Brazilian Journal of health Review

An overview of machine learning in health related areas: pitfalls and opportunities
/
Uma visão geral do aprendizado de máquina em áreas relacionadas à saúde: armadilhas e oportunidades

DOI:10.34119/bjhrv3n2-204

Recebimento dos originais: 05/03/2019
Aceitação para publicação: 29/04/2020

Renato de Lima Vitorasso
Doctoral student in Biomedical Engineering field in Escola Politécnica - USP
Institution: Universidade Estadual de São Paulo
Address: Av. Professor Luciano Gualberto, travessa 3, 158. 05508-010, São Paulo, SP, Brazil
E-mail: renatovitorasso@usp.br

Carolina de Souza Ribeiro Vitorasso
Medical Resident in Plastic Surgery
Institution: Hospital Geral de Vila Penteado
Address: Av. Ministro Petrônio Portela, 1642, 02802-120, São Paulo, SP, Brazil
E-mail: carolinasribeiro@hotmail.com

ABSTRACT
Machine learning techniques are on the spotlight in current scientific literature and these methods are gaining prominence in the health field. However, there are a few considerations that must be taken before conducting a study with machine learning techniques. This paper aims to provide an overview of machine learning methods applied to studies of health related areas. Additionally, this article will discuss important points about data preparation that may influence on the prediction outcome; comparison with statistical analysis; and potential applications. A literature search was carried out, using IEEE xplore and Pubmed, of publications from the last 10 years. Undoubtedly machine learning is becoming more and more present in science. However, the unfamiliarity with this technology may hinder or jeopardize its application. As any scientific tool, machine learning presents positive points along with limitations and both aspects should be considered in every analysis. The researcher must select the most adequate method and consider all repercussions of data preparation on the predictive model. A special attention should be given towards distance based techniques. ML techniques are full with potential applications; however these methods did not replace classical statistical analysis and, yet, they will continue to be an important tool to in health areas.

Keywords: Machine learning; medical informatics; clustering, classification analyses, biostatistics.

RESUMO
As técnicas de aprendizado de máquina estão em destaque na literatura científica atual e esses métodos estão ganhando destaque no campo da saúde. No entanto, existem algumas considerações que devem ser tomadas antes de realizar um estudo com técnicas de aprendizado de máquina. Este artigo tem como objetivo fornecer uma visão geral dos métodos de aprendizado de máquina aplicados a estudos de áreas relacionadas à saúde. Além disso, este

Brazilian Journal of health Review

artigo discutirá pontos importantes sobre a preparação de dados que podem influenciar no resultado da previsão; comparação com análise estatística; e aplicações em potencial. Foi realizada uma pesquisa bibliográfica, utilizando IEEE xplore e Pubmed, de publicações dos últimos 10 anos. Sem dúvida, o aprendizado de máquina está se tornando cada vez mais presente na ciência. No entanto, o desconhecimento dessa tecnologia pode prejudicar ou comprometer sua aplicação. Como qualquer ferramenta científica, o aprendizado de máquina apresenta pontos positivos, além de limitações, e ambos os aspectos devem ser considerados em todas as análises. O pesquisador deve selecionar o método mais adequado e considerar todas as repercussões da preparação dos dados no modelo preditivo. Uma atenção especial deve ser dada às técnicas baseadas na distância. As técnicas de ML estão cheias de possíveis aplicações; no entanto, esses métodos não substituíram a análise estatística clássica e, ainda assim, continuarão sendo uma ferramenta importante nas áreas da saúde.

Palavras chave: Aprendizado de máquina; informática médica; agrupamentos, análises de classificação, bioestatística.

1 INTRODUCTION

Machine Learning (ML) is ubiquitous in modern society. We are surrounded by applications, web sites and a countless examples of platforms that use ML. Health areas are not an exception. However, despite a few comparisons of ML and statistical analysis [1,2], ML techniques still seem to be less prevalent than classical statistical analysis in most health research centers.

There are a few reasons that may answer why ML techniques might be neglected: familiarity with statistical analysis; unfamiliarity with ML methods; singularities related with each method; the fact that in most ML methods the user can not know why computer's routines predict the outcome based on the features.

