ In addition, some ML techniques enable their machine to learn properly from the data of numerous variables compared to a small number of cases.^{15,16} Thus, ML can be regarded as ‘an essential part of big data analytics,’^{14,17} and has contributed to resolving issues in healthcare such as early diagnosis of disease, real time patient monitoring, patient centric care, and enhancement of treatment.¹⁸

However, as only successes of ML such as AlphaGo beating a human Go champion has been highlighted in press,¹⁹ ML started to seem like a magical wand to some people outside of ML fields. In our recent mental health research, a clinician expected that ML algorithms could be a remedy for small sample size or gave a proper diagnosis without expertise’s diagnosis even for training. In addition, Deep Learning or Neural Networks algorithm, well-known as the main algorithm of Artificial Intelligence (AI), was thought of as all about ML. These sorts of confusion obstacles clinical researches using ML techniques which are usually conducted collaboratively by domain experts and ML researchers.

Despite their great advantages, it is obvious that ML techniques are ‘not a panacea that would automatically’ yield a solution of generalizability or higher accuracy without large quality dataset nor any human instruction for training.^{20,21} There are huge numbers of ML algorithms used for the analysis of clinical data other than deep learning, and each of them has its own advantages. Therefore, it is meaningful to organize and offer the prior knowledge about the ML algorithms necessary for applied clinical research with empirical examples of how ML has actually been employed in the clinical field. This information can be a useful guideline to facilitate communication between clinical and ML researchers and help their collaborative research more efficient. Section 2 introduces the basic concept of ML and traditional ML algorithms of supervised learning for clinical research. In Section 3, we focus on several examples of mental health research using ML algorithm and investigate ML algorithms used for those researches and their performance. In Section 4, their implications are summarized, and the points to be noted using those algorithms are discussed.

TYPES OF MACHINE LEARNING TECHNIQUES

Primary purposes of ML techniques are to analyze data, to predict target features of data, or to derive meanings of the given data. Here, we introduce two main types of ML, super-

vised learning and unsupervised learning, in terms of given data and the purpose of analysis. The ML algorithms used in previous works for mental health data are mostly categorized into these two types. We also introduce representative algorithms of each type with examples.

Besides those two main types mentioned above, another main type of ML is reinforcement learning (RL). We do not cover RL in this paper because the mental health data have not been formulated into RL settings in general. The main purpose of RL is for agents to learn optimal behaviors about given environments through repetitive simulations of interacting with environments. This is not applicable to our case where given data is a set of attributes and the corresponding values.

SUPERVISED LEARNING

The supervised learning (SL) is the most commonly used way in ML-based diagnosis. In the supervised learning setting, given data should be labeled. In other words, all data instances should be represented with attributes and corresponding values. Attributes are a set of features, representing data instances. For example, personal characteristics such as height, body weights, eye color, etc. can be attributes to describe a person, or a data instance. A label is the value of a data instance’s specific target attribute that we want to predict from other attributes’ information. The main purpose of supervised ML model is to predict labels, or values, of unseen instances with their corresponding given attributes. The SL model can be viewed as a mapping from attributes to a label. A label of a data instance can be either a discrete value, or a class, or a representing number in a continuous space. If labels are classes, the ML task is a classification problem. On the other hand, if labels are numbers in a continuous space, the task is a regression problem.

An example of a SL setting is described: in case of classifying diseases of patients, each patient is a data instance, a patient’s measured conditions are attributes, and the patient’s disease is a class. The ML algorithm employs attributes to predict a disease class. Given the measured data of a number of patients, the ML model can be built and then can be used to pre-screening of patient’s disease before medical professionals’ diagnosis. To be specific, the SL algorithms require subjects with measured values for attributes and labels that represent their diagnosis. The real diagnosed labels supervise the ML model to be trained. Once a model completed its training, it can predict diagnosis labels of new patients with the measured attributes.

UNSUPERVISED LEARNING

The unsupervised learning requires no supervision, unlike the supervised learning that requires target labels of data instances for the model to predict. The main purpose of unsupervised learning is not to predict target attributes but to handle data without supervision. Some examples of unsupervised learning are figuring out similarities between data instances, discovering relationships of attributes, reshaping attributes to reduce dimensionality, etc. The unsupervised learning methods have not been frequently used in the clinical field since a majority of clinical research using ML aimed to develop a tool to diagnose diseases whose diagnostic criteria were already determined as golden standards. Based on the literature review in our study, it was hard to find a study directly using unsupervised learning method for the diagnostic purpose.

