Predictive analysis using machine learning: Review of trends and methods

Conference Paper · November 2020

DOI: 10.1109/ISAECT50560.2020.9523703

3 authors:

Loola Bokonda
Mohammadia School of Engineers
9 PUBLICATIONS **62** CITATIONS
SEE PROFILE

Ouazzani Touhami Khadija
Mohammadia School of Engineers
20 PUBLICATIONS **150** CITATIONS
SEE PROFILE

Nissrine Souissi
MINES-RABAT School
96 PUBLICATIONS **694** CITATIONS
SEE PROFILE

Predictive analysis using machine learning: Review of trends and methods

1st Patrick Loola Bokonda
Department of Computer Science
Mohammed V University in Rabat. EMI, Siweb team
Rabat, Morocco
loola.bokonda@gmail.com

2nd Khadija Ouazzani-Touhami
Department of Computer Science
MINES-RABAT School
Rabat, Morocco
ouazzani@enim.ac.ma

3rd Nissrine Souissi
Department of Computer Science
MINES-RABAT School
Rabat, Morocco
souissi@enim.ac.ma

***Abstract*—Artificial Intelligence (AI) has been growing considerably over the last ten years. Machine Learning (ML) is probably the most popular branch of AI to date. Most systems that use ML methods use them to perform predictive analysis. This paper aims to conduct a literature review of trends and methods of machine learning used for predictive analysis. To do this, we carried out a collection of research papers from three scientific databases. We then considered selection criteria in order to study only papers published in the last five years, prioritizing those published in peer-reviewed scientific journals. This process led to the selection of 30 research papers that were considered for this review. The purpose of this study is to provide researchers, companies or anyone wishing to perform predictive analysis with clues that will enable them to choose the best ML method(s) according to its field of application, based on the latest research works in the literature. This study highlighted the most used methods by field of application: DT and ANN in education, LR, RF and DT in building, DT in botany, RF and ANN in social science and RF in medicine.**

***Keywords*—artificial intelligence, machine learning methods, predictive analysis, supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, medicine**

I. Introduction

For the past ten years Artificial Intelligence (AI) has experienced a renaissance and particularly Machine Learning (ML) has been the subject of great attention. The ultimate goal of AI is to make machines capable of performing tasks previously considered intelligent [1], [2] better than humans. Is this not the promise of a new revolution that will overturn our relationship to work and knowledge [2] ?

While some see this as an unexpected opportunity to improve our performance in several areas, others see it as the danger of an adversary ready to take the place of man in various sectors. But never mind, researchers, politicians, companies and governments all use AI for a variety of reasons: cancer control, voter profiling, search engines, programmatic advertisements, soldiers' weapons, weather, data processing , etc.

A common element in most AI use cases is prediction. Prediction of high risk of cancer, profiling of voters most likely to vote for a particular candidate, prediction of driver behavior [3], prediction of videos and/or advertisements that may be of interest to a particular person, etc. Predictive analysis is increasingly used.

But a recurring question arises every time a predictive analysis needs to be performed: Which method should be used? This paper is a literature review of trends and methods of ML used for predictive analysis in the most recent studies.

Since ML is a discipline with its own terminology, and in order to familiarize researchers newly interested in ML with concepts specific to this discipline, we have devoted section II to the definition of terms used in ML. The aim is to facilitate the understanding and use of the analysis and results of this study.

The approach undertaken and explained in section III has made it possible to select and study only the research carried out over the last five years. The aim is to provide researchers, companies or anyone wishing to perform predictive analysis with clues to help them choose the best ML method(s) for their field of application, based on the latest research in the literature. Only papers from refereed journals were included in this review.

The result of this study is presented in section IV. These are the disciplines where ML methods are frequently used for predictive analysis, the most used learning techniques, the most used ML methods, and even the most used ML methods by discipline. The result presented in this way does not dictate which method to choose, but rather what other researchers around the world are using in a specific discipline to do predictive analysis.

Our prospects for future research and the conclusion of this work are presented in Section V.