Regardless of the reason, this work aims to provide an overview of ML methods along with studies of health related areas. Additionally, this work will discuss important points about data preparation that may influence on the prediction outcome; comparison with statistical analysis; and potential applications.

2 METHODS

A literature search was carried out, using IEEE xplore and Pubmed, of publications from the last 10 years. References of relevant studies outside this period were also included. The strategy of search consisted in the combination of the terms: "Machine Learning", "Health", "Cardiovascular", "Respiratory System", "Neuroscience", "Biomedical", "Imaging", "Death", "Statistical Analysis", "K-means", "DBSCAN", "Naive Bayes", "K-NN", "Decision trees",

Brazilian Journal of health Review

"Classifier ensemble", "Random Forest", "Suport Vector Machine", "Neural Network", "Deep Learning", "Overfitting". Only articles published in English were included.

3 CLUSTERING ALGORITHMS

ML may be categorized on the utilization of labels in training data. There are three categories of ML algorithms: supervised learning, unsupervised learning, and semi-supervised learning [3]. Basically, the analysis is unsupervised if the researcher does not have the classification or label of each item of the data base.

The unsupervised learning is propitious scenario to studies that need to separate individuals in clusters, through *p* features, but the researcher does not know *a priori* the individual condition.

3.1 K-MEANS

K-means is simple, iterative and a very influential in ML [4]. There are several variations in this technique and there are health studies using both classical form and variations [5,6]. K-means operates with p-dimensional vectors, where p is the number of features. Initially, k points in p-dimensional space may be pseudo-randomly chosen by a computer routine. The user sets the k number of clusters where k is the number of groups. This number is different in each study and a visual plot may help when this plot is possible due to the limitation of dimensions.

Once the k points or centroids in space are selected, the routine verifies the distance, e.g. Euclidean, of each patient to these centroids. Thus, if the patient A is closer to centroid 1 than 2, this patient belongs to the cluster 1. Then, after each individual or point is located in its particular cluster, the average of all data becomes a new centroid.

Since this is an iterative method, the distance of each individual is measured again to acquire a new data average and a new centroid until the data average converges, i.e. the new centroid or data mean does not vary anymore over iterations. It is possible to implement K-means and DBScan using sklearn.cluster module of Python.

3.1.1 Limitations

K-means present a few limitations [4], it is sensitive to the initialization, hence it may be interesting to initialize with different starting points. Since K-means uses distance among points or patients and the centroids, the data scale interferes with the prediction. Finally, it performs better if the data is reasonably separated spherical balls.

Brazilian Journal of health Review

3.2 DBSCAN

DBSCAN (Density-Based Spatial Clustering of Application with Noise) is a density based algorithm [7]. There are numerous algorithms based on density and this section will discuss DBSCAN. These methods separate clusters with high density from regions with low densities. One difference and positive point over K-means is the fact that distribution of the data on the p-dimensional space does not need to present a reasonably separated spherical balls.

In order to fully understand DBSCAN technique there are a few important concepts that need to be presented [7]: 1) Eps-neighborhood of point p is the circular area with EPS radius centered in p; 2) density of points p is the number of points within Eps-neighborhood of point p; 3) core point is a point with a density of points greater than a minimum threshold of points (MinPts) and this threshold is specified by the user; 4) border point is a point contained in Eps-neighborhood of a core point, but it is not one core point; 5) noise point is neither a border or a core point.

DBSCANs core idea is the border limitation, with the border points and all core points inside, signalizing a cluster. All points that do not fulfill the border or core points criteria are noise and they are not classified as any cluster.

This method does not need a prior number of cluster (k), only the EPS radius and MinPts. Since this method does not need a prior number of clusters, it is a good technique for clustering online medical decision systems and recently it has been associated with a Piece-wise Aggregate Approximation, a popular dimension reduction method in data stream mining [8]. Reduction of the p-dimensional space is an important topic on data preparation. Finally, it works well in different shapes of data base and is less sensitive to outliers than k-means.

