International Journal of Advanced Technology in Engineering and Science
Vol. No.9, Issue No. 05, May 2021
www.ijates.com
ijates
ISSN 2348 - 7550

A CONCEPTUAL STUDY OF MACHINE LEARNING TECHNIQUES FOR PREDICTION OF DISEASES

Prof.Kshirsagar Sarika, Dr.Sanjay S.Solanki

1(MCA Dept,JSPM's Jayawantrao Sawant college of Engineering, Hadapsar, Pune 28, Maharashtra, India)

2(MCA DEPT,JSPM's,Narhe Campus Pune, Maharashtra, India)

ABSTRACT

In today's world machine learning is the most important and powerful technology.It is a tool which convert the information into knowledge using various algorithms.It will find the model, patterns etc.It can predict the future things using the data for decision making. Intoday's era proportion of diseases highly increasing due to pollution, living of standards and many more reasons.Its data is also available so from that data machine learning techniques predict the no of diseases provided the symptoms or characteristics into dataset.It is very useful techniques for doctors for decision making and also useful for patients.ML promises upgraded accuracy and diagnosis of diseases.This paper studies the analysis of different machine learning algorithm for prediction of heart disease,thyroid, diabetes and breast cancer diseases.It carries prognosis of diseases and decision-making system consequently.

Keywords: *Machine learning,* prediction, diseases, techniques

I. INTRODUCTION

Machine learning is a methodology for devolving computational, mathematical, statically patterns from data. Nowadays it is used in various sectors like healthcare, agriculture, business etc. Machine learning algorithms are organized into taxonomy, based on the desired outcome of the algorithm. Common algorithm types include:

1.1Supervised learning

Where the algorithm generates a function that maps inputs to desired outputs. One standard formulation of the supervised learning task is the classification problem: the learner is required to learn (to approximate the behaviour of) a function which maps a vector into one of several classes by looking at several input-output examples of the function[1].

1.2Unsupervised learning

Which models a set of inputs: labelled examples are not available[1].

1.3Semi-supervised learning

Which combines both labelled and unlabelled examples to generate an appropriate function or classifier[1].

1.4Reinforcement learning

Where the algorithm learns a policy of how to act given an observation of the world. Every action has some impact in the environment, and the environment provides feedback that guides the learning algorithm[1].

1.5Transduction

Similar to supervised learning, but does not explicitly construct a function: instead, tries to predict new outputs based on training inputs, training outputs, and new inputs[1].

1.6Learning to learn

Where the algorithm learns its own inductive bias based on previous experience[1]

Following figure shows types of machine learning according to taxonomy.

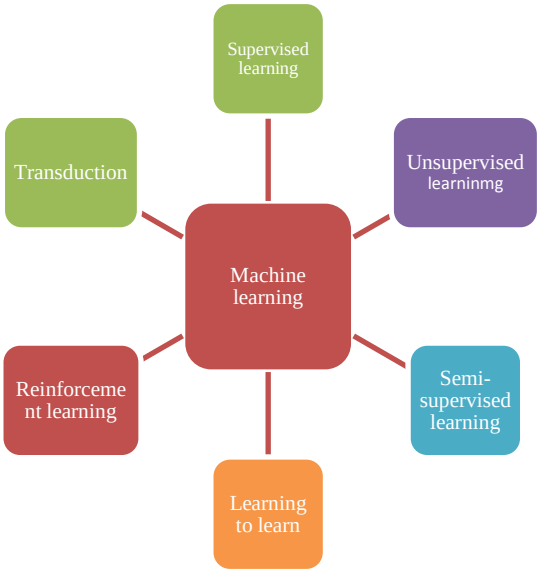

Fig.1.Types of Machine Learning

Machine learning algorithms can also be classified according to the type of computation they can perform on a given dataset. Common types of computation include[1]:

- Classification: learn to put instances into pre-defined classes
- Association: learn relationships between the attributes
- Clustering: discover classes of instances that belong together
- Numeric prediction: learn to predict a numeric quantityinstead of a class