II. BACKGROUND

Machine Learning is defined by Melo Lima and Dursun Delen in [4] as: "A subset of artificial intelligence, which is often applied when computing devices attempt to mimic human cognitive functions related to learning and problem solving processes in order to achieve "optimal" results". For their part, Harleen Kaur and Vinita Kumari [5] define Machine Learning as "the development of algorithms and techniques that enable computers to learn and acquire intelligence based on past experience. This is a branch of Artificial Intelligence (AI) and is closely related to statistics. Learning means that the system is able to identify and understand the data entered, so that it can make decisions and predictions based on them".

In the context of this study, we use the following summary definition: Machine Learning consists of training systems capable of understanding the data entered in order to predict responses or extract useful information from them. It is a subset of artificial intelligence and is closely related to statistics.

In terms of the number of learning techniques that Machine Learning understands, we have found that not all authors speak with one voice on this subject. Some like Jorge Castañón in [6] or Harleen Kaur and Vinita Kumari in [5] distinguish between two types of learning, supervised and unsupervised. Others such as Paul Lanier et al. in [7] consider three types of learning, supervised, unsupervised and semi-supervised. Nirav J. Patel and Rutvij H. Jhaveri in [8] consider reinforcement learning to be the third type by removing the semi-supervised from the list. Abdallah Moujahid et al. in [9] distinguish four types of learning, supervised, unsupervised, reinforcement and deep learning.

In this study, we consider four types of learning: supervised, unsupervised, semi-supervised and reinforcement learning.

Supervised learning is used when historical data is available for a certain problem. The system is trained with the respective inputs and responses, and then used to predict responses for new inputs [5].

Supervised learning is subdivided into two sub-types: classification and regression [10].

Classification involves finding a relationship between discrete inputs and discrete outputs. Output variables are also called categories or labels. A mapping function (classifier) is constructed by analyzing training data in the learning step, and this classifier is adopted to predict categorical class labels in the classification step [10]. Regression, on the other hand, involves estimating or predicting continuous quantities. Regression relies on input statistical characteristics to establish the relationship between two or more independent variables, [10].

Unsupervised learning, unlike supervised learning, does not come with labels (no output vectors). The objective of unsupervised learning is to analyze the structure of the data and extract useful information from it without any explicit indication of the expected result [10].

Unsupervised learning includes two sub-types: clustering and dimensionality reduction [10].

clustering consists of dividing a set of objects into different groups so that the objects in each group are as similar as possible to each other, and the different groups are as different as possible from each other [10]. While dimensionality reduction aims to transform a large data space into a smaller space without losing the useful information from the original data [10].

Semi-supervised learning, as its name suggests, is a hybrid of the two approaches mentioned above. Semi-supervised learning is commonly used when some cases (problems) have values for both covariates (inputs) and outcomes (outputs), but the majority of cases have values only for the covariates and lack data on the expected outcome [7].

Reinforcement learning is a particular area of Machine Learning that is based on taking certain actions followed by numerical rewards to achieve a goal. The important point is that whoever undertakes an action, called an agent, in a particular world, called an environment, does not know which action is good or bad, but he will learn which ones will give the greatest rewards by trying them out [9].

The word, **technique** will be used to refer to a type of learning (supervised, unsupervised, semi-supervised, or reinforcement) [5], [11], [12]. For the term, **method**, it will be used to refer to different methods of Machine Learning: ANN, SVM, DT, etc. [13]. We will retain the nuance between algorithm and model proposed by [13]. A **model** is a set of hypotheses about a problem domain, expressed in a precise mathematical form, which is used to create a Machine Learning solution [13]. Whereas an **algorithm** is simply a set of instructions used to implement a model to solve a problem or perform a calculation.

III. STUDIED MACHINE LEARNING METHODS

This study focuses on the use of Machine Learning (ML) methods in predictive analysis. To do this, we carried out a research in three scientific databases: Science Direct, Springer Link and IEEE Xplore. This search yielded 10 095 papers for the three databases.

We then considered selection criteria in order to limit the number of papers to only those published in the last five years, prioritizing those published in peer-reviewed scientific journals and using the relevance ranking tools provided by the scientific databases.

