

Machine Learning in Public Health: A Review

Md. Asadullah¹, Mamunar Rashid², Priyanka Bosu³, Emon Ahmed⁴ and Sabeha Tamanna⁵

¹ Bangabandhu Sheikh Mujibur Rahman Science and Technology University

Received: 13 June 2021 Accepted: 1 July 2021 Published: 15 July 2021

Abstract

In recent years Machine learning has been used for disease diagnosis and prediction in the public healthcare sector. It plays an essential role in healthcare and is rapidly being applied to education. It is one of the driving forces in science and technology, but the emergence of big data involves paradigm shifts in the implementation of machine learning techniques from traditional methods. Computers are now well equipped to diagnose many health issues with large health care datasets and progressions in machine learning techniques. Researchers have been used several machine learning techniques in public health. Several methods, including Support Vector Machines (SVM), Decision Trees (DT), Naïve Bayes (NB), Random Forest (RF), and K-Nearest Neighbors (KNN), are widely used in predictive model design research, resulting in effective and accurate decision-making. The predictive models discussed here are based on different supervised ML techniques and various input characteristics and data samples. Therefore, the predictive models can be used to support healthcare professionals and patients globally to improve public health as well as global health. Finally, we provide some problems and challenges which face the researcher in public health.

Index terms— machine learning, prediction, classification, public health, disease

1 Introduction

Machine learning, a method of developing a prototype that learns to enhance its quality through experience, belongs to the context of artificial intelligence and is increasingly being used in various fields of science [1]. Such algorithms can be applied to help track the progress of a person, what variables make their symptoms worse, predict how long they would take etc. [2]. It is likely to deliver technically superior results, but it is not going to be perfect. While machine learning can deliver technical performance, inequities can be compounded [3]. The intervention was particularly among the group with a moderate likelihood of participation. Targeting the results of the prediction model using the machine-learning method has been identifying suitable intervention targets [4]. Traditional machine-learning approaches have been successful because the complexity of molecular interactions has been reduced by investigating only one or two dimensions of the molecular structure in the feature descriptors. Several different ML classifiers are experimentally validated into the data set in the present study [5]. Machine learning is involved in many of these, but streaming data is only addressed in a few plays. The machine learning library consists of common learning algorithms such as classification, clustering, collaborative sorting, etc. useful when dealing with problems with machine learning [6].

Machine learning typically extends these methods to cope with high dimensionality and nonlinearity, which in wearable sensor data is of particular importance. It overlaps with artificial intelligence, but traditional biomedical statistics usually recognize the problems it seeks to solve. Extraction of the function renders machine-learning traceable because it reduces the number of data dimensions [7]. These techniques can help enhance the ability to discriminate by combining multiple metabolites' predictive abilities. However, these methods are monitor, and therefore, various validations are key factors in preventing over fitting [8]. In this paper, a new approach is

proposed to automatically identify fund us objects. The method uses pre-processing techniques for images, and data to improve the performance of classifiers for machine learning [9]. Machine learning techniques are applied to these data, which are useful for data analysis and are used in specific fields [10].

Recently it can use to analyze medical data, and for medical diagnosis to identify various complex diagnostic problems. We can improve the accuracy, speed, reliability, and performance of the diagnosis on the current system by using machine learning classification algorithms for any particular disease [11]. It is used to estimate vegetation parameters and to detect disease, with less consideration give to the effects of disease symptoms on their performance [12].

II.

Machine Learning in Public Health

Machine learning plays a role in the healthcare field and it is rapidly apply to healthcare, including segmentation of medical images, authentication of images, a fusion of multimodal images, computer-aided diagnosis, image-guided therapy, image classification, and retrieval of image databases, where failure could be fatal [15]. Statistical models developed using machinelearning methods can view in many ways as extensions from epidemiology and health econometrics of more conventional health services research methodologies [16]. Given the wide availability of free packages to support this work, many researchers have been encouraged to apply deep learning to any data mining, and pattern recognition topic related to health informatics [17]. In medical fields, machine learning has also shown promise when the aim is to discover clusters in the data, such as therapeutic choice imaging research. Here, the new features can check with a radiologist or neurologist expert assessment which varies from the prediction environment where observed marks exist in the data [18]. Screening and prognosis of patients with cancer use methods for pattern recognition and identification such as machine learning [19].

The repository should highlight the specifications of clinical machine learning tasks and motivate the ML community by providing a platform for the publication, collection of data sets, benchmarking of statistical evaluators, and methods for challenging machine learning problems [19]. The main of applying the classification method is to allow healthcare organizations to provide accurate medication quantities [20]. At every stage of development and application of machine learning in advancing health, ethical design thinking is essential. To this end, honesty and innovation physicians will work closely with software and data scientists to re-imagine clinical medicine and foresee its ethical implications. It is crucial that data from mobile health and consumer-facing technologies be systematically validated, especially when dynamic intervention is provided [22]. Three developments in machine learning may be of interest to public health researchers and practitioners [25]. Machine Learning techniques have shown success in the prediction and diagnosis of numerous critical diseases. Some sets of features use in this strategy to represent each instance in any dataset [26]. Research comparing the quality of different prediction methods to predict disease, disease etiology, or disease subtype is minimal. For many types of medical diagnoses, a good machine learning approach to classification will apply [27].

III.

Challenges in Public Health

Overall, health systems face multiple challenges: rising disease burden, multimorbidity, and disability drove by aging and epidemiological transition, increased demand for health services, higher social expectations, and increased health spending [3]. Healthcare offers unique machine learning challenges where the requirements for explaining ability, model fidelity, and performance, in general are much higher than in most other fields. Ethical, legal, and regulatory challenges are unique to health care since health care decisions can have an immediate impact on a person's well-being or even life [28]. The primary focus in health informatics is on computational aspects of big data, including challenges, current Big Data Mining techniques, strengths and limitations of works, and an outline of directions for future work. A challenge is pose the high volume of healthcare data, the need for flexible processing, and support for decentralized queries across multiple data sources. Global health as an approach to the current situation and challenges, and the use of digital health as an ideal way to address health challenges associated with conflict-affected environments [30]. There are several ways in which the proposed models of machine learning can help address public health challenges. The regularity, reliability, and granularity of available data is a challenge in tracking population health. Model estimates can play a role in strategic decision-making if they can achieve sufficient precision, and machine learning models can provide a route to this required level of precision [31]. Several writers describe different challenges in public health.