The unsupervised learning, however, can be utilized for the additional analysis along with applying supervised learning techniques. We can bring the example described in the previous section for supervised learning as an example of unsupervised learning. Although disease types are labels for supervised learning, the label information can be neglected, and the extra analysis can be conducted with the unsupervised learning. In case of grouping data instances with a smaller number of groups than the number of diseases, from given data, generated groups can be viewed as an indicator to check similarities of diseases. By contrast, in case of grouping with a greater number, even a disease can be subdivided for the further analysis.

SUPERVISED LEARNING EXAMPLES

In this paper, we focus on the SL setting as described in the example for the supervised learning above. The supervised learning fits to the purpose of diagnosis since diagnosis can be viewed as evaluation/prediction of patients with measured quantities. We do not cover details of all supervised learning algorithms due to the large number of existing SL methods. Instead, we briefly describe representative SL algorithms frequently used in many works: Support Vector Machines (SVM), Gradient Boosting Machine (GBM), Random Forest, Naïve Bayes, and K-Nearest Neighborhood (KNN).

The SVM is one of the most famous and utilized supervised learning methods.²² The fundamental concept of SVM is based on the binary classification case. For the binary classification, we want to divide given data points into two classes. Given data points distributed on the feature space, SVM finds a margin that divides feature space most well. The linear space division is the simplest form to divide a space. However, gen-

erally, data points are not distributed to be well linearly-divided. To cover the complicated non-linear cases, the advanced kernel trick and regularizations are used. Also, SVM variants for the multi-class classification and regression exist. SVM is known as well-working in practice. In order to work well, careful selection of kernel and corresponding data pre-processing are required.

GBM is a boosting method, which is an ensemble technique that leverages a set of weak learners to create a strong learner to obtain better performance.²³ Generally, decision trees, which use a set of hierarchical conditions to divide given data by stages, are used as weak learners due to its simplicity. For example, a set of weak learners may not fit well to given data. We add another weak learner that works well on a part of data that the current set of weak learners do not work well. Then, the new set of weak learners including the newly added weak learner perform better on the given data. The boosting method determines the next newly added weak learners based on the data where the current set of weak learners do not perform well. This procedure can be seen as adding weak learners that work well on data points difficult to handle with the current weak learners. For the aggregation of weak learners, the weights for weak learners are determined when they are newly added to the set. In terms of training, the model requires longer time than other methods because it requires to distinguish which data points are hard to cover with current set of weak learners at every step before adding a new weak learner. The model usually works well in practice.

Random Forest is another ensemble technique using decision trees.²⁴ Unlike GBM which is a boosting method, Random Forest is a bagging method that handles weak classifiers in a different way. The basic procedure that the model uses weak learners and adding new weak learners is same. However, when adding a new weak classifier, the bagging method including Random Forest searches for the best feature among a random subset of data, instead of specific data that are difficult to the current set of weak classifiers. In other words, many randomly generated different decision trees are merged into one learner, and this is why it's called Random Forest. Surprisingly, the combination of many randomly sampled data-based decision trees works well in practice because the randomness helps the model avoid to overfit on data. However, its prediction speed can be slow if it contains a large number of trees and the model performance depends on the model parameter that represents how deeply the model has random trees. This also affects the prediction time of the model.

Naïve Bayes is a naïve probabilistic model to find the value that achieves the maximum probability computed from a conditional probability chain.²⁵ This model is called 'naïve' because it assumes independency between all measured at-

tributes. In real world, data attributes are not always perfectly independent. With this assumption, conditional probabilities of attributes can be easily computed as long as all values of attributes are given. In Naïve Bayes model, the model designer is required to decide which attributes to view as dependent attributes to other attributes when computing conditional probabilities. In practice, in some cases, the model works surprisingly well even with this independency assumption, but in other cases, it totally fails to work. The model solely depends on characteristics of data.