At the end of the application of these criteria, 246 papers stood out from the batch. Of these 246 papers, 13 duplicates were detected and removed. The remaining 233 papers were read in full and 30 were included in this study. This approach allowed only the most recent publications to be retained. Of the 30 papers, six were published in 2020, twenty-one in 2019 and three in 2018. Table I lists the papers considered in this study, their years of publication and their fields of application.

We can see that ML methods have been used for predictive analysis in several different areas. The papers studied cover five domains: construction and botany for 3.3% each, education for 10%, social science for 26.7% and medicine for 56.7% of the papers.

Medicine being a very broad field, we have listed five specialties in particular: cardiology [14], [15]; oncology [16], [17], [18]; diabetology [11], [5], [19]; psychiatry [20] and pediatric surgery [12], in addition to which it will be necessary to add general medicine which groups together the largest number of papers [21], [22], [14], [20], [23], [24], [25].

The review of these papers identified 23 ML methods used to perform predictive analysis. We then organized these methods by learning type, learning sub-type, and field of application. Table II provides a summary of this organization.

IV. ANALYSIS AND DISCUSSION

As stated above, we distinguish four learning techniques. It was observed during this study that supervised learning is

TABLE I
STUDIED PAPERS BY YEAR AND APPLICATION FIELD

Paper	Application field	Year
Predicting adults likely to develop heart failure using readily available clinical information [15]	Medicine - Cardiology	
Using machine learning to predict opioid misuse among U.S. adolescents [26]	Social Science	
Machine learning models for credit analysis improvements: Predicting low-income families' default [27]	Social Science	
A Mobile Application for Early Prediction of Student Performance Using Fuzzy Logic and Artificial Neural Networks [28]	Education	2020
learning predictive model based on national data for fatal accidents of construction workers [29]	Building	
Testing the convergent- and predictive validity of a multi-dimensional belief-based scale for attitude towards personal safety on public bus/ minibus for long-distance trips in Ghana: A SEM analysis [30]	Social Science	
Predictors of length of stay in the coronary care unit in patient with acute coronary syndrome based on data mining methods [14]	Medicine - Cardiology	
Application of the Albumin-Bilirubin Grade in Predicting the Prognosis of Patients With Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis [16]	Medicine - Oncology	
Computational models for predicting anticancer drug efficacy: A multi linear regression analysis based on molecular, cellular and clinical data of oral squamous cell carcinoma cohort [17]	Medicine - Oncology	
Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study [18]	Medicine - Oncology	
Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review [31]	Botany	
Can we predict lesion detection rates in second-look ultrasound of MRI detected breast lesions? A systematic analysis [21]	General Medicine	
Predictive analytics for hospital admissions from the emergency department using triage information [22]	General Medicine	
A predictive analytics framework for identifying patients at risk of developing multiple medical complications caused by chronic diseases [32]	General Medicine	
Using Machine Learning Applied to Real-World Healthcare Data for Predictive Analytics: An Applied Example in Bariatric Surgery [11]	General Medicine	
Predictive Analytics and Modeling Employing Machine Learning Technology: The Next Step in Data Sharing, Analysis, and Individualized Counseling Explored With a Large, Prospective Prenatal Hydronephrosis Database [12]	Medicine - Pediatric Surgery	2019
Residential demand response program: Predictive analytics, virtual storage model and its optimization [33]	Social Science	
Predicting and explaining corruption across countries: A machine learning Approach [4]	Social Science	
Utilizing early engagement and machine learning to predict student outcomes [34]	Education	
Identifying predictors of probable posttraumatic stress disorder in children and adolescents with earthquake exposure: A longitudinal study using a machine learning approach [35]	Social Science	
Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records [23]	General Medicine	
An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data [24]	General Medicine	
Educational data mining: Predictive analysis of academic performance of public-school students in the capital of Brazil [36]	Education	
Ensemble method based predictive model for analyzing disease datasets: a predictive analysis approach [25]	General Medicine	
A Predictive Analytics-Based Decision Support System for Drug Courts [37]	Social Science	
Preventing Infant Maltreatment with Predictive Analytics: Applying Ethical Principles to Evidence Based Child Welfare Policy [7]	Social Science	
Predictive models for diabetes mellitus using machine learning techniques [19]	Medicine Diabetology	
Predictive modelling and analytics for diabetes using a machine learning Approach [5]	Medicine Diabetology	
Processing electronic medical records to improve predictive analytics outcomes for hospital readmissions [38]	General Medicine	2018
Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review [20]	Medicine - Psychiatry	

the most widely used technique when it comes to predictive analysis. We can see this in Fig. 1, which shows the rates of use of the different learning techniques for the studied papers. 70% of the methods used to perform predictive analysis are supervised learning, followed by unsupervised learning and very little semi-supervised learning. None of the authors used reinforcement learning to make prediction.