Table 1: Public health Challenges

Challenges

Description	References	Development	Challenges
In the acquisition of talent and growth capital	[64]	Data schema	
Increasing the burden of disease, multimorbidity and disability driven by aging and epidemiological transition	[3]		
Ethics, laws and regulations			

Health care choices can have an immediate impact on a person’s well-being or even life. [28] Epidemic Social health inequalities, a small number of local healthcare professionals, and a weak infrastructure for healthcare. [13] Big Data Data mining methods, advantages and weaknesses of current works and recommendations for future work [29][32][19] [61] Treatment effect Treatment of patient outcomes in order to select the correct treatment [16] Clinical Data Real clinical information environment, incomplete and erroneous data. [65]

8 Data regularity, timing and reliability

The regularity, pacing and granularity of available data is the control of population health. [31] Characteristics identifying

The features of communities, ecosystems and policies are defined in population health [47] Health Tackling Health as an approach to the existing situation and challenges. [30] Dataset imbalance Forming an ensemble of multiple models with matched numbers of positive and negative slides trained on data subsets. [66] Biomarkers identify Build diagnostic, prognostic or guided therapy predictive models [19][59]

9 Screening

The area of early detection of cancer is packed with highlighting cautionary tales. [67]

10 Diagnosis

The nuanced essence of the disease and its patient heterogeneity [68][69] Image data Modern imaging technology will surpass the capabilities of human pattern recognition [70] Diagnosis, treatment and monitoring The growing number of patient data in the form of medical images [9] Decision making Prediction of disease is one of the most important medical problems because it is one of the leading causes of death.. [26] Monitoring of disease The progression of the disease and the estimate of the level of fibrosis of the patient [71] High-dimension image data Imaging evidence was a problem in the treatment of diseases based on brain imaging. [72] Accurate prediction Things that recur within a binary outcome [69] IV.

11 Problem Statement

In public health, they are reducing constraints such as lack of resources (human and logistic) in healthcare centers, high population dispersion, and lack of infrastructure. One problem with the concept of ”data health” is the lack of a practical idea of effective and efficient healthcare programs: each insurer has sought effective strategies through trials and errors [4].The main problem is the unstructured of the medical reports. High complexity and noise issues result from the multisource and multimodal nature of healthcare data. Additionally, the high-volume data also has problems with impurity and missing values. These issues are to handle in terms of both size and reliability, although a range of methods has developed to improve data accuracy and usability [29].Machine learning methods are the leading option for achieving a better result in classification and prediction problems. In a wide range of machine learning (ML) problems, classification plays a role. Another major issue with the collection of data is the potential lack of label accuracy. Overfitting is a potential problem in machine learning. The general problem is that several existing datasets are difficult to use in terms of permission [34]. Table 2 displays the numerous public health issues facing them.

12 Clinical unstructured notes

The multisource and multimodality of health care data leads to high complexity and noise problems [29] Impurity and missing The high-volume data also has problems with impurity and missing values [29] Missing variables This results in the normal multivariate methods, while [16] machine-learning approaches can still be appealing for other reasons

13 Prediction

The computer is equipped with a set of data to improve the classification model after it can be used for future predictions [33][82]

14 Mobility

The problem of visual, hearing, flexibility also affects the disease. [50] Dose management Use machine learning approaches to the SCD drug problem [20] Segmentation That pixels can be marked as belonging to a particular segment or category ??

15 Missing data and Class imbalance

For the context, the performance metrics selected are most often inappropriate. [46] V.

16 Dataset

To generate the most effective results, machine learning algorithms use to analyze data repeatedly. Machine learning currently provides the machine for scrutinizing imaginative information. Today, medical clinics very well equippe with fully automatic machines, and these machines produce tremendous amounts of data, then collect, and exchange these data with information systems or doctors to take the necessary steps. Machine learning methods can used to examine medical data and various technical diagnostic conditions in medical diagnosis. Using machine learning, systems take patient data as an input such as symptoms, laboratory data, and some of the at tributes and produce reliable diagnostic results. Depending on the reliability of the test, the computer must determine the information for the future reference will be used as a learning and qualified dataset [11]. Different Authors are used to several data determine the quality of the proposed classifiers which display as. ILPD Data [83][45]

Cleveland Data

NHA-NES Data [121] VI.

17 Classification Technique

In many real-world issues, classification is one of the most decision-making techniques. The higher number of samples selected for many classification problems, but does not lead to higher classification accuracy [35]. Supervised machine-learning algorithms are mainly use for classification or regression issues where the patient sample class label is already available [19]. Classification tasks are found in a various decisionmaking tasks in various fields such as medicine, science, industry, etc. Several approaches are suggest in the literature on how to solve classification problems [5]. In the medical context, the identification quality of commonly used machine learning models, including k-Nearest Neighbors, Nave Bayes, Decision Tree, Random Forest, Support Vector Machine, and Logistic Regression [36]. In this research paper, we conclude various research papers in a tabular form showing different methodologies and compare the accuracy VII.

18 Cross-Validation Technique

The predictive performance of the models is evaluated using the Cross-Validation technique to estimate how each model performs outside the sample to a new dataset also identified as to test data. The reason for using the cross-validation techniques is to fit it into a training dataset when we fit a model [33].Crossvalidation was applied to achieve the best results to measure the numerical performance of a learning operator [10]. This was not achieved to properly isolate and compare the performance of the different methods concering the weighting of the propensity score. Through several steps, we measured the quality of the various propensity score matching methods [53].The classifier's accuracy calculation is the average accuracy of k-folds. Subsampling is done in bootstrap validation with equivalent substitution from the training dataset [59]. Effective use of the 10-fold cross-validation was found to be a good and reasonable compromise between offering accurate performance estimates and being computationally feasible and preventing overfitting [57].