KNN is a type of instance-based learning, where no computations are required before actual classification or regression.²⁶ In other words, given data itself can be considered as 'learnt' model and then computations are conducted when actual predictions are needed. The basic idea is that a label of a data point can be predicted from same/similar labels of nearest neighbor data points. Therefore, in order to use this algorithm, the selection of parameter K and attribute-distance computation metric to compute which other data points are nearest neighbors are required. For attribute-distance computation, relationships among attributes should be carefully considered to achieve good performance. However, this algorithm is usually not best because of its simplicity to model real world settings. Also, the computation time rapidly increases as the number of data or the number of attributes increase. In the general sense that prediction time should be short and that training time does not matter as much as prediction time, this disadvantage may be critical.

Here, we introduced five supervised learning models. The summary of these models is shown in Table 1. Generally, the performance of SVM, GBM, and Random Forest are usually

better than the performance of other simple models, Naïve Bayes and KNN.^{24,27-31} But sometimes simple models better depend on the data. The model adaptation always depends on the data characteristics. In most real-world cases, it is hard to predict which model will work the best before actually applying data to models. In some cases, may only a small number of measured attributes contribute to solving given problems even though a large number of attributes were measured. In some other cases, a complex ML model may overfit to given data so that it does not work well for unseen data while simpler ML models work better as they avoid overfitting to training data. Therefore, one of the best ways to figure out which ML model works is to apply given data to all models. Then we can choose the best-working model and analyze the data reversely from the model performance. Otherwise, we need a thorough investigation of data before selecting which ML model to use.

DIAGNOSIS OF HEALTH DISEASES BY USING DIFFERENT MACHINE LEARNING ALGORITHMS

We searched research papers for diagnosis of mental illnesses using ML by article search engine. All literatures were selected in SCOPUS, RISS, and PubMed. Keywords were mental disorders, mental illness, diagnosis, machine learning, and big data. We examined a total of 59 articles, of which 10 papers related to mental illness were analyzed and summarized. We categorized them according to the type of mental illnesses and summarized each of them by their purpose of using ML techniques, which techniques they use, data type, sample

Table 1. The summary of characteristics of 5 machine learning algorithms frequently used in the field of mental health

| Algorithm | Advantage | Disadvantage |
|---------------|--|--|
| SVM | - Working relatively accurate in general in practice | - An appropriate kernel and careful data-preprocessing is required to precisely analyze complicated data |
| GBM | - Working relatively accurate in general in practice | - Training learner in GBM takes relatively more time than in other algorithms. |
| Random forest | - The algorithm itself automatically covers its previous weakness in dealing with some attributes by adding more weak learners | - Difficult to interpret the relationship between attributes and the target |
| Naïve bayes | - Need relatively short time for training and prediction | - Prediction speed can be relatively slow for a complex random forest |
| | - Can provide intermediate conditional probabilities of attributes | - Difficult to interpret the relationship between attributes and the target |
| KNN | - Training process is not required | - Unrealistic assumption: independency of all attributes |
| | | - Prediction time rapidly increases as the number of data or attributes increases |
| | | - Relatively large additional data is required for even a small increase in the number of features to consider |
| | | - Highly dependent on The distance metric to be used |

SVM: Support Vector Machine, GBM: Gradient Boosting Machine, KNN: K-Nearest Neighborhood

size, and the accuracy as their performances. Table 2 depicts the summary of each research.

Post-traumatic stress disorder

Augsburger and Elbert³² performed a ML technique to predict risk-taking behavior for refugees that have experienced