This observation gives a first clue in the choice of the type of learning to do predictive analysis.

Another clue can be found in the fields of application of these methods. Fig. 2 shows the areas of application of ML methods for predictive analysis. We can see that medicine is the most used discipline as an application area for ML methods for predictive analysis. But it is not only medicine, there are also social science, followed by construction, then botany and finally education.

It should be noted that the use of a method in a field is not exclusive. A method can be used in several areas. We see for example the ANN which is used in medicine [14], [22], [32], [5], [20], [23], in social science [26], [27], [4] and in education [28]; or the CA which is used both in botany [31] and in social science [33].

TABLE II
ML METHODS BY LEARNING TECHNIQUE, PAPER AND APPLICATION FIELD

Method	Description	Type of learning	Sub-Type	Papers	Application fields
ANN	Artificial Neural Network	Supervised and Unsupervised	Regression and clustering	[14] [22] [32] [26] [27] [28] [4] [5] [20] [23]	Medicine, Social Science, Education
SVM	Support Vector Machine	Supervised	Classification	[14] [32] [4] [5] [20] [22] [27] [25]	Medicine, Social Science
DT	Decision Tree	Supervised	Classification and Regression	[14] [15] [31] [21] [32] [27] [34] [29] [25] [19]	Medicine, Building, Education, Botany
AC	Auto Classifier Node	Supervised	Classification	[14]	Medicine
MLR	Multi Linear Regression	Supervised	Regression	[17]	Medicine
Cox modele		Supervised	Regression	[15] [18]	Medicine
MRMR	Maximum Relevance Minimum Redundancy	Semi-Supervised		[18]	Medicine
RF	Random Forest	Supervised	Classification and Regression	[18] [22] [26] [4] [29] [15] [24] [38] [25] [37] [19]	Medicine, Social Science, Building
PCA	Principal Component Analysis	Unsupervised	Dimentionnality Reduction	[31] [25]	Medicine, Botany
LDA	Linear Discriminant Analysis	Supervised	Classification	[31] [20]	Medicine, Botany
CA	Cluster Analysis	Unsupervised	Clustering	[31] [33]	Social Science, Bontany
XGBoost	eXtreme Gradient Boosting algorithm	Supervised	Classification and Regression	[22] [35]	Medicine, Social Science
LR	Logistic Regression	Supervised	Regression	[22] [32] [11] [29] [20] [37] [19]	Medicine, Building
GBM	Gradient Boosting Machine	Supervised	Regression and classification	[26] [20] [36] [19]	Medicine
Fuzzy		Supervised and Unsupervised	Classification and Clustering	[28]	Education
RBF	Radial Basis Function	Supervised and Unsupervised	Classification and Clustering	[5]	Medicine
K-NN	K-Nearest Neighbour	Supervised	Classification	[5] [25]	Medicine
Adaboost		Supervised	Classification	[29]	Building
K-Means	Clustering	Unsupervised	Clustering	[23]	Medicine
NB	Naïve Bayes	Supervised	Classification	[25]	Medicine
PR	Poisson Regression	Supervised	Regression	[37]	Social Science
OLSR	Ordinary Least Squares Regression	Supervised	Regression	[37]	Social Science
Bayesian methods		Supervised and Unsupervised	Regression and Clustering	[7]	Social Science

It is therefore important to have an overview of the most commonly used ML methods for predictive analysis, all fields combined. For this we have Fig. 3.