19 Model Evaluation Technique

After the estimation, the performance of the predictive models is evaluate in terms of accuracy, accuracy, and recall of unseen data using the k-fold cross-validation technique to test their abilities [33]. Classification performance is evaluating the precision, sensitivity, and specificity of each system as it is a widely accepted tool of classification performance evaluation and generalization error estimation [60]. It is mention that the F1 score can be affect distorted class ratios when used as a quality indicator. Both AUC and F1 scores are compared using paired t-tests to updated Bonferroni inference thresholds [59].Here we can summarize different methods of performance evaluation as below, IX.

20 Limitations

While the application of machine learning approaches to healthcare problems is unavoidable given the complexity of processing massive amounts of data, the need to standardize standards of interpretable ML in this field is critical [28]. Although very broad, these data sets can also be very limited (e.g., system data can only be accessible for a small subset of individuals). Several machine learning methods effectively address these restriction but are still subject to the usual sources of bias commonly found in experimental studies [62]. The limitation of using SVM is its interpretation, computational costs for larger datasets, and SVM is essentially a binary classifier. A Simplified decision tree with four attributes for a multi-class decision problem [16]. A model that is overfitted is more complicated than the data can explain. For a genuine disease-related structure, an overfitted model may have too many free parameters and thus risk confusing random noise or another confounding in the training data. This is a pervasive problem in numerical machine learning because it is often possible to set the complexity of the model as high as required to achieve arbitrarily high prediction accuracy [7]. Some limitations of traditional medical scoring systems are the presence of the input set of intrinsic linear combinations of variables, therefore they are not able to model complex nonlinear interactions in medical domains. In this study, this weakness is addressed

by using classification models that can implicitly detect complex nonlinear associations between independent and dependent variables as well as the ability to identify any potential correlations between predictor variables [63].
X.

21 Conclusion

To inform clinicians and policymakers, systems powered by machine learning will have to deliver results of interest in action through clinical trials or real-world performance observations. Eventually, classification approaches such as clustering and artificial neural networks would require a complete set of experiments. Most of the researchers used the traditional machine learning algorithm to analyze public health data like SVM, RF, NB, LR, NN, KNN, ANN, and DT and 10-fold cross-validation provide the better results. But in public health, the challenge is pose the high volume of healthcare data. As a result, the challenge in public health to handle big data. Besides, there are a lot of public health researchers facing problems. Most of that found in a different research paper are classification problems in public health data. Overfitting and data imbalances are problems in public health. In our review paper, we find some challenges which keep in mind every public health researcher because most of the research paper discussed these problems, and most of the researchers have faced these problems.

2

Problem	Description	References
Classification	The situation was linear in nature for all armed and unarmed group datasets	[57][33][42][50][19][5][21][55][73][74][60][75][76][77][69]
Scalability	Exists with two of the most widely used interpretable machine learning models	[28]
Lack of infrastructure	Lack of resources in health care centers (human and logistic), high population dispersion	[13]
Effective and Efficient	Through trial and error, every insurer tried effective strategies	[4][78]
Exchange health information securely	Scientists and clinicians across institutional, across a given healthcare organization. provincial, or even national jurisdictional boundaries	[29]
Overfitting	Because of its storage limitation, it may not be appropriate for very large datasets with high dimensional features	[29][32][17][34][24][79][80]
Data Imbalanced	Which are commonly used to resolve big data clinical databases.	[29][81][27][82][83]

Figure 1: Table 2 :

¹© 2021 Global Journals
²Machine Learning in Public Health: A Review
³() F © 2021 Global Journals Machine Learning in Public Health: A Review

3

Data Description	References
Patient data	[2]
Parkinson's Disease Data	[35]
Clinical Data	[3][87][88][55][24][89][10][71][90][91][38][92]
RGB-D Data	[8]
Diabetes Data	[54][93][52][81][23][36][86][43]
Malaria Data	[1]
TB Data	[13]
Health Data	[4][29]
Biomedical Data	[94]
Heart Disease Data	[62][95][48]
EMR data	[16]
Chronic Disease Data	[96]
Breast Cancer Data	[33][42][46]
Stress Data	[65]
S1,BRFSS & ACS Data	[31]

Figure 2: Table 3 :

4

Technique	Disease Name	Highest Accuracy	References
SVM,RF,KNN,DT	Parkinson's	SVM=97.22%	[2]
NB,KNN,C4,5DT,RF,SVM	Liver Disease	KNN=98.6%	[57]
LR,Adaboost,SVM,DT,	DB	SVM=94.4%	[54]
SVM,ANN	Malaria	SVM=89%	[63]
DNN	Diabetes	DNN=83.67%	[62]
MLP, KNN, CART, SVM,	Breast cancer	MLP=96.70%	[33]
NB.			

Figure 3: Table 4 :

5

Disease Name	Validation Methods	References
Parkinson Disease	10 fold	[35]
Liver Disease	10 fold	[57][45]
Diabetes Disease	10 fold	[54][52][23]
Malaria Disease	5 fold	[1]
Heart Disease	5 fold	[7]
Breast cancer Disease	10 fold	[33][73][115]

Figure 4: Table 5 :

6

Performance Evaluation Method	References
Specificity, Sensitivity, F-Measure, Accuracy, Precision, Cohen-Kappa Statistic	[8]
RMSE,ROC	[10]
Specificity, Sensitivity, F-Score,Accuracy,ROC,K-S Test	[18]
Accuracy, Precision, Recall	[24]
MSE,MAE,NMSE	[44]
RMSE,RSE,RAE,MSE	[45]
Specificity, Sensitivity,PLR,NLR,DP.PPV,NPV	[57]
Specificity, Sensitivity, Accuracy	[60]
AUC, Specificity, Sensitivity,F1-Score,Precision,Recall	[61]
Specificity, Sensitivity,ROC	[63][68]
Fishers Exact Test	[64]
ROC	[69][115][98][75]
Accuracy, Recall, Precision, TP rate, Precision, NPV, FP rate, RME,	[70]
F1-measure, G-mean	
MAE,RAE,RMSE,RSE	[71]
FROC	[74]
Accuracy, Specificity, Sensitivity, Precision, (ROC)	[75]
F1 score, Precision, and Recall,NPV	[81]
DR, Specificity, Sensitivity	[95]
RRSE,Accuracy	[96]
RMSE,R-Square	[114][91]

Figure 5: Table 6 :

[Platelets] , *Platelets* 199 p. .

