International Journal of Innovative Research in Engineering
Volume 4, Issue 2 (March-April 2023), PP: 544-546.
www.theijire.com ***ISSN No: 2582-8746***

Machine Learning Algorithms for Healthcare Sector: A Survey

Sukhwinder Kaur[1], Mani Arora[2], Anureet Kaur[3]
[1,2,3]Department of Computer Science and Applications, Khalsa College, Amritsar, India.

How to cite this paper:
Sukhwinder Kaur[1], Mani Arora[2], Anureet Kaur[3], "Machine Learning Algorithms for Healthcare Sector: A Survey", IJIRE-V4I02-544-546.

https://www.doi.org/10.59256/ijire.2023040224

Abstract: *Machine learning algorithms are used in numerous real- world applications. In Healthcare, Artificial intelligence and machine learning is implemented to improve diagnostic systems. Since the healthcare industry handles a lot of data, including patient information, these large amounts of records are challenging for an individual to keep and analyse. Machine learning systems developed a way around this by automatically identifying patterns and projecting outcomes based on data. Machine learning algorithms can effectively handle the enormous amount of medical data. The rapid growth of machine learning, and big data has facilitated the deployment of machine intelligence algorithms in healthcare applications like precise disease diagnosis, novel methods of treatment, remote healthcare monitoring, drug discovery etc. In this review paper, we review several ML algorithms and their applications in the healthcare sector.*

Key Word : *Machine Learning, Support vector machine, Artificial Neural Network, Decision Tree, Naive Bayes*

I.INTRODUCTION

With advancing technology, machine learning is being used in the healthcare sector to diagnose many diseases early. Since healthcare domain deals with massive data, with the help of machine learning systems, the massive amount of medical data is handled efficiently .machine learning systems automatically discover patterns and predict the results on data. ML algorithms have been very useful in extracting, analyzing data and making predictions to know whether a person is having a disease based on the given data sets. Machine learning techniques provide numerous algorithms for data classification and prediction resulting in better decisions with high rate of accuracy [1].

II.MACHINE LEARNING METHODS

Machine learning techniques are based on algorithms – sets of mathematical procedures which describe the relationships between variables. Data plays a big part in machine learning. Machine learning methods classify the data into classes. Machine learning methods work by training an algorithm with a training dataset before applying it to actual datasets. These methods frame data in the form of a hypothetical function (g) that maps input to output. The machine learning algorithm will learn to predict output variable given some input variable.

Ouput variable=g(input variable)

This section briefly describes the most commonly used machine learning algorithms .

Artificial Neural Networks (ANN)

Artificial Neural Network (ANN) is mathematical models based on the operative process of human brain. These are presented as systems of interconnected "neurons" which can be applied to machine learning problem and it can learn mapping from input to output. Neural networks clusters and classifies the data. They group unlabeled data according to similarities among the inputs, and they classify data when they have a labeled dataset to train on. When the relationships between inputs and outputs are complex i.e. nonlinear then they are modeled with ANN.

Support Vector Machine

It is a discriminative classifier in which large amount of training dataset is required. It defines separating hyper plane which categorizes new dataset . This hyper plane divides a plane into two halves in two dimensions, with each class located on one side.

Here original training datasets are mapped by a kernel function. After mapping the mapped outputs are linearly separable. Previously SVMs are used for binary classification, but now SVM can perform a multiclass classification.

Naive Bayes

The Naive Bayes classifier is a probabilistic classifier used in most applications of supervised machine learning .It works by figuring out the probability of different attributes of the data being associated with a certain classIt is based on a Bayes theorem of probability for making a probabilistic model of data. It is simple, easy to understand, used when size of training set is less.

$$P(X/Y) = P(Y/X)\ P(X)/P(Y)$$

This fundamentally states the probability of X given that Y is true equals the probability of Y given that X is true times the probability of X being true, divided by the probability of Y being true.

Maximum entropy Classifier

The Max Entropy classifier is a probabilistic classifier belonging to the class of exponential models, it takes into account that the features are not conditionally independent of each other. It selects the model with maximum entropy from all models that fits training data.

Decision Tree

The decision tree (DT) classifier represents the training data in hierarchical form. The tree is composed of nodes representing the input values, edges represent possible moves and path from node to leaf gives the target values from which classification can be created.