From Fig. 3, it is clear that RF is the most widely used ML method for predictive analysis across all domains, followed by ANN and DT, then SVM and then LR and the others. That said, although RF is the most widely used method, it is interesting to note its absence in some fields, such as education and botany. The same is true for ANN which is not used in building and botany, DT which is not used in social science and botany, SVM which is not used in building, botany and education, and LR which is not used in education and botany.

Thus, to make a good choice, one must consider the use of methods in each area. For this reason, Fig. 4 shows the usage rate of the five most commonly used ML methods for predictive analysis by application domain. These five methods alone cover 22 of the 30 papers, as well as the five application areas identified in this study.

Fig. 4 shows, for the studied papers in this review, that:

- RF method is the most widely used in medicine;
- ANN and DT are the most used in education;
- DT is the only method used in the only botany paper reviewed in this review;
- RF has the same usage rate as ANN in social science;
- LR, RF and DT have the same rate of use in building.

V. CONCLUSION AND FUTURE WORK

Following our previous research work [39], [40], [41], and in order to select the most appropriate predictive analysis methods to perform predictive analysis for epidemiological diseases, we conducted a literature review.

This study presents a literature review of trends and methods of ML used for predictive analysis. In order to provide a review that reflects the current state of research, we adopted

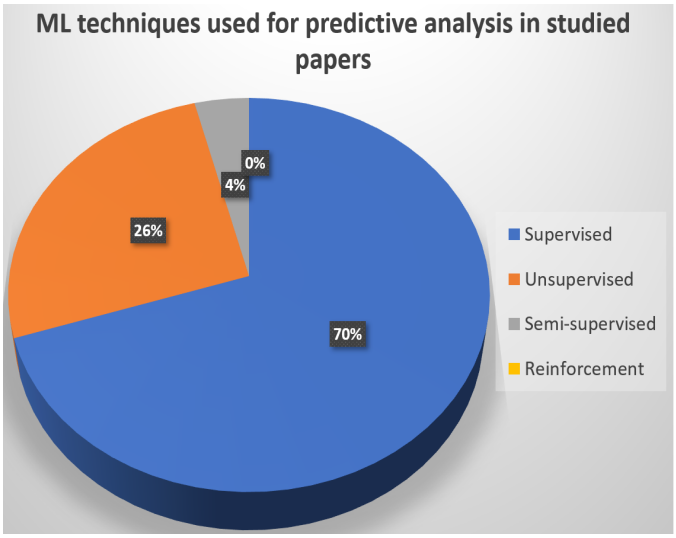

Fig. 1. ML techniques used for predictive analysis in studied papers

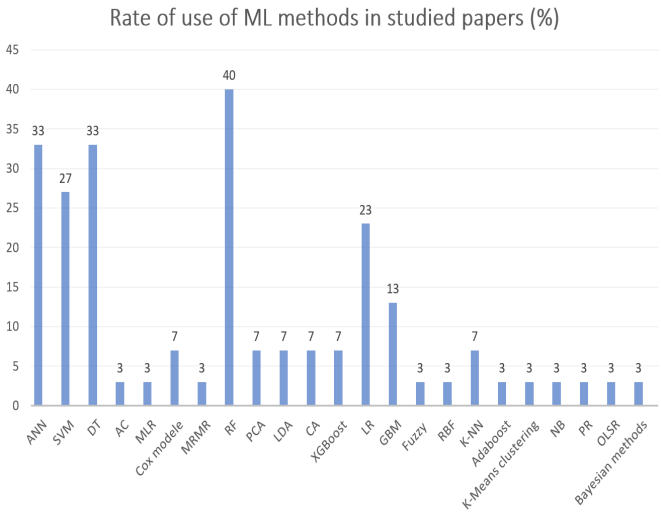

Fig. 3. Rate of use of ML methods in studied papers

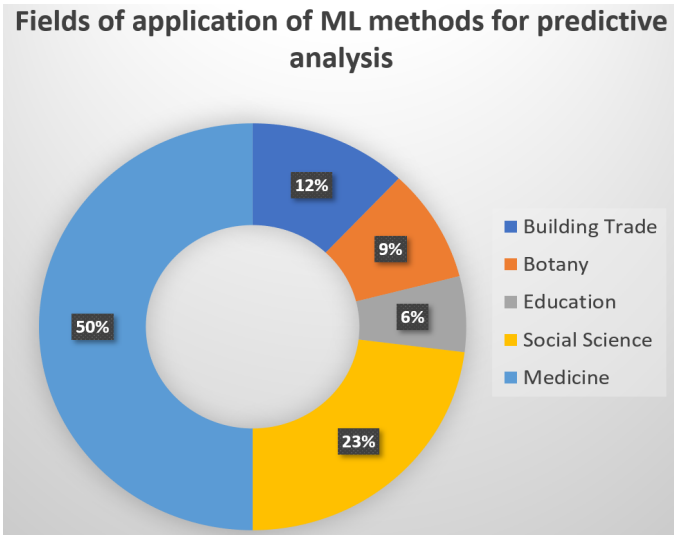

Fig. 2. Fields of application of ML methods for predictive analysis

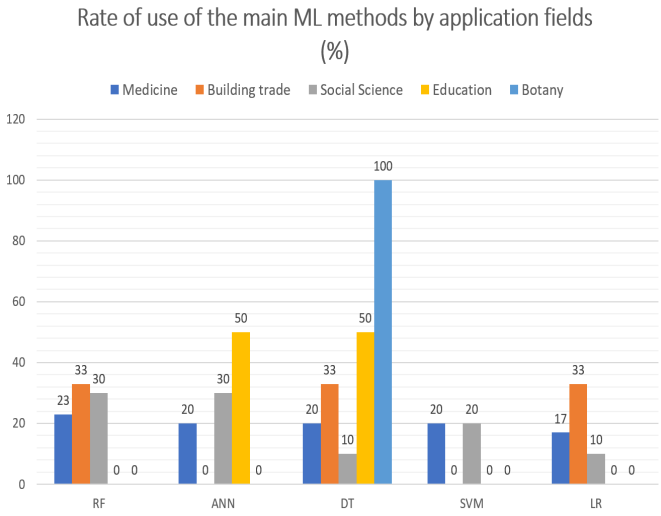

Fig. 4. Rate of use of the main ML methods by application fields

an approach that allowed us to select only the most recent research papers in the literature.