Today, in medical sciences disease diagnostic test is a serious task. It is very important to understand the exact diagnosis of patients by clinical examination and assessment. For effective diagnosis and cost-effective management, decision support systems that are based upon computer may play a vital role. Health care field generates big data about clinical assessment, report regarding patient, cure, follow-ups, medication etc. It is complex to arrange in a suitable way. Quality of the data organization has been affected due to inappropriate management of the data. Enhancement in the amount of data needs some proper means to extract and process data effectively and efficiently[2].The aim of machine learning research in healthcare is not, of course, to replace human doctors or nurses, but rather to supplement and provide support where humans struggle. By doing precisely what human can't, namely processing huge amounts of data quickly, machine learning methods can both improve the quality and consistency of care on a large scale. Additionally, machine learning has promise in aiding more basic research in healthcarerelated fields, such as automated drug discovery, genomics, and computational biology[3].Machine learning helps computers to learn and act accordingly. It helps the

computer to learn the complex model and predict the data and also has the ability to calculate complex mathematics on big data. The machine learning based heart disease predicting systems will be precise and will reduce the risk. The value of machine learning technology is recognized well in health care industry which has large pool of data. It helps medical experts to predict the disease and lead to improvise the treatment. Machine learning predictive models such as decision tree, k-nearest neighbour, logistic regression, random forest, support vector machine are utilized to predict whether a person is having disease or not[4].

II.PREDICTION OF DIFFERENT DISEASES USING DIFFERENT ALGORITHMS

There are many researchers used many algorithms for prediction of diseases.this paper study different diseases heart, breast cancer,thyroid,diabetesdiseases and analysis of the different paper. Following table shows the diseases and algorithm used for prediction.

TABLE I

Sr.No	Name of disease	ML algorithms
1	Thyroid disease	Decision tree,SVM, naïve Bayes
2	Breast cancer disease	Random forest, SVM, KNN, NB, GA, DT
3	Heart disease	Hybrid ML like SVM,RF, DT, KNN
4	Diabetes disease	SVM,Naïve Bayes,RF

1.1Thyroid Disease

Barik et.al.,[5] presented thyroid disease classification using differential evolution with SVM. The results show that the proposed Evolutionary Multivariate Bayesian Prediction classifier model achieves remarkable dimensionality reduction from among the 7200 medical datasets obtained from the UCI repository with 21 attributes (Continuous -15; Discrete - 6). 21 epochs (runs) are carried out for the data and after stabilization, the data are classified as Hyper, Hypo and Normal classes. The results are evaluated based on ten evaluation metrics and the accuracy of classification is 97.97%.

Ioniță et.al.,[6] discussed data mining techniques for prediction of thyroid disease analysed and compared four classification models: Naive Bayes, Decision Tree, Multilayer Perceptron and Radial Basis Function Network. The results indicate a significant accuracy for all the classification models mentioned above, the best classification rate being that of the Decision Tree model (accuracy 96.91%). The data set used to build and to validate the classifier was provided by UCI machine learning repository and by a website with Romanian data. The framework for building and testing the classification models was KNIME Analytics Platform and Weka, two data mining software.

Raghuraman et.al.,[7] presented thyroid disorder blood discovery and prediction using machine learning algorithms.an experiment carried out by SVM,decision tree and naïve Bayes algorithms.The dataset contains the record of 3500+ patients of almost all age groups. It is observed that the Decision Trees accuracy (97.69%) outperformed the SVM accuracy of the network to diagnose the thyroid disease.

From above literature decision tree have high accuracy than other classifiers for prediction of thyroid disease.

1.2Breast Cancer Disease

Asri et.al.,[8]worked on prediction and diagnosis of breast cancer risk using machine learning.They have used DT,SVM,KNN and Naïve Bayes classifier and Wisconsin Breast Cancer (original) datasets is used as data for ML. Experiment conducted on WEKA Tools.SVM gives highest accuracy (97.13%) with lower error rate.

Kumar et al.,[9] discussed an efficient prediction of breast cancer data using data mining technique. For thus study they used WBC dataset. Comparing six classification technologies in WEKAsoftware. According to