3.2.1 Limitations

This process is sensitive to the EPS radius and MinPts; it depends on the method used to assess the distance among points (when k-means it is usually Euclidian distance); it cannot categorize cluster with different densities.

4 CLASSIFICATION ALGORITHMS

Supervised-learning is the most common form in ML and is widely used for classification [3]. Data is usually collected and labeled in categories. These categories may be: diseases; radiological findings; death; and many others. The main purpose of supervised

learning is to find an appropriate input-output function from training data, which generalizes well against the testing data [3].

4.1 NAIVE BAYES (NB)

It is a simple model of ML and yet a very influential one [4]. NB is based on Bayes' theorem (eq. 1):

$$P(c|x) = \frac{P(x|c)\cdot P(c)}{P(x)} \quad (1)$$

where $P(c|x)$ and $P(x|c)$ are the probability of c given x and x given c; and $P(c)$ and $P(x)$ are the probability of c and x respectively. Furthermore c is the class (label) and x the predictor or the vector with the features. NB assumes that the all features are independent [9] and have the same weight. In order to calculate the probability of an outcome based on a vector X we have:

$$P(c|X) = P(x_1|c) \cdot P(x_2|c) \ldots P(x_n|c) \cdot P(c) \quad (2)$$

Below in table 1 is presented a fictitious example:

Patient	Feature 1	Feature 2	Feature 3	Disease
1	A	C	E	Yes
2	A	C	E	No
3	A	D	F	No
4	B	C	F	Yes
5	B	D	F	Yes
6	A	C	F	Yes
7	A	C	E	Yes
8	B	D	E	No
9	A	C	F	?

Table1 - Fictitious example for Naive Bayes calculation

Brazilian Journal of health Review

In order to predict the probability of having a disease with $\{A, C, F\}$:

$$P(Yes|\{A, C, F\}) = \frac{3}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} \cdot \frac{5}{8} = 0.18 \quad (3)$$

Calculating the probability for $P(No|\{A, C, F\}) = 0.028$ we realize that it is more probable that the patient 9 will have this particular disease. It is possible to transform these values in a scale in which both probabilities summed provide 1 or 100%. Despite its simplicity, NB is a competitive predictive model and it was used as a predictor in recent studies [9,10]. In addition, a large number of authors have proposed modifications to NB [4]. But, overall, this method remains with a low computational cost.

As for the most techniques, there are methods or libraries implemented in Python responsible to calculate NB. Besides that, it is possible, due to the simplicity of this method, to implement directly a NB routine to assess these probabilities. Finally, it is possible to use a R package *naiveBayesian* [10].

4.1.1 Limitations

The Gaussian assumptions and the independence criteria are not always fulfilled. However, once more, it is a robust a simple method with new a variety of alterations [4].

4.2 K-NEAREST NEIGHBORS (K-NN)

K-NN is, also, a method based on distance among points in a p-dimensional space, in which the p is the number of features in the model. However, this is a classification method, thus it needs the label in advance. For example, if the outcome is surgical intervention, the outcome would be: intervention or not intervention.

This classifier measures the new point distance to the pre-established classes. The distance of the new point is evaluated to each component of the data set. If the number of "nearest neighbors" of the class C_1 is higher than $C_2, C_3 \ldots C_n$ this new point belongs to C_1. The core of this method is intuitive, but may present the same complications of any distance based ML method. As any Euclidian distance method, the distance between a vector X and Y is:

$$d_E(X, Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (3)$$

Brazilian Journal of health Review

One good example of a study that used K-NN is the work of Jian and Chen [11]. In this paper, they describe the acquisition of accelerometry using a developed device, the feature extraction and, by K-NN, they present a set up to predict fall.

4.2.1 Limitations

There are a few key issues that may affect the performance of k-NN: the choice of k number of neighbors; the radius size; the number of parameters and the scale of the features since it is a distance based method.

4.3 DECISION TREES (DT)

This technique is a wide applied and a particularly informative method in medical and health areas. In most cases of ML models the researcher does not know which feature presents a higher weight or how to predict the outcome based solely on the features. Thus, in most scenarios the ML is a "black box" and it depends on computer's applications.

With DT it is possible to draw a flow chart based on the features and it is possible to predict the outcome and the probability of this outcome visually. This is possible since DT algorithms is a top down technique which selects the most important feature to top and split the path to the next important feature (Figure 1).

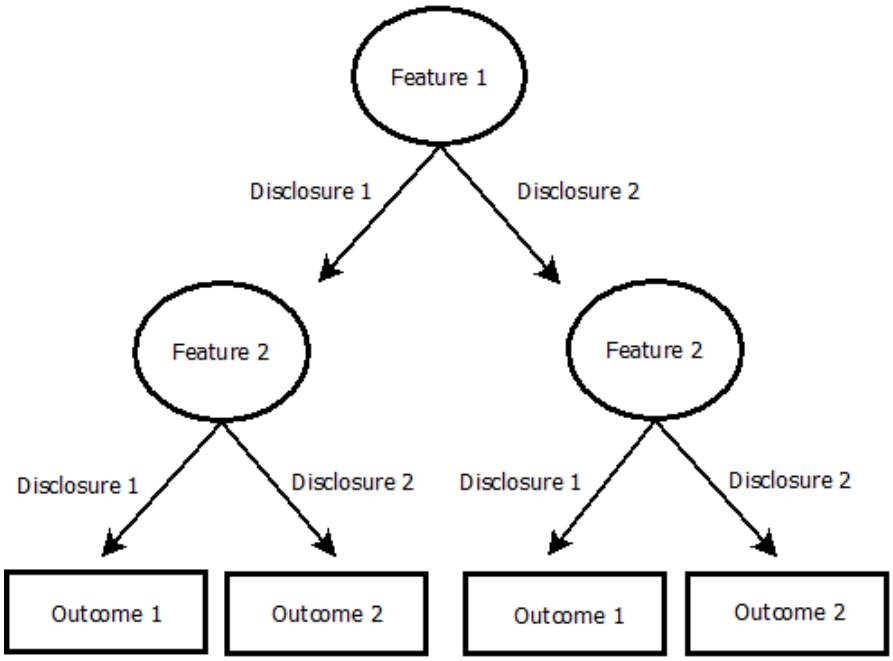

Figure 1 – An example of decision tree with 2 "branches". Feature 1 is the most important feature and each outcome has a probability associated.

Brazilian Journal of health Review

In the figure above there were two divisions or branches of this tree. This number of branches may be determined in computer's routine. Each outcome presents a probability of happening. In order to provide a better perspective, the outcome may be a disease instauration and the features could be laboratorial exams.

One point remains obscured. How to determine which feature is more important? One of the most common algorithms, or maybe the most common, is the C4.5 [4,12]. This algorithm selects or classifies the feature based on its "information gain" (IG) and, in case of C4.5, the IG is based on the entropy [12]:

$$IG(X) = H(Y) - H(Y|X) \quad (4)$$

$$H(Y) = -\sum_{i=1}^{2} P(y_i) log_2 P(y_i) \quad (5)$$

$$H(Y|X) = -\sum_{j=1}^{v} P(x_j) \sum_{i=1}^{2} P(y_i|x_j) log_2 P(y_i|x_j) \quad (6)$$

where $H(Y)$ is the entropy of the decision variables, if it is a disease, the outcomes can be presence or absence. The $H(Y|X)$ is the entropy the decision variable (Y) given the observed features values $x_j \in \{x_1, \dots x_v\}$. This features may be IMC, gender, weight, among others and in case of binary features or features categorized in 2 groups, the $v = 2$. $P(y_i|x_j)$ is the conditional probability of y_i given x_j. Finally, the IG(X) is the information gain of a particular feature, such as gender.

Therefore, the feature that provides the more information gain (IG), i.e. the higher capacity of predicting the outcomes, appears on top of the flow chart followed by the feature with the second highest IG and so on.

Recently two relevant studies used DT in order to: 1) predict right ventricular failure and 2) improve a model of decision tree in assistive devices for people with disabilities [12,13].