This made it possible to identify: The most used disciplines as application fields of ML methods for predictive analysis: medicine, social science, building, botany and education; The most commonly used learning techniques: Supervised and Unsupervised; The most commonly used ML methods: RF, ANN, DT, SVM and LR; And the most used ML methods by discipline.

The result presented in this way does not dictate which method to choose, but rather what methods other researchers around the world are using in a specific discipline to do predictive analysis. This can serve as an indicator for researchers interested in predictive analysis. It should be note that the use of ML method could depend closely on the specification of the problem regardless of its domain.

For our part, we plan to integrate ML methods into a mobile data collection system to perform predictive analysis for an epidemiological investigation. For example an investigation about covid-19.

REFERENCES

[1] Montreal University: http://www.iro.umontreal.ca/ nie/IFT3335/Introduction.html, accessed on 13 june 2020.

[2] CNRS LE JOURNAL, Yaroslav Pigenet : https://lejournal.cnrs.fr/articles/des-machines-enfin-intelligentes.

[3] Zinebi, K., Souissi, N., and Tikito, K. (2018). Driver Behavior Analysis Methods: Applications oriented study. In Proceedings of the 3rd International Conference on Big Data, Cloud and Application (BDCA 2018), April 2018 in Kenitra Morocco.

[4] Lima, M. S. M., and Delen, D. (2020). Predicting and explaining corruption across countries: A machine learning approach. Government Information Quarterly, 37(1), 101407.

[5] Kaur, H., and Kumari, V. (2018). Predictive modelling and analytics for diabetes using a machine learning approach. Applied computing and informatics.

[6] Castañón, J. (10). Machine Learning Methods that Every Data Scientist Should Know. Consultado em Outubro, 16, 2019.

[7] Lanier, P., Rodriguez, M., Verbiest, S., Bryant, K., Guan, T., and Zolotor, A. (2020). Preventing infant maltreatment with predictive analytics: applying ethical principles to evidence-based child welfare policy. Journal of family violence, 35(1), 1-13.