[Ahmed et al. (2018)] ‘A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms’. M R Ahmed , S H Mahmud , M A Hossin , H Jahan , S R H Noori . *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, 2018. December. IEEE. p. .

[Bataineh ()] ‘A comparative analysis of nonlinear machine learning algorithms for breast cancer detection’. A A Bataineh . *International Journal of Machine Learning and Computing* 2019. 9 (3) p. .

[Chandel et al. ()] ‘A comparative study on thyroid disease detection using K-nearest neighbor and Naive Bayes classification techniques’. K Chandel , V Kunwar , S Sabitha , T Choudhury , S Mukherjee . *CSI Transactions on ICT* 2016. 4 (2-4) p. .

[Padmaja et al. ()] ‘A Machine Learning Approach for Stress Detection using a Wireless Physical Activity Tracker’. B Padmaja , V R Prasad , K V Sunitha . *Int. J. Mach. Learn. Comput* 2018. 8 p. .

[S K and P ()] ‘A Machine Learning Ensemble Classifier for Early Prediction of Diabetic Retinopathy’. S S K , A & P . *Journal of Medical Systems* 2017. 41 (12) p. .

[Mannini et al. ()] ‘A Machine Learning Framework for Gait Classification Using Inertial Sensors: Application to Elderly, Post-Stroke and Huntington’s Disease Patients’. A Mannini , D Trojaniello , A Cereatti , A Sabatini . *Sensors* 2016. 16 (1) p. .

[Abdelaziz et al. ()] ‘A machine learning model for improving healthcare services on cloud computing environment’. A Abdelaziz , M Elhoseny , A S Salama , A M &riad . *Measurement* 2018. 119 p. .

[Abdelaziz et al. ()] ‘A machine learning model for improving healthcare services on cloud computing environment’. A Abdelaziz , M Elhoseny , A S Salama , A M &riad . *Measurement* 2018. 119 p. .

[Zheng et al. ()] ‘A machine learning-based framework to identify type 2 diabetes through electronic health records’. T Zheng , W Xie , L Xu , X He , Y Zhang , M You , Y Chen . *International Journal of Medical Informatics* 2017. 97 p. .

[Gatos et al. ()] ‘A Machine-Learning Algorithm Toward Color Analysis for Chronic Liver Disease Classification, Employing Ultrasound Shear Wave Elastography’. I Gatos , S Tsantis , S Spiliopoulos , D Karnabatidis , I Theotokas , P Zoumpoulis , G C Kagadis . *Ultrasound in Medicine & Biology* 2017. 43 (9) p. .

[Kumar and Gandhi ()] ‘A novel three-tier Internet of Things architecture with machine learning algorithm for early detection of heart diseases’. P M Kumar , U Gandhi . *Computers & Electrical Engineering* 2018. 65 p. .

[Bind et al. ()] ‘A Survey of Machine Learning Based Approaches for Parkinson Disease Prediction’. Shubham Bind , Arvind Kumar Tiwari , Anil Kumar Sahani . *International Journal of Computer Science and Information Technologies* 2015. 6 (2) p. .

[Thiyagarajan et al. ()] ‘A survey on diabetes mellitus prediction using machine learning techniques’. C Thiagarajan , K A Kumar , A Bharathi . *International Journal of Applied Engineering Research* 2016. 11 (3) p. .

[Choudhury and Gupta ()] ‘A Survey on Medical Diagnosis of Diabetes Using Machine Learning Techniques’. A Choudhury , D Gupta . *Machine Learning and Data Analytics*, 2018. 9 p. . (Recent Developments in)

[Perveen et al. ()] ‘A Systematic Machine Learning Based Approach for the Diagnosis of Non-Alcoholic Fatty Liver Disease Risk and Progression’. S Perveen , M Shahbaz , K Keshavjee , A &guergachi . *Scientific Reports* 2018. 8 (1) p. .

[Nilashi et al. ()] ‘An analytical method for diseases prediction using machine learning techniques’. M Nilashi , O Ibrahim , H Ahmadi , L Shahmoradi . *Computers & Chemical Engineering* 2017. 106 p. .

[Fu et al.] *An Ensemble Machine Learning Model For the Early Detection of sepsis from Clinical Data*, M Fu , J Yuan , M Lu , P Hong , M Zeng .

[Ashourloo et al. ()] ‘An Investigation Into Machine Learning Regression Techniques for the Leaf Rust Disease Detection Using Hyperspectral Measurement’. D Ashourloo , H Aghighi , A A Matkan , M R Mobasheri , A M Rad . *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2016. 9 (9) p. .

[Ashourloo et al. ()] ‘An Investigation Into Machine Learning Regression Techniques for the Leaf Rust Disease Detection Using Hyperspectral Measurement’. D Ashourloo , H Aghighi , A A Matkan , M R Mobasheri , A M Rad . *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2016. 9 (9) p. .

[Lim et al. ()] ‘An unsupervised machine learning model for discovering latent infectious diseases using social media data’. S Lim , C S Tucker , S &kumara . *Journal of Biomedical Informatics* 2017. 66 p. .