Random Forest

Random Forest is a bunch of Decision Trees combined. They Random Forest algorithms are often used for each classification and regression tasks and can handle categorical features very well.This algorithm can handle large number of training examples and high dimensional spaces . The random forest technique is applied in problems in which many trees are formed and trained by dividing the training sets, and outcomes are generated through aggregation of all trees.

K-nearest neighbor

(KNN)The K–nearest neighbor algorithm is based on the principle that the classification of an instance will be close to its similar instances (neighbors) and therefore It keeps data in training and separates data continuously when it receives new data by searching through entire training set for K most similar instances.

III. LITERATURE SURVEY

In daily life people have severe physical disorders and psychological stress due to a many internal and external factors leading to depression, therefore In the healthcare it is necessary to handle this situation, Baek et al. [2] used deep neural networks to predict depression.

Qiong Bai, [3] trained and tested five machine learning algorithms (logistic regression, naive Bayes, random forest, decision tree, and Knearest neighbors) and then compared the performance of each model with Kidney Failure Risk Equation (KFRE).Three ML models, including logistic regression, naive Bayes, and random forest, showed similarity and high sensitivity compared to KFRE.

Gazi Mohammed Ifraz [4] used three machine learning techniques(logistic regression , decision tree classification, and k-nearest neighbor) to predict Chronic kidney disease. Logistic regression technique outperformed the others.

YongFeng Wang. [5] for the classification of thyroid nodules from ultrasound images ,authors compared the performance of radiomics and deep learning based approaches, by dividing dataset into 80% training and 20% testing data. Results showed that deep learning-based methods outperformed radiomics in terms of performance.

In case of diabetes, Machine Learning plays a crucial role in the diagnosis process. Using Machine Learning algorithms for diagnosis of diabetes can give quick and accurate results. Here authors used machine learning algorithms(Decision Tree, ANN, Naive Bayes and SVM algorithms) to predict the diabetic risk level of a patient with a better accuracy and obtained precision of 85% for decision tree, Naive Bayes 77% and 77.3% for Support Vector Machine[6] .

Diabetes that is not recognized and treated can cause damage to the heart, kidneys, eyes, nerves, blood vessels, and many other organs.. Deepti Sisodia et. al., [7] experimentally used three classification algorithms(Naive Bayes, Decision Tree and SVM) to detect diabetes . The performances of algorithms are evaluated using Weka tool and compared using various statistical measures namely precision, recall and f measure. Naive Bayes gave better performance than other algorithms.

In this paper authors calculated accuracy of machine learning algorithms(k-nearest neighbor, decision tree, linear regression and support vector machine) for predicting heart disease and used UCI repository dataset for training and testing[8].

Qin et. al.[9] proposed machine learning methodology for diagnosis of chronic kidney disease. They used KNN method for filling the missing values in dataset. To create Models they used six machine learning algorithms (logistic regression, random forest, support vector machine, k-nearest neighbor, naive Bayes classifier and feed forward neural network) ,and concluded that Among these machine learning models, random forest achieved the best performance.

Dilip Kumar Choubey[10] proposed a model for diagnosis of diabetic patients by two stage process. In first stage, features selction is done by using genetic algorithm from Pima Indian diabetes dataset. Second stage, predicts diabetes using multi-layer Perceptron neural network. The proposed method improved the results for diagnosing diabetes.

Vijiya Kumar, et.al. [11] Using the Random Forest approach, this study proposed a method for more accurate early diabetes prediction for a patient. When compared to other algorithms, the accuracy level is higher and The system is capable of predicting the diabetes disease accurately and quickly.

Chronic Kidney Disease is a progressive loss in function of kidneys over a several years of times. G. Kaur et al. [12] used various data mining algorithms in a Hadoop environment to predict chronic kidney disease. Two classifiers KNN (K-Nearest Neighbor) and SVM (Support Vector Machine) are used in the model.

To predict heart disease, A hybrid strategy is used by combining the characteristics of Random Forest and linear model according to Mohan et al. [13]. Maurya et. al.[14] presents a machine learning algorithm that uses a medical test record classification algorithm to suggest suitable meal plans for chronic kidney disease patients. To slow progression of chronic kidney disease, patients are given a recommended diet according to the potassium zone calculated from blood potassium levels.