Table 2. The summary of mental health researches using ML techniques

| Disease | Machine learning | Data type | Sample size | Accuracy | Author | Year |
|---------------|---|---|-------------|--|-----------------------|--------------------|
| PTSD | Stochastic GBM | Balloon analogue risk task (BART), questionnaires in semi-structured interviews | 56 | | Augsburger and Elbert | 2017 ³² |
| | SVM | Longitudinal data (evaluations during emergency room admission) | 152 | 64% (using background information alone) 82% (using background and ER variables) 88% (using data through 1 week) 93% (using data through 1 month) | Galatzer-Levy et al. | 2017 ³³ |
| | SVM, Target information equivalence algorithm | Event characteristics, emergency department records, early symptoms | 957 | 75% | Karstoft et al. | 2015 ³⁴ |
| Schizophrenia | SVM | MRI | 198 | 90% (differentiating between schizophrenia patients and healthy controls) 88% (differentiating between schizophrenia patients and bipolar disorder ones) 53%, 67% (differentiating between bipolar disorder patients and healthy controls) | Schnack et al. | 2014 ³⁵ |
| | Random forest | MRI scans (anatomic brain measures), clinical and genetic information | 197 | 73.7% | Greenstein et al. | 2012 ³⁶ |
| | SVM with an ensemble (SVME) method | fMRI, single nucleotide polymorphism (SNP) data | 40 | 74% (applying SVM to fMRI and SNP data) 82% (building SVME with SNP data) 83% (building SVME with fMRI data) 87% (combining above three models) | Yang et al. | 2010 ³⁹ |
| Depression | GBM | Patient-reported data | 1,949 | 64.6% (predicting clinical remission) 59.6% (predicting escitalopram treatment group in cross validation test) 59.7% (predicting combined escitalopram-bupropion treatment group in cross validation test) | Chekroud et al. | 2016 ⁴⁰ |
| ASD | SVM | Single-word utterance data | 81 | 76% | Nakai et al. | 2017 ⁴¹ |
| | Random forest | Words and phrases from the evaluations | 1,162 | 86.5% | Maenner et al. | 2016 ⁴² |
| | KNN, SVM, naïve bayes | Modified multiscale entropy (mMSE) computed from resting state EEG data | 79 | 77% (applying KNN and SVM to 9 months infants data) 80% (applying naïve bayes to 18 months infants data) | Bosl et al. | 2011 ⁴³ |

ASD: Autism Spectrum Disorders, EEG: electroencephalogram, MRI: magnetic resonance imaging, fMRI: functional magnetic resonance imaging, PTSD: post-traumatic stress disorder, SVM: support vector machine, GBM: gradient boosting machine, KNN: K-Nearest Neighbors

traumatic experiences. Data sets consisted of Balloon analogue risk task (BART) and questionnaires in interviews of 56 cases. ML technique used in this experiment was stochastic GBM and R was employed for analysis. This method was effective compared to conventional methods since lots of variables can be tested at the same time, even in relatively few samples. This experiment had the limitation that a small dataset is used to construct a model.

Galatzer-Levy et al.³³ predicted post-traumatic stress using Support Vector Machines (SVM). The data set was longitudinal data of 152 subjects collected during emergency room admission following the traumatic incident. Latent growth mixture modeling (LGMM) was used to identify PTSD symptom severity trajectories, and its result was used for outcome variable in models built by ML method. SVM-based recursive feature elimination was employed for feature selection, and SVM was used for predicting trajectory membership. MATLAB was applied by using SVM. Accuracy reported as the mean area under the receiver operator characteristic curve (AUC) was 0.64 only on the basis of back ground information, 0.82 on the basis of background and ER variables, 0.88 on the basis of data through 1 week, and 0.93 on the basis of data through 1 month.

Karstoft et al.³⁴ used a ML method for the purpose of prediction of PTSD. Data set comprised of 957 trauma survivors and 68 features about even characteristics, emergency department records and early symptoms. A Target Information Equivalence Algorithm was applied in this experiment; An optimal set of variables was selected from a large set of variables, and SVM was used for prediction by MATLAB. Area Under the ROC curve (AUC) was provided as accuracy, and mean AUC was 0.75.

Schizophrenia

Schnack et al.³⁵ employed a ML method to classify patients with schizophrenia, those with bipolar disorder and healthy controls by using magnetic resonance imaging (MRI) scans. MRI scans of 66 schizophrenia patients, 66 patients with bipolar disorder and 66 healthy subjects were used, and SVM was used to construct three models. The average accuracy of the model to discriminate between schizophrenia patients and healthy controls was 90%, and the average accuracy of the model separating patients with schizophrenia and bipolar disorder was 88%. The model to distinguish patients with bipolar disorder and healthy controls offered lower accuracy than other models, correctly classifying 53% of the patients with bipolar disorder and 67% of the healthy subjects.

Greenstein et al.³⁶ performed work to classify childhood onset schizophrenia (COS) groups and healthy controls. The employed data set consisted of 98 patients who suffer from

COS and 99 controls. Data set comprised of 74 anatomic brain magnetic resonance imaging (MRI) sub regions and clinical and genetic information. Random Forest method is applied on the data set since it has lower error rates compared to other methods,³⁷ and can determine the probability of diseases on the basis of the feature set of brain regions.³⁸ 73.7% accuracy was attained after classification.