[Thottakkara et al. ()] ‘Application of Machine Learning Techniques to High-Dimensional Clinical Data to Forecast Postoperative Complications’. P Thottakkara , T Ozrazgat-Baslanti , B B Hupf , P Rashidi , P Pardalos , P Momcilovic , A Bihorac . *PLOS ONE* 2016. 11 (5) p. .

[Nair et al. ()] ‘Applying spark based machine learning model on streaming big data for health status prediction’. L R Nair , S D Shetty , S D Shetty . *Computers & Electrical Engineering* 2018. 65 p. .

- [Nair et al. ()] ‘Applying spark based machine learning model on streaming big data for health status prediction’. L R Nair , S D Shetty , S D Shetty . *Computers & Electrical Engineering* 2018. 65 p. .
- [Hamoud and Alanazi] ‘Artificial Intelligence and Healthcare’. Salem Hamoud , H Alanazi . *International Journal of Artificial Intelligence and Machine Learning* 1 p. .
- [Basham ()] ‘Artificial Intelligence in Health, Human Service Delivery and Education: A Brief Conceptual Overview’. Randy Basham . *Journal of Health Science* 2019. 7 p. .
- [Panch et al. ()] ‘Artificial intelligence, machine learning and health systems’. T Panch , P Szolovits , R &atun . *Journal of global health* 2018. 8 (2) p. .
- [Panch et al. ()] ‘Artificial intelligence: opportunities and risks for public health’. T Panch , J Pearson-Stuttard , F Greaves , R &atun . *The Lancet Digital Health* 2019. 1 (1) p. .
- [Gargeya and leng ()] ‘Automated Identification of Diabetic Retinopathy Using Deep Learning’. R Gargeya , T &leng . *Ophthalmology* 2017. 124 (7) p. .
- [Beam and kohane ()] ‘Big Data and Machine Learning in Health Care’. A L Beam , I S &kohane . *JAMA* 2018. 319 (13) p. .
- [Mooney and Pejaver ()] ‘Big Data in Public Health: Terminology, Machine Learning, and Privacy’. S J Mooney , V Pejaver . *Annual Review of Public Health* 2018. 39 (1) p. .
- [Mooney and pejaver ()] ‘Big data in public health: terminology, machine learning, and privacy’. S J Mooney , V &pejaver . *Annual review of public health* 2018. 39 p. .
- [Bryan ()] R N Bryan . *Machine Learning Applied to Alzheimer Disease. Radiology*, 2016. 281 p. .
- [Wu et al. ()] ‘Classifier Model Based on Machine Learning Algorithms: Application to Differential Diagnosis of Suspicious Thyroid Nodules via Sonography’. H Wu , Z Deng , B Zhang , Q Liu , J Chen . *American Journal of Roentgenology* 2016. 207 (4) p. .
- [Ko et al. ()] ‘Combining Machine Learning and Nanofluidic Technology To Diagnose Pancreatic Cancer Using Exosomes’. J Ko , N Bhagwat , S S Yee , N Ortiz , A Sahmoud , T Black , D Issadore . *ACS Nano* 2017. 11 (11) p. .
- [Fang et al. ()] ‘Computational health informatics in the big data age: a survey’. R Fang , S Pouyanfar , Y Yang , S C Chen , S S Iyengar . *ACM Computing Surveys (CSUR)* 2016. 49 (1) p. .
- [Muchiri et al.] ‘Concealed Firearm Detection in Male and Female on Video using Machine Learning Classification: A Comparative Study’. H Muchiri , I Ateya , G &wanyembi . *Age (Years)* 20 p. .
- [Masetic and Subasi ()] ‘Congestive heart failure detection using random forest classifier’. Z Masetic , A Subasi . *Computer Methods and Programs in Biomedicine* 2016. 130 p. .
- [Park et al. ()] ‘Construction of environmental risk score beyond standard linear models using machine learning methods: application to metal mixtures, oxidative stress and cardiovascular disease in NHANES’. S K Park , Z Zhao , B Mukherjee . *Environmental Health* 2017. 16 (1) p. .
- [Miaoa and miaoa ()] ‘Coronary Heart Disease Diagnosis using Deep Neural Networks’. K H Miaoa , J H &miaoa . *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS* 2018. 9 (10) p. .
- [Vijayarani and dhayanand ()] ‘Data mining classification algorithms for kidney disease prediction’. S Vijayarani , S &dhayanand . *International Journal on Cybernetics & Informatics (IJCI)* 2015. 4 (4) p. .
- [Suk et al. ()] ‘Deep ensemble learning of sparse regression models for brain disease diagnosis’. H-I Suk , S.-W Lee , D Shen . *Medical Image Analysis* 2017. 37 p. .
- [Ravi et al. ()] ‘Deep learning for health informatics’. D Ravi , C Wong , F Deligianni , M Berthelot , J Andreu-Perez , B Lo , G Z Yang . *IEEE journal of biomedical and health informatics* 2016. 21 (1) p. .
- [Ravi et al. ()] ‘Deep learning for health informatics’. D Ravi , C Wong , F Deligianni , M Berthelot , J Andreu-Perez , B Lo , G Z Yang . *IEEE journal of biomedical and health informatics* 2016. 21 (1) p. .
- [Wang et al. ()] ‘Detecting Cardiovascular Disease from Mammograms With Deep Learning’. J Wang , H Ding , F A Bidgoli , B Zhou , C Iribarren , S Molloy , P &baldi . *IEEE Transactions on Medical Imaging* 2017. 36 (5) p. .
- [Zhang and Wang ()] ‘Detection of Alzheimer’s disease by displacement field and machine learning’. Y Zhang , S Wang . *Peer J* 2015. 3 p. .
- [Guo et al. ()] ‘Developing a dengue forecast model using machine learning: A case study in China’. P Guo , T Liu , Q Zhang , L Wang , J Xiao , Q Zhang , W Ma . *PLOS Neglected Tropical Diseases* 2017. 11 (10) p. .
- [Kim et al. ()] ‘Development of machine learning models for diagnosis of glaucoma’. S J Kim , K J Cho , S Oh . *PLOS ONE* 2017. 12 (5) p. .