It is possible to implement a routine in python importing "tree" from "sklearn" library. Specifically for the C4.5 it is possible to create a C45 object.

Brazilian Journal of health Review

4.3.1 Limitations

The gain of information can prioritize features with a great number of values. Nevertheless, the problems that may occur with DT, like overfitting, can be attenuated by selecting a proper number of divisions.

4.4 ENSEMBLE CLASSIFICATION(EC)

This technique is worthy of a study on its own and indeed this year a review was published discussing EC methods in breast cancer [14]. ECs are methods that combine more than one technique to solve a single task.

Ensemble techniques can be Homogeneous or Heterogeneous [14,15]. Heterogeneous ensemble is a model that combines at least two different base methods. Whereas, the homogeneous ensemble is used to refer to: 1) an ensemble that combines one base method with at least two different configurations or different variants, 2) an ensemble that combines one base method with one meta ensemble, such as Bagging or Boosting [14].

Another difference among ensemble techniques is the association of predictive models (series or parallel). There are numerous ensembles models in the literature and yet, based on a systematic review of Ensemble Effort Estimation, there is no particular ensemble technique that outperforms other ensemble method [15]. However, also based on this systematic review, in the majority of cases, ensemble techniques are more accurate than any single model.

One popular example of an ensemble model is the random forest. This technique is a homogeneous one and is based on DT. A recent study related to prediction of colorectal cancer compared different predictive models and they found out that the random forest was the most stable method, despite being outperformed by, for example, Support Vector Machine [16].

4.5 SUPPORT VECTOR MACHINE (SVM)

Support Vector Machines (SVM) offers one of the most robust and accurate methods among all algorithms. It has a solid theoretical foundation, requires few examples for training, and is insensitive to the number of dimensions [4].

SVM can be used both on supervised ML and regression problems. The objective of SVM is to find a hyperplane which divides the two or more classes of data. The function's equation of the hyperplane depends on the kernel used such as: sigmoid, radial basis function or polynomial [17]. In order to create a routine in Python the researcher could, once more, import

Brazilian Journal of health Review

from “sklearn” and this time import the module “svm” and carefully read the documentation for “svm.svc”.

An interesting study regarding SVM in detection of lung nodules was carried out [17]. In this paper, the researchers used tomography imaging and, additionally, this study became relevant not only for the SVM but because of the features extracted from the tomography images.

4.5.1 Limitations

One important limitation of SVM is related with the fact that the kernel selection is chosen by the user and interferes directly with the generalization of the model. This phenomenon is similar to the selection of predictive equation in a regression analysis.

4.6 NEURAL NETWORK

Neural network is a robust and very prevalent model in current literature. This predictive model is inspired by an animal neural network in which a neuron via synapses participates in an intricate network. A neural network model can be interpreted as below:

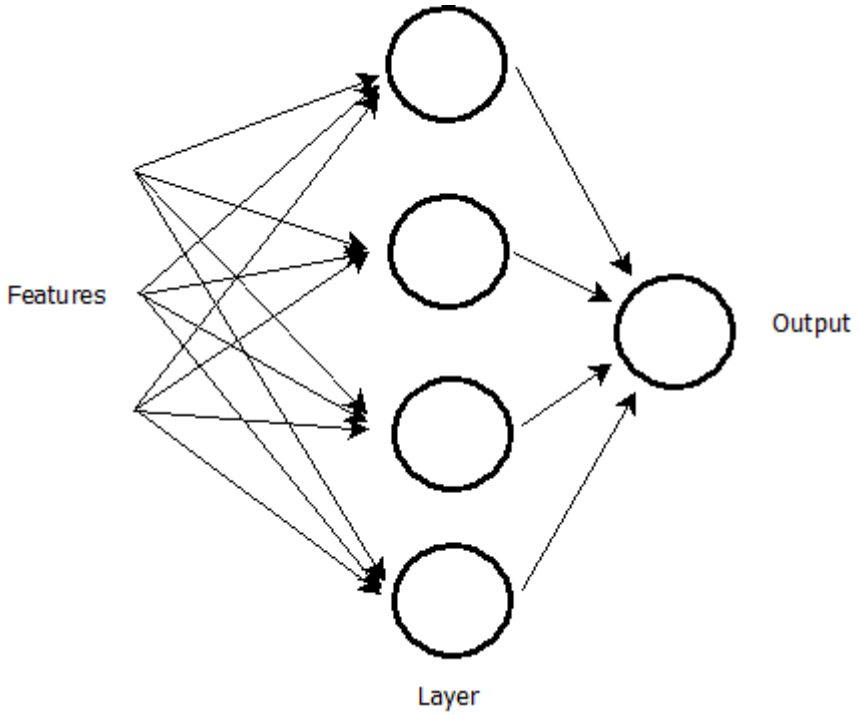

Figure 2 – An example of a small neural network. Each circle represents a neuron. In this network there is only one hidden layer separating the features from the last neuron.

Brazilian Journal of health Review

As any ML method, the neural network needs a dataset in order to predict an outcome. The neural network is composes by layers and the number of layer and nodes vary from each situation. A initial suggestion of nodes in one hidden layer is half of the length of the feature vector [18].

Each neuron performs a weighted sum of the entry and applies an activation function. There are several activation functions, however two are mention worthy: the sigmoid activation function [18] and, perhaps the most used one, ReLu (Rectified Linear Unit). The latter basically attributes value of 0 to each negative value and is the identity function to positive values. Neural networks operates trying to find the best weighted values in order to predict the outcome. There are many ways to minimize the error and find the best set of weights, but in any scenario the goal is the same.

4.6.1 Limitations

This technique depends on the initial set up of nodes and activation function. This model can be very expensive computationally. In addition, this method can be usually interpreted as a black box.

4.7 DEEP LEARNING

Deep learning describes a set of computational models composed of multiple layers of data processing, which makes it possible to learn by representing these data through several levels of abstraction [19]. These levels of representation, obtained by composing simple, but nonlinear modules that transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level [20,21].

While artificial neural networks are usually composed by three layers and one transformation toward the final outputs, deep learning architectures are constituted by several layers of neural networks [21]. Therefore, this model presents the same limitations that neural networks and it is more expensive computationally.

In health areas, deep learning is commonly used in imaging studies since there are numerous features that need to be extracted from an image. So, there is a huge spotlight on tomography and radiology fields. Last year a review on deep learning and ophthalmology [22] and, also last year, a review on deep learning and radiotherapy [19] were carried out.

One important issue regarding deep learning was discussed by Meyer, P. et al [19], they discussed that one of the main criticisms of deep learning is the lack of theory concerning and

the fact that standard general principles are mostly empirically obtained. The choice of the general architecture of the network (such as: the layout of the layers; their number; and the size of filters) best suited to the problem to be solved is mostly guided by intuition and carried out experimentally.

5 DATA PREPARATION (DATAPREP)

DataPrep is as important as the selection of the model itself. Usually, regarding methods as decision tree, for example, or Naive Bayes, one could have a sheet with several features and the disclosures.

Several cautions should be taken as: the proper selection of the features; the elimination of features that do not aggregate information to the model (this can be done by eliminating strongly associated features); the proper scaling of features (the features can be transformed into a 0 to 1 scale), specially to distance based models; the correct use of the selected method, e.g. the proper number of division in a decision tree avoiding overfitting.

However, DataPrep goes beyond the correct treatment of the features. The features extraction *per se* is highly relevant. Imaging processing and extraction are a whole research field [23].

In a sleep staging study the features needed to be extracted from EEG signals. The authors separated the features in: Time domain; Time-frequency domain; Non-linear. The researchers extracted and used, in the prediction models, 30 features among those 3 previous groups.

6 EVALUATION OF THE MODELS

A common evaluation method is cross-validation, in which the data is split into 'training' and 'test' datasets. Models are developed using only the training set, and model performance is assessed using the test set. The training and test set must be kept separate for all analysis steps [24].

Additionally, it is possible to verify the accuracy based on a confusion matrix [25], something very common in epidemiological studies. Another possible way to assess how good a model is performing is through the assessment of Area Under the Curve (AUC) of a ROC curve.

One of the main concerns of evaluating a predictive model is how capable of generalizing the prediction the model is. In other words, how good the model is without

overfitting. Overfitting is the characteristic of a highly specific model in which the technique performs very well in the test dataset, however is incapable of predicting in a real world situation.

7 MACHINE LEARNING VS STATISTICAL ANALYSIS

There are studies comparing ML with classical statistical analysis used to predict outcomes in the health field. The question "is machine learning more capable to predict than a classical statistical analysis?" may not have a complete answer yet. In a study of fetal growth abnormalities the ML did not present a better performance than classical analysis [1]. However, in a multicenter comparison predicting clinical deterioration on the wards the selected ML techniques were more accurate than logistic regression [2].

The researcher must have the knowledge of the potentialities and limitations of each method at the moment of selecting the model. Characteristics like the size of the sample, the number of features, if the "black box" paradigm is a problem, the computational cost, among others should be considered in the model selection.

8 POTENTIALLY USEFUL ML APPLICATIONS

We will probably perceive an increase in the number of studies with ML techniques and an increase in clinical application of ML as well. Maybe in the same way that we have user friendly applications destined to robust statistical analysis, we may have, on an also close future, more platforms and applications dedicated to health related fields.

Processing medical data with multi-layer neural networks increased the predictive power for several specific applications in different clinical domains [21]. Deep learning performs analysis that classical statistics or non-multi-layer neural network cannot, thus one tool specifically designed to feature extraction of medical images can facilitate the analysis.

9 CONCLUSIONS

The unfamiliarity with Machine Learning may hinder or jeopardize its application. As any scientific tool, ML presents positive points and limitations that should be considered in every analysis. The researcher must select the most adequate method and consider all repercussions of data preparation on the predictive model. A special attention should be given towards distance based techniques. ML techniques are full with potential applications; however

these methods did not replace classical statistical analysis and, yet, they will continue to be an important tool to in health areas.

ACKNOWLEDGMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior - Brazil (CAPES) - Finance Code 001 (88882.333348/2019-01)

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

1. Kuhle S, Maguire B, Zhang H, Hamilton D, Allen AC, Joseph KS, Allen VM, Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. *BMC Pregnancy Childbirth*. 2018; 18: 333.

2. Churpek MM, Yuen TC, Winslow C, Meltzer DO, Kattan MW, Edelson DP, Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit. Care Med.* 2016; 44: 368–374.

3. Zhang Z, Sejdić E, Radiological images and machine learning: Trends, perspectives, and prospects. *Comput. Biol. Med.* 2019; 108: 354–370.

4. Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, McLachlan GJ, Ng A, Liu B, Yu PS, Zhou ZH, Steinbach M, Hand DJ, Steinberg D, Top 10 algorithms in data mining. *Knowl. Inf. Syst.* 2008; 14: 1–37.

5. Kakushadze Z, Yu W, *K-means and cluster models for cancer signatures. *Biomol. Detect. Quantif.* 2017; 13: 7–31.

6. Zampieri FG, Costa EL, Iwashyna TJ, Carvalho CRR, Damiani LP, Taniguchi LU, Amato MBP, Cavalcanti AB, Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial Investigators, Heterogeneous effects of alveolar recruitment in acute respiratory distress syndrome: a machine learning reanalysis of the Alveolar Recruitment for Acute Respiratory Distress Syndrome Trial. *Br. J. Anaesth.* 2019; 123: 88–95.

7. Meng'Ao L, Dongxue M, Songyuan G, Shufen L, Research and Improvement of DBSCAN Cluster Algorithm, in: 2015 7th Int. Conf. Inf. Technol. Med. Educ., IEEE, 2015; pp. 537–540.

8. Al-Shammari A, Zhou R, Naseriparsaa M, Liu C, An effective density-based clustering and dynamic maintenance framework for evolving medical data streams. *Int. J. Med. Inform.* 2019; 126: 176–186.

9. Miasnikof P, Giannakeas V, Gomes M, Aleksandrowicz L, Shestopaloff AY, Alam D, Tollman S, Samarikhalaj A, Jha P, Naive Bayes classifiers for verbal autopsies: comparison to physician-based classification for 21,000 child and adult deaths. *BMC Med.* 2015; 13: 286.

10. Cui S, Zhao L, Wang Y, Dong Q, Ma J, Wang Y, Zhao W, Ma X, Using Naive Bayes Classifier to predict osteonecrosis of the femoral head with cannulated screw fixation. *Injury*. 2018; 49: 1865–1870.

11. Jian H, Chen H, A portable fall detection and alerting system based on k-NN algorithm and remote medicine. *China Commun.* 2015; 12: 23–31.

12. Wang Y, Simon M, Bonde P, Harris BU, Teuteberg JJ, Kormos RL, Antaki JL, Prognosis of Right Ventricular Failure in Patients With Left Ventricular Assist Device Based on Decision Tree With SMOTE. *IEEE Trans. Inf. Technol. Biomed.* 2012; 16: 383–390.

13. Chi C-F, Tseng L-K, Jang Y, Pruning a Decision Tree for Selecting Computer-Related Assistive Devices for People With Disabilities. *IEEE Trans. Neural Syst. Rehabil. Eng.* 2012; 20: 564–573.

14. Hosni M, Abnane I, Idri A, Carrillo de Gea JM, Fernández Alemán JL, Reviewing ensemble classification methods in breast cancer. *Comput. Methods Programs Biomed.* 2019; 177: 89–112.

15. Idri A, Hosni M, Abran A, Systematic literature review of ensemble effort estimation. *J. Syst. Softw.* 2016; 118: 151–175.

16. Cueto-López N, García-Ordás MT, Dávila-Batista V, Moreno V, Aragonés N, Alaiz-Rodríguez R, A comparative study on feature selection for a risk prediction model for colorectal cancer. *Comput. Methods Programs Biomed.* 2019; 177: 219–229.

17. Khan SA, Nazir M, Khan MA, Saba T, Javed K, Rehman A, Akram T, Awais M, Lungs nodule detection framework from computed tomography images using support vector machine. *Microsc. Res. Tech.* 2019; 82: 1256–1266.

18. Zhang Z, Yang K, Qian J, Zhang L, Real-Time Surface EMG Pattern Recognition for Hand Gestures Based on an Artificial Neural Network. *Sensors*. 2019; 19: 3170.

19. Meyer P, Noblet V, Mazzara C, Lallement A, Survey on deep learning for radiotherapy. *Comput. Biol. Med.* 2018; 98: 126–146.

20. LeCun Y, Bengio Y, Hinton G, Deep learning. *Nature*. 2015; 521: 436–44.

21. Miotto R, Wang F, Wang S, Jiang X, Dudley JT, Deep learning for healthcare: review, opportunities and challenges. *Brief. Bioinform.* 2018; 19: 1236–1246.

22. Grewal PS, Oloumi F, Rubin U, Tennant MTS, Deep learning in ophthalmology: a review. *Can. J. Ophthalmol.* 2018; 53: 309–313.

23. Tian DP, A Review on Image Feature Extraction and Representation Techniques, 2013; https://pdfs.semanticscholar.org/a820/b7882ea2365e0396d100a412126e93611a6f.pdf (accessed July 22, 2019).

24. Jollans L, Boyle R, Artiges E, Banaschewski T, Desrivières S, Grigis A, Martinot J-L, Paus T, Smolka MN, Walter H, Schumann G, Garavan H, Whelan R, Quantifying performance of machine learning methods for neuroimaging data. *Neuroimage*. 2019; 199: 351–365.

25. Bartholomai JA, Frieboes HB, Lung Cancer Survival Prediction via Machine Learning Regression, Classification, and Statistical Techniques, in: 2018 IEEE Int. Symp. Signal Process. Inf. Technol., IEEE, 2018; pp. 632–637.

Brazilian Journal of health Review