[8] Patel, N. J., and Jhaveri, R. H. (2015, January). Detecting packet dropping nodes using machine learning techniques in Mobile ad-hoc network: A survey. In 2015 International Conference on Signal Processing and Communication Engineering Systems (pp. 468-472). IEEE.

[9] Moujahid, A., Tantaoui, M. E., Hina, M. D., Soukane, A., Ortalda, A., ElKhadimi, A., and Ramdane-Cherif, A. (2018, June). Machine learning techniques in ADAS: A review. In 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE) (pp. 235-242). IEEE.

[10] Yang, H., Xie, X., and Kadoch, M. (2020). Machine Learning Techniques and A Case Study for Intelligent Wireless Networks. IEEE Network, 34(3), 208-215.

[11] Johnston, S. S., Morton, J. M., Kalsekar, I., Ammann, E. M., Hsiao, C. W., and Reps, J. (2019). Using machine learning applied to real-world healthcare data for predictive analytics: an applied example in bariatric surgery. Value in Health, 22(5), 580-586.

[12] Lorenzo, A. J., Rickard, M., Braga, L. H., Guo, Y., and Oliveria, J. P. (2019). Predictive Analytics and Modeling Employing Machine Learning Technology: The Next Step in Data Sharing, Analysis, and Individualized Counseling Explored With a Large, Prospective Prenatal Hydronephrosis Database. Urology, 123, 204-209.

[13] Model-based Machine Learning: http://www.mbmlbook.com/Introduction.html

[14] Rezaianzadeh, A., Dastoorpoor, M., Sanaei, M., Salehnasab, C., Mohammadi, M. J., and Mousavizadeh, A. (2020). Predictors of length of stay in the coronary care unit in patient with acute coronary syndrome based on data mining methods. Clinical Epidemiology and Global Health, 8(2), 383-388.

[15] Bergsten, T. M., Nicholson, A., Donnino, R., Wang, B., Fang, Y., and Natarajan, S. (2020). Predicting adults likely to develop heart failure using readily available clinical information: An analysis of heart failure incidence using the NHEFS. Preventive Medicine, 130, 105878.

[16] Xu, L., Wu, J., Lu, W., Yang, C., and Liu, H. (2019, December). Application of the Albumin-Bilirubin Grade in Predicting the Prognosis of Patients With Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. In Transplantation Proceedings (Vol. 51, No. 10, pp. 3338-3346). Elsevier.

[17] Robert, B. M., Brindha, G. R., Santhi, B., Kanimozhi, G., and Prasad, N. R. (2019). Computational models for predicting anticancer drug efficacy: A multi linear regression analysis based on molecular, cellular and clinical data of oral squamous cell carcinoma cohort. Computer methods and programs in biomedicine, 178, 105-112.

[18] Ji, G. W., Zhu, F. P., Xu, Q., Wang, K., Wu, M. Y., Tang, W. W., ... and Wang, X. H. (2019). Machine-learning analysis of contrast-enhanced CT radiomics predicts recurrence of hepatocellular carcinoma after resection: A multi-institutional study. EBioMedicine, 50, 156-165.

[19] Lai, H., Huang, H., Keshavjee, K., Guergachi, A., and Gao, X. (2019). Predictive models for diabetes mellitus using machine learning techniques. BMC endocrine disorders, 19(1), 1-9.

[20] Lee, Y., Ragguett, R. M., Mansur, R. B., Boutilier, J. J., Rosenblat, J. D., Trevizol, A., and Chan, T. C. (2018). Applications of machine learning algorithms to predict therapeutic outcomes in depression: a meta-analysis and systematic review. Journal of affective disorders, 241, 519-532.

[21] Bumberger, A., Clauser, P., Kolta, M., Kapetas, P., Bernathova, M., Helbich, T. H., ... and Baltzer, P. A. (2019). Can we predict lesion detection rates in second-look ultrasound of MRI-detected breast lesions? A systematic analysis. European journal of radiology, 113, 96-100.

[22] Araz, O. M., Olson, D., and Ramirez-Nafarrate, A. (2019). Predictive analytics for hospital admissions from the emergency department using triage information. International Journal of Production Economics, 208, 199-207.