[Mercaldo et al. ()] ‘Diabetes mellitus affected patients classification and diagnosis through machine learning techniques’. F Mercaldo , V Nardone , A Santone . *Procedia computer science* 2017. 112 p. .

[Subasi et al. (ed.) ()] *Diagnosis of Chronic Kidney Disease by*, A Subasi , E Alickovic , J Kevric . Using Random Forest. CMBEBIH (ed.) 2017. 2017 p. .

[Ehteshamibejnordi et al. ()] ‘Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer’. B Ehteshamibejnordi , M Veta , P Johannes Van Diest , B Van Ginneken , N Karssemeijer , G Litjens , Balkenhol . *JAMA* 2017. 318 p. .

[Jadhav et al. ()] *Disease Prediction by Machine Learning from Healthcare Communities*, S Jadhav , R Kasar , N Lade , M Patil , S &kolte . 2019. p. .

[Chen et al. ()] ‘Disease Prediction by Machine Learning Over Big Data from Healthcare Communities’. M Chen , Y Hao , K Hwang , L Wang , L Wang . *IEEE Access* 2017. 5 p. .

[So et al. ()] ‘Early Diagnosis of Dementia from Clinical Data by Machine Learning Techniques’. A So , D Hooshyar , K Park , H Lim . *Applied Sciences* 2017. 7 (7) p. 651.

[Ortiz et al. ()] ‘Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer’s Disease’. A Ortiz , J Munilla , J M Górriz , J &ramírez . *International Journal of Neural Systems* 2016. 26 (07) p. .

[Martins ()] ‘Evaluation of diagnostic tests in human health and disease’. I J Martins . *J Clin Path Lab Med* 2018. 2018. 2 (1) p. .

[Bhowal ()] ‘Extensive Study Of Iot in Healthcare Based On Machine Learning And Cloud’. Asmita Bhowal . *International Journal of Innovations in Engineering and Technology (IJJET)* 2019. 12 p. .

[Dipnall et al. ()] ‘Fusing Data Mining, Machine Learning and Traditional Statistics to Detect Biomarkers Associated with Depression’. J F Dipnall , J A Pasco , M Berk , L J Williams , S Dodd , F N Jacka , D Meyer . *PLOS ONE* 2016. 11 (2) p. .

[Abdulhay et al. ()] ‘Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease’. E Abdulhay , N Arunkumar , K Narasimhan , E Vellaippan , V Venkatraman . *Future Generation Computer Systems* 2018. 83 p. .

[Nashif et al. ()] ‘Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System’. S Nashif , Md R Raihan , Md R Islam , M H Imam . *World Journal of Engineering and Technology* 2018. 6 p. .

[Learning ()] ‘Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review’. M Learning . *Advances in Computational Sciences and Technology* 2017. 10 (7) p. .

[Patel et al. ()] ‘Heart disease prediction using machine learning and data mining technique’. J Patel , D Tejalupadhyay , S Patel . *Heart Disease* 2015. 7 (1) p. .

[Mathotaarachchi et al. ()] ‘Identifying incipient dementia individuals using machine learning and amyloid imaging’. S Mathotaarachchi , T A Pascoal , M Shin , A L Benedet , M S Kang , T Beaudry , P Rosa-Neto . *Neurobiology of Aging* 2017. 59 p. .

[Leightley et al. ()] ‘Identifying probable post-traumatic stress disorder: applying supervised machine learning to data from a UK military cohort’. D Leightley , V Williamson , J Darby , N T Fear . *Journal of Mental Health* 2019. 28 (1) p. .

[Shimoda et al. ()] ‘Identifying the Most Appropriate Intervention Targets Using Prediction Model Based on a Machine-Learning Method: A Retrospective Analysis of a Health Promotion Program for Improving Participation in General’. A Shimoda , Y Saito , C Ooe , D Ichikawa , A Igarashi , T Nakayama , . . Oyama , H . *Health Check-Up. EJBI* 2018. 14 (4) p. .

[Mirzaei et al. ()] ‘Imaging and machine learning techniques for diagnosis of Alzheimer’s disease’. G Mirzaei , A Adeli , H Adeli . *Reviews in the Neurosciences* 2016. 27 (8) p. .

[Schubach et al. ()] ‘Imbalance-Aware Machine Learning for Predicting Rare and Common Disease-Associated Non-Coding Variants’. M Schubach , M Re , P N Robinson , G &valentini . *Scientific Reports* 2017. 7 (1) p. .

[Cao et al. (2016)] ‘Improving tuberculosis diagnostics using deep learning and mobile health technologies among resource-poor and marginalized communities’. Y Cao , C Liu , B Liu , M J Brunette , N Zhang , T Sun , W H Curioso . *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies*, 2016. June. 9 p. .

[Ahmad et al. (2018)] ‘Interpretable machine learning in healthcare’. M A Ahmad , C Eckert , A Teredesai . *Proceedings of the 2018 ACM International Conference on Bioinformatics*, (the 2018 ACM International Conference on Bioinformatics) 2018. August. 19 p. .

[Luo et al. ()] ‘Is demography destiny? Application of machine learning techniques to accurately predict population health outcomes from a minimal demographic dataset’. W Luo , T Nguyen , M Nichols , T Tran , S Rana , S Gupta , . . Allender , S . *PloS one* 2015. 10 (5) p. .