Yang et al.³⁹ used a hybrid ML method to classify schizophrenia patients and healthy controls. The data used was gathered from 20 schizophrenia patients and 20 controls. SVM algorithm was implemented in this experiment using functional magnetic resonance imaging (fMRI) and single nucleotide polymorphism (SNP) data. The accuracy was 0.74 for the method that SNPs were used to build a SVM ensemble (SVME), and 0.82 for the method that voxels in the fMRI map were used to construct another SVME. Also, the accuracy of the method to attain a SVM classifier using components of fMRI activation obtained with independent component analysis was 0.83, and the accuracy of the way to combine above three models was 0.87.

Depression

Chekroud et al.⁴⁰ developed a ML algorithm to predict clinical remission from a 12-week course of citalopram. Data set consisted of 1949 patients with depression from level 1 of the Sequenced Treatment Alternatives to Relieve Depression. 25 variables were selected from 164 patient-reportable variables to make the most predictive outcome. GBM was implemented for prediction due to its merit that it combines a few weakly predictive models when built. By using GBM, 64.6% accuracy was attained. In cross validation test, the accuracy of 59.6% was obtained in the escitalopram treatment group (n=151) of Combining Medications to Enhance Depression Outcomes (COMED), and accuracy of 59.7% was offered in a combined escitalopram-bupropion treatment group (n=134) in COMED.

Autism spectrum disorders

Nakai et al.⁴¹ performed a ML method to classify children with autism spectrum disorders (ASD) and typical development (TD). Data set was single-word utterance data and consisted of 81 cases, 30 with ASD and 51 with TD. The machine-learning-based voice analysis method was SVM and SVM was applied by using MATLAB. The accuracy offered by SVM was 0.76. According to the result of this project, the ML method was superior to speech therapist judgments if only single-word utterance was used for stimuli.

Maenner et al.⁴² classified autism spectrum disorder (ASD) by ML algorithms. Random Forest was employed for feature selection and classification. Data set used in this project was

words and phrases from the evaluations collected from the 2008 Georgia Autism and Developmental Disabilities Monitoring (ADDM) site, and consisted of 5,396 evaluations of 1,162 children. Data set about 9,811 evaluations of 1,450 children from the 2010 Georgia ADDM surveillance was employed for evaluating the classifier. Random Forest obtained 86.5% accuracy for prediction.

Bosl et al.⁴³ applied ML methods to classify infants into higher risk group at Autism Spectrum Disorder (ASD) or control. Data set was modified multiscale entropy (mMSE) computed based on resting state EEG data and consisted of 79 infants; 46 are at high risk for ASD and 33 are healthy controls. The k-nearest neighbors (k-NN), SVM, Naïve Bayes algorithms were used for classification to obtain the best classifiers for their data. The k-NN and SVM both provided 0.77 accuracy at age 9 months infants, and the result was statistically significant. Accuracy of 0.80 which was obtained at age 18 months was the highest among the statistically significant results Naïve Bayes offered.

DISCUSSION

The purpose of this paper is to provide information about basic concepts of ML algorithms frequently used in mental health area and their actual application examples in this domain. The analysis of mental health using ML techniques is mainly focused on the supervised learning setting for classifications. We reviewed PTSD, schizophrenia, depression, ASD, bipolar disease as domains where ML techniques were employed in mental health, and SVM, GBM, Random Forest, KNN, Naïve Bayes have been applied in these domains.

Among many ML techniques, SVM is commonly used technique in mental health. It has been employed for all domains in mental health, and most of the SVM classifiers built in the papers revealed more than 75% high accuracy, as expected. In the mental health domain, data are sparse, meaning measured feature values are discrete and features represent just a few characteristics of a data point. This data sparsity is one reason why SVM works well in this domain.²⁷ In some cases, different ML techniques such as GBM were utilized together for feature selection. The Ensemble methods of GBM and Random Forest were also employed as the main algorithm to classify some mental health patients for their merits that they can handle many features simultaneously without feature selection. They showed good performance as SVM on average, but the accuracy of one classifier was slightly less than 60% on the test sample. It implies ML algorithm itself may not guarantee classification in high accuracy. KNN and Naïve Bayes were used only once with SVM under our investigation, but their performance was equivalent or greater than

to that of SVM in some cases.