[23] Huang, L., Shea, A. L., Qian, H., Masurkar, A., Deng, H., and Liu, D. (2019). Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. Journal of biomedical informatics, 99, 103291.

[24] Hill, B. L., Brown, R., Gabel, E., Rakocz, N., Lee, C., Cannesson, M., and Maoz, U. (2019). An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data. British Journal of Anaesthesia, 123(6), 877-886.

[25] Ramesh, D., and Katheria, Y. S. (2019). Ensemble method based predictive model for analyzing disease datasets: a predictive analysis approach. Health and Technology, 9(4), 533-545.

[26] Han, D. H., Lee, S., and Seo, D. C. (2020). Using machine learning to predict opioid misuse among US adolescents. Preventive medicine, 130, 105886.

[27] de Castro Vieira, J. R., Barboza, F., Sobreiro, V. A., and Kimura, H. (2019). Machine learning models for credit analysis improvements: Predicting low-income families' default. Applied Soft Computing, 83, 105640.

[28] Nosseir, A., and Fathy, Y. (2020). A Mobile Application for Early Prediction of Student Performance Using Fuzzy Logic and Artificial Neural Networks.

[29] Choi, J., Gu, B., Chin, S., and Lee, J. S. (2020). Machine learning predictive model based on national data for fatal accidents of construction workers. Automation in Construction, 110, 102974.

[30] Sam, E. F., Brijs, K., Daniels, S., Brijs, T., and Wets, G. (2020). Testing the convergent-and predictive validity of a multi-dimensional belief-based scale for attitude towards personal safety on public bus/minibus for long-distance trips in Ghana: A SEM analysis. Transport policy, 85, 67-79.

[31] Maione, C., Barbosa Jr, F., and Barbosa, R. M. (2019). Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: A review. Computers and Electronics in Agriculture, 157, 436-446.

[32] Talaei-Khoei, A., Tavana, M., and Wilson, J. M. (2019). A predictive analytics framework for identifying patients at risk of developing multiple medical complications caused by chronic diseases. Artificial Intelligence in Medicine, 101, 101750.

[33] Basnet, S. M., Aburub, H., and Jewell, W. (2019). Residential demand response program: Predictive analytics, virtual storage model and its optimization. Journal of Energy Storage, 23, 183-194.

[34] Gray, C. C., and Perkins, D. (2019). Utilizing early engagement and machine learning to predict student outcomes. Computers & Education, 131, 22-32.

[35] Ge, F., Li, Y., Yuan, M., Zhang, J., and Zhang, W. (2020). Identifying predictors of probable posttraumatic stress disorder in children and adolescents with earthquake exposure: a longitudinal study using a machine learning approach. Journal of affective disorders, 264, 483-493.

[36] Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., and Van Erven, G. (2019). Educational data mining: Predictive analysis of academic performance of public-school students in the capital of Brazil. Journal of Business Research, 94, 335-343.

[37] Zolbanin, H. M., Delen, D., Crosby, D., and Wright, D. (2019). A Predictive Analytics-Based Decision Support System for Drug Courts. Information Systems Frontiers, 1-20.

[38] Zolbanin, H. M., and Delen, D. (2018). Processing electronic medical records to improve predictive analytics outcomes for hospital readmissions. Decision Support Systems, 112, 98-110.

[39] Loola Bokonda P.,Ouazzani-Touhami K., and Souissi N. (2020) Mobile Data Collection Using Open Data Kit. In: Serrhini M., Silva C., Aljahdali S. (eds) Innovation in Information Systems and Technologies to Support Learning Research. EMENA-ISTL 2019. Learning and Analytics in Intelligent Systems, vol 7. Springer, Cham.

[40] Bokonda, P. L., Ouazzani-Touhami, K., and Souissi, N. (2019). Open Data Kit: Mobile Data Collection Framework For Developing Countries. 8. 4749-4754. 10.35940/ijitee.L3583.1081219.

[41] Bokonda, P. L., Ouazzani-Touhami, K., and Souissi, N. (2020). A Practical Analysis of Mobile Data Collection Apps. International Journal of Interactive Mobile Technologies (iJIM), Vol. 14, No. 13, 2020. In press