- [Johnson et al. ()] K M Johnson , H E Johnson , Y Zhao , D A Dowe , L H Staib . *Scoring of Coronary Artery Disease Characteristics on Coronary CTAngiograms by Using Machine Learning. Radiology, 182061*, 2019. 29 p. .
- [Vijayarani and Dhayanand ()] ‘Liver disease prediction using SVM and Naïve Bayes algorithms’. S Vijayarani , S Dhayanand . *International Journal of Science, Engineering and Technology Research (IJSETR)* 2015. 4 (4) p. .
- [Boursalie et al. ()] ‘M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease’. O Boursalie , R Samavi , T E Doyle . *Procedia Computer Science* 2015. 63 p. .
- [Zhang et al. ()] ‘Machine Learning Algorithms for Risk Prediction of Severe Hand-Foot-Mouth Disease in Children’. B Zhang , X Wan , F Ouyang , Y Dong , D Luo , J Liu , S Zhang . *Scientific Reports* 2017. 7 (1) p. .
- [Kannan and vasanthi ()] ‘Machine Learning Algorithms with ROC Curve for Predicting and Diagnosing the Heart Disease’. R Kannan , V &vasanthi . *Springer Briefs in Applied Sciences and Technology*, 2018. 7 p. .
- [Kavakiotis et al. ()] ‘Machine Learning and Data Mining Methods in Diabetes Research’. I Kavakiotis , O Tsave , A Salifoglou , N Maglaveras , I Vlahavas , I Chouvarda . *Computational and Structural Biotechnology Journal* 2017. 15 p. .
- [Darcy et al. ()] ‘Machine Learning and the Profession of Medicine’. A M Darcy , A K Louie , L W Roberts . *JAMA* 2016. (6) p. .
- [Kourou et al. ()] ‘Machine learning applications in cancer prognosis and prediction’. K Kourou , T P Exarchos , K P Exarchos , M V Karamouzis , D I &fotiadis . *Computational and Structural Biotechnology Journal* 2015. 13 p. .
- [De Bruijne ()] ‘Machine learning approaches in medical image analysis: From detection to diagnosis’. M De Bruijne . *Medical Image Analysis* 2016. 33 p. .
- [Khalaf et al. ()] ‘Machine learning approaches to the application of disease modifying therapy for sickle cell using classification models’. M Khalaf , A J Hussain , R Keight , D Al-Jumeily , P Fergus , R Keenan , P Tso . *Neurocomputing* 2017. 228 p. .
- [Mahmud et al. (2018)] ‘Machine Learning Based Unified Framework for Diabetes Prediction’. S M Mahmud , M A Hossin , M R Ahmed , S R H Noori , M N I Sarkar . *Proceedings of the 2018 International Conference on Big Data Engineering and Technology*, (the 2018 International Conference on Big Data Engineering and Technology) 2018. August. ACM. p. .
- [Jagga and Gupta ()] ‘Machine learning for biomarker identification in cancer researchdevelopments toward its clinical application’. Z Jagga , D Gupta . *Personalized medicine* 2015. 12 (4) p. .
- [Kubota et al. ()] ‘Machine learning for large-scale wearable sensor data in Parkinson’s disease: Concepts, promises, pitfalls, and futures’. K J Kubota , J A Chen , M A Little . *Movement Disorders* 2016. 31 (9) p. .
- [Rose ()] ‘Machine learning for prediction in electronic health data’. S Rose . *JAMA network open* 2018. 1 (4) p. .
- [Hijazi et al. ()] ‘Machine Learning in Cardiac Health Monitoring and Decision Support’. S Hijazi , A Page , B Kantarci , T Soyata . *Computer* 2016. 49 (11) p. .
- [Hijazi et al. ()] ‘Machine Learning in Cardiac Health Monitoring and Decision Support’. S Hijazi , A Page , B Kantarci , T &soyata . *Computer* 2016. 49 (11) p. .
- [Awan et al. ()] ‘Machine learning in heart failure’. S E Awan , F Sohel , F M Sanfilippo , M Bennamoun , G Dwivedi . *Current Opinion in Cardiology* 2017. 32 p. .
- [Samant and Agarwal ()] ‘Machine learning techniques for medical diagnosis of diabetes using iris images’. P Samant , R Agarwal . *Computer Methods and Programs in Biomedicine* 2018. 157 p. .
- [Razia and Rao ()] ‘Machine learning techniques for thyroid disease diagnosis-a review’. S Razia , M N Rao . *Indian J Sci Technol* 2016. 9 (28) p. .
- [Khare et al. ()] ‘Machine Learning Theory and Applications for Healthcare’. A Khare , M Jeon , I K Sethi , B Xu . *Journal of Healthcare Engineering* 2017. 2017 p. .
- [Narula et al. ()] ‘Machine-Learning Algorithms to Automate Morphological and Functional Assessments in 2D Echocardiography’. S Narula , K Shameer , A M Salem Omar , J T Dudley , P P Sengupta . *Journal of the American College of Cardiology* 2016. 68 (21) p. .
- [Sharma et al. ()] ‘Malaria outbreak prediction model using machine learning’. V Sharma , A Kumar , D Lakshmipanat , G Karajkhede . *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)* 2015. 4 (12) p. .

442 [Zia and Khan ()] 'May, Predicting Diabetes in Medical Datasets Using Machine Learning Techniques'. U A Zia
443 , N Khan . *International Journal of Scientific and Engineering Research* 2017. 8 (5) p. .

444 [Subbulakshmi and deepa ()] 'Medical Dataset Classification: A Machine Learning Paradigm Integrating Particle
445 Swarm Optimization with Extreme Learning Machine Classifier'. C V Subbulakshmi , S N &deepa . *The*
446 *Scientific World Journal* 2015. 2015 p. .

447 [Raval et al. ()] 'Medical diagnosis system using machine learning'. D Raval , D Bhatt , M K Kumhar , V Parikh
448 , D &vyas . *International Journal of Computer Science & Communication* 2016. 7 (1) p. .

449 [Lange et al. ()] 'Methodology to develop machine learning algorithms to improve performance in gastrointestinal
450 endoscopy'. T Lange , De , P Halvorsen , M Riegler . *World Journal of Gastroenterology* 2018. 24 p. .

451 [Aravanis et al. ()] 'Next-Generation Sequencing of Circulating Tumor DNA for Early Cancer Detection'. A M
452 Aravanis , M Lee , R D Klausner . *Cell* 2017. 168 (4) p. .