Table 1

ML techniques and their applications in the healthcare sector

ML techniques	Applications	References
Deep neural networks	predict depression	[2]
logistic regression, naive Bayes, random forest, decision tree, and Knearest neighbors	Kidney disease	[3]
logistic regression , decision tree classification, and k-nearest neighbor	predict kidney disease	[4]
deep learning-based methods	Diagnosing heart disease	[5]
Decision Tree, ANN, Naive Bayes and SVM algorithms	predict Diabetes	[6]
Naive Bayes, Decision Tree and SVM	predict Diabetes	[7]
k-nearest neighbor, decision tree, linear regression and support vector machine)	predict heart disease	[8]
logistic regression, random forest, support vector machine, k-nearest neighbor, naive Bayes classifier and feed forward neural network	Diagnosing kidney disease	[9]
genetic algorithm and multi-layer Perceptron neural network	predict Diabetes	[10]
Random Forest	predict Diabetes	[11]
K-Nearest Neighbor) and SVM (Support Vector Machine)	predict chronic kidney disease	[12]
Hybrid of Random Forest and linear model	predict heart disease	[13]
machine learning algorithm	predict kidney disease	[14]

IV. CONCLUSION

Machine Learning is transforming all spheres of our life, including the healthcare sector. The main objective is data classification and prediction, and machine learning techniques offer a variety of algorithms for these tasks. Machine learning (ML) has been shown to be effective in assisting in making decisions and predictions from the large quantity of data produced by the healthcare industry. This paper discusses various machine learning methods for healthcare sector.

References

[1]. *K. Kalaiselvi, M. Deepika, Machine Learning for Healthcare Diagnostics, in: Machine Learning with Health Care Perspective, Springer, Cham, 2020, pp. 91–105, https://doi.org/10.1007/978-3-030-40850-3*

[2]. *J.W. Baek, K. Chung, Context deep neural network model for predicting depression risk using multiple regression, IEEE Access 21 (8) (2020 Jan)18171–18181, https://doi.org/10.1109/ACCESS.2020.2968393.*

[3]. *Bai, Qiong, et al. "Machine learning to predict end stage kidney disease in chronic kidney disease."Scientific reports 12.1 (2022): 1-8.*

[4]. *Ifraz, G. M., Rashid, M. H., Tazin, T., Bourouis, S., & Khan, M. M. (2021). Comparative Analysis for Prediction of Kidney Disease Using Intelligent Machine Learning Methods.Computational and Mathematical Methods in Medicine, 2021.*

[5]. *YongFeng Wang,(2020). "Comparison Study of Radiomics and Deep-Learning Based Methods for Thyroid Nodules Classification using Ultrasound Images" published on IEEEAccess.*

[6]. *Priyanka Sonar, Prof. K. JayaMalini," DIABETES PREDICTION USING DIFFERENT MACHINE LEARNING APPROACHES", 2019 IEEE ,3rd International Conference on Computing Methodologies and Communication (ICCMC)*

[7]. *DeeptiSisodia, Dilip Singh Sisodia, Prediction of Diabetes using Classification Algorithms, International Conference on Computational Intelligence and Data Science (ICCIDS 2018), Procedia Computer Science 132 (2018) 1578–1585.pp 1578-1585.*

[8]. *ArchanaSingh ,Rakesh Kumar, "Heart Disease Prediction Using Machine Learning Algorithms", 2020 IEEE, International Conference on Electrical and Electronics Engineering (ICE3)*

[9]. *Qin, Jiongming& Chen, Lin & Liu, Yuhua& Liu, Chuanjun&Feng, Changhao& Chen, Bin.(2019). A Machine Learning Methodology for Diagnosing Chronic Kidney Disease. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2963053.*

[10]. *Dilip Kumar Choubey, Sanchita Paul, GA_MLP NN: A Hybrid Intelligent System for Diabetes Disease Diagnosis, I.J. Intelligent Systems and Applications, 2016, 1, 49-59*

[11]. *VijiyaKumar, K., Lavanya, B., Nirmala, I., & Caroline, S. S. (2019). Random Forest Algorithm for the Prediction of Diabetes. 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN). doi:10.1109/icscan.2019.8878802.*

[12]. *G. Kaur and A. Sharma, "Predict chronic kidney disease using data mining algorithms in hadoop," 2017 International Conference on Inventive Computing and Informatics (ICICI), pp. 973-979.*

[13]. *Mohan, S., Thirumalai, C., Srivastava, G.: Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 2019.*

[14]. *Maurya, A., Wable, R., Shinde, R., John, S.B., Jadhav, R., & Dakshayani, R. (2019). Chronic Kidney Disease Prediction and Recommendation of Suitable Diet Plan by using Machine Learning. 2019 International Conference on Nascent Technologies in Engineering (ICNTE), 1-4.*