However, the studies we have examined show several limitations. As can be seen from the above, some studies used a small size of instances, less than 100, to build their classifier through ML techniques. The accuracy of their classifier estimated on the samples may not be generalized since a small sample is not easy to represent the entire population. This may be unavoidable limitations due to finite resources in real-world clinical/diagnostic settings. Still, researchers in practice need to be aware of the fact that ML itself cannot resolve this issue.

Furthermore, many works did not describe even why they chose specific ML techniques. As reviewed above, each method has its own strength and weakness, and its performance depends on the purpose of studies and properties of data. Also, generally ML algorithms are affected a lot by human preprocessing such as hyperparameter tuning of models and data processing for models to be fitted into optimality. That's why we cannot say that one specific algorithm is the best for all domains at all times. Researchers need to be circumspect about the choice of ML algorithm and to elucidate the ground for their choice with other procedures to be explained. In the case where researchers cannot be sure of which ML algorithm is the best for their research in advance, various classifiers need to be fitted to the data and selected by cross-validation. Researchers may analyze their data using several sorts of different classifiers such as SVM, RF and Naïve Bayes and choose the algorithm with the highest accuracy, as in Bosl et al.⁴³

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2017R1C1B1 008002).

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

REFERENCES

- Ozgur C, Kleckner M, Li Y. Selection of Statistical Software for Solving Big Data Problems: A Guide for Businesses, Students, and Universities. *SAGE Open* 2015;1-12.
- OECD. What is PISA? Available at: <http://www.oecd.org/pisa/aboutpisa/>. Accessed Aug 21, 2018.
- McAfee A, Brynjolfsson E. Big data: the management revolution. *Harv Bus Rev* 2012;90:60-66.
- Sagiroglu S, Sinanc D. Big Data: A Review. *IEEE Int Conf Collab Technol Syst (CTS)* 2013;42-47.
- Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and opportunities of big data in health care: a systematic review. *JMIR Med Inform* 2016;4:e38.
- Tam D. Facebook processes more than 500 TB of data daily. Available at: <https://www.cnet.com/news/facebook-processes-more-than-500-tb-of-data-daily/>. Accessed Aug 23, 2012.
- Constine J. How Big Is Facebook's Data? 2.5 Billion Pieces of Content And 500+ Terabytes Ingested Every Day. Available at: [268 Psychiatry Investig 2019;16\(4\):262-269](https://tech-

</div>
<div data-bbox=)

- crunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/. Accessed Aug 23, 2012.
8. Feldman B, Martin E, Skotnes T. Big Data in Healthcare Hype and Hope. Available at: https://www.ghdonline.org/uploads/big-data-in-healthcare_B_Kaplan_2012.pdf. Accessed Nov 9, 2016.
 9. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) National Human Genome Research Institute. Available at: www.genome.gov/sequencingcostsdata. Accessed Apr 25, 2018.
 10. Goense J, Bohraus Y, Logothetis NK. fMRI at High Spatial Resolution: Implications for BOLD-Models. *Front Comput Neurosci* 2016;10:66.
 11. Nuwer MR. The development of EEG brain mapping. *J Clin Neurophysiol* 1990;7:459-471.
 12. Turing AM. Computing Machinery and Intelligence. In: Epstein R, Roberts G, Beber G, Editors. *Parsing the Turing Test*. 1. Dordrecht: Springer, 2009, p.23-65.
 13. Kohavi R, Provost F. Glossary of terms. *Mach Learn* 1998;30:271-274.
 14. Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on big data: opportunities and challenges. *Neurocomputing* 2017;237:350-361.
 15. Mullainathan S, Spiess J. Machine learning: an applied econometric approach. *J Econ Perspect* 2017;31:87-106.
 16. Obermeyer Z, Emanuel EJ. Predicting the future—big data, machine learning, and clinical medicine. *N Engl J Med* 2016;375:1216-1219.
 17. Tsai CW, Lai CF, Chao HC, Vasilakos AV. Big data analytics: a survey. *J Big Data* 2015;2:1-32.
 18. Kaura P, Sharmab M, Mittal M. Big data and machine learning based secure healthcare framework. *Procedia Comput Sci* 2018;132:1049-1059.
 19. Borowiec S. AlphaGo seals 4-1 victory over Go grandmaster Lee sedol. Available at: <https://www.theguardian.com/technology/2016/mar/15/googles-alphago-seals-4-1-victory-over-grandmaster-lee-sedol>. Accessed Oct 12, 2016.
 20. Zhang GP. Avoiding pitfalls in neural network research. *IEEE Trans Syst Man Cybern C Appl Rev* 2007;37:3-16.
 21. Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and machine learning forecasting methods: concerns and ways forward. *PLoS ONE* 2018;13:e0194889.
 22. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intell Syst Appl* 1998;13:18-28.
 23. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat* 2001;29:1189-1232.
 24. Liaw A, Wiener M. Classification and regression by RandomForest. *R News* 2002;2:18-22.
 25. Rish I. An empirical study of the naive Bayes classifier. *IJCAI 2001 Work Empir Methods Artif Intell* 2001;3:41-46.
 26. Cover T, Hart P. Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 1967;13:21-27.
 27. Duan KB, Keerthi SS. Which Is the Best Multiclass SVM Method? An Empirical Study. In: Oza NC, Polikar R, Kittler J, Roli F, Editors. *Multiple Classifier Systems*. MCS 2005. Lecture Notes in Computer Science, vol 3541. Berlin, Heidelberg: Springer, 2005, p.278-285.
 28. Khoshgoftaar TM, Golawala M, Van Hulse J. An empirical study of learning from imbalanced data using random forest. *ICTAI 2007* 2007; 2:310-317.
 29. Lin Y, Jeon Y. Random forests and adaptive nearest neighbors. *J Am Stat Assoc* 2006;101:578-590.
 30. Mason L, Baxter J, Bartlett PL, Frean MR. Boosting algorithms as gradient descent. *Adv Neural Inf Process Syst* 2000;512-518.
 31. Schapire RE. The Boosting Approach to Machine Learning: An Overview. In: Denison DD, Hansen MH, Holmes CC, Mallick B, Yu B, Editors. *Nonlinear Estimation and Classification*. Lecture Notes in Statistics, vol 171. New York: Springer, 2003, p.149-171.
 32. Augsburg M, Elbert T. When do traumatic experiences alter risk-taking behavior? A machine learning analysis of reports from refugees. *PLoS ONE* 2017;12:e0177617.
 33. Galatzer-Levy IR, Ma S, Statnikov A, Yehuda R, Shalev AY. Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD. *Transl Psychiatry* 2017;7:e0.
 34. Karstoft KI, Galatzer-Levy IR, Statnikov A, Li Z, Shalev AY. Bridging a translational gap: using machine learning to improve the prediction of PTSD. *BMC Psychiatry* 2015;15:30.
 35. Schnack HG, Nieuwenhuis M, van Haren NE, Abramovic L, Scheewe TW, Brouwer RM, et al. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage* 2014;84:299-306.
 36. Greenstein D, Malley JD, Weisinger B, Clasen L, Gogtay N. Using multivariate machine learning methods and structural MRI to classify childhood onset schizophrenia and healthy controls. *Front Psychiatry* 2012;3:53.
 37. Malley JD, Kruppa J, Dasgupta A, Malley KG, Ziegler A. Probability machines: consistent probability estimation using nonparametric learning machines. *Methods Inf Med* 2012;51:74-81.
 38. Malley JD, Malley KG, Pajevic S. *Statistical Learning for Biomedical Data*. New York: Cambridge University Press; 2011.
 39. Yang H, Liu J, Sui J, Pearlson G, Calhoun VD. A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia. *Front Hum Neurosci* 2010;4:192.
 40. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* 2016;3:243-250.
 41. Nakai Y, Takiguchi T, Matsui G, Yamaoka N, Takada S. Detecting abnormal word utterances in children with autism spectrum disorders: machine-learning-based voice analysis versus speech therapists. *Percept Mot Skills* 2017;124:961-973.
 42. Maenner MJ, Yeargin-Allsopp M, Van Naarden Braun K, Christensen DL, Schieve LA. Development of a Machine Learning Algorithm for the Surveillance of Autism Spectrum Disorder. *PLoS ONE* 2016;11: e0168224.
 43. Bosl W, Tierney A, Tager-Flusberg H, Nelson C. EEG complexity as a biomarker for autism spectrum disorder risk. *BMC Med* 2011;9:18.