453 [Abdar et al. ()] 'Performance analysis of classification algorithms on early detection of liver disease'. M Abdar
454 , M Zomorodi-Moghadam , R Das , I.-H Ting . *Expert Systems with Applications* 2017. 67 p. .

455 [Crown ()] 'Potential application of machine learning in health outcomes research and some statistical cautions'.
456 W H Crown . *Value in health* 2015. 18 (2) p. .

457 [Abreu et al. ()] 'Predicting Breast Cancer Recurrence Using Machine Learning Techniques'. P H Abreu , M S
458 Santos , M H Abreu , B Andrade , D C Silva . *ACM Computing Surveys* 2016. 49 (3) p. .

459 [Norouzi et al. ()] 'Predicting Renal Failure Progression in Chronic Kidney Disease Using Integrated Intelligent
460 Fuzzy Expert System'. J Norouzi , A Yadollahpour , S A Mirbagheri , M M Mazdeh , S A Hosseini .
461 *Computational and Mathematical Methods in Medicine* 2016. 2016 p. .

462 [Kumar ()] 'Prediction of chronic kidney disease using random forest machine learning algorithm'. M Kumar .
463 *International Journal of Computer Science and Mobile Computing* 2016. 5 (2) p. .

464 [Bhakta and Sau ()] 'Prediction of depression among senior citizens using machine learning classifiers'. I Bhakta
465 , A Sau . *International Journal of Computer Applications* 2016. 144 (7) p. .

466 [Nikhar and karandikar ()] 'Prediction of heart disease using machine learning algorithms'. S Nikhar , A M
467 &karandikar . *International Journal of Advanced Engineering* 2016. 2 (6) p. . (Management and Science)

468 [Dai et al. ()] 'Prediction of hospitalization due to heart diseases by supervised learning methods'. W Dai , T S
469 Brisimi , W G Adams , T Mela , V Saligrama , I C &paschalidis . *International Journal of Medical Informatics*
470 2015. 84 (3) p. .

471 [Betancur et al. ()] 'Prognostic Value of Combined Clinical and Myocardial Perfusion Imaging Data Using
472 Machine Learning'. J Betancur , Y Otaki , M Motwani , M B Fish , M Lemley , D Dey , P J Slomka .
473 *JACC: Cardiovascular Imaging* 2018. 11 (7) p. .

474 [Umesh et al. ()] 'Review of image processing and machine learning techniques for eye disease detection and
475 classification'. L Umesh , M Mrunalini , S Shinde . *International Research Journal of Engineering and*
476 *Technology* 2016. 3 (3) p. .

477 [Levine et al. ()] 'Rise of the machines: advances in deep learning for cancer diagnosis'. A B Levine , C Schlosser
478 , J Grewal , R Coope , S J Jones , S Yip . *Trends in cancer* 2019. 5 p. .

479 [Ozkan and Koklu ()] 'Skin Lesion Classification using Machine Learning Algorithms'. I A Ozkan , M Koklu .
480 *International Journal of Intelligent Systems and Applications in Engineering* 2017. 5 (4) p. .

481 [Stanly and Malar ()] S Stanly , R K Malar . *Earlier Diabetic Retinopathy Detection Using Advanced Pre-*
482 *Processing Methods and SVM Classification*, 2019. 6 p. .

483 [Biccler ()] 'Statistical and machine learning methods for the dynamic prediction of prognosis in haematological
484 malignancies (Doctoral dissertation). J L Biccler . *Aalborg Universitetsforlag* 2019. 2019. 7 p. .

485 [Subha et al. ()] 'Study on Cardiovascular Disease Classification Using Machine Learning Approaches'. R Subha
486 , K Anandakumar , A &bharathi . *International Journal of Applied Engineering Research* 2016. 11 (6) p. .

487 [Fatima and Pasha ()] 'Survey of machine learning algorithms for disease diagnostic'. M Fatima , M Pasha .
488 *Journal of Intelligent Learning Systems and Applications* 2017. 9 (01) p. .

489 [Koyner et al. ()] 'The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model*'.
490 J L Koyner , K A Carey , D P Edelson , M M &churpek . *Critical Care Medicine* 2018. 46 (7) p. .

491 [Novillo-Ortiz et al. ()] 'The role of digital health in supporting the achievement of the Sustainable Development
492 Goals (SDGs)'. D Novillo-Ortiz , H De Fátima Marin , F Saigi-Rubió . *International Journal of Medical*
493 *Informatics* 2018. 114 p. .

494 [Arganda-Carreras et al. ()] 'Trainable Weka Segmentation: a machine learning tool for microscopy pixel
495 classification'. I Arganda-Carreras , V Kaynig , C Rueden , K W Eliceiri , J Schindelin , A Cardona , H
496 Sebastian Seung . *Bioinformatics* 2017. 33 (15) p. .

- 497 [Nakajima et al. ()] ‘Urinary Polyamine Biomarker Panels with Machine-Learning Differentiated Colorectal
498 Cancers, Benign Disease, and Healthy Controls’. T Nakajima , K Katsumata , H Kuwabara , R Soya ,
499 M Enomoto , T Ishizaki , M Sugimoto . *International Journal of Molecular Sciences* 2018. 19 (3) p. .
- 500 [Asri et al. ()] ‘Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis’. H Asri ,
501 H Mousannif , H A Moatassime , T Noel . *Procedia Computer Science* 2016. 83 p. .
- 502 [Valdes-Donoso et al. ()] ‘Using Machine Learning to Predict Swine Movements within a Regional Program to
503 Improve Control of Infectious Diseases in the US’. P Valdes-Donoso , K Vanderwaal , L S Jarvis , S R Wayne
504 , A M Perez . *Frontiers in Veterinary Science* 2017. 4 p. .
- 505 [Abdel-Ilah and ?ahinbegovi? ()] ‘Using machine learning tool in classification of breast cancer’. L Abdel-Ilah ,
506 H ?ahinbegovi? . *CMBEBIH* 2017. 2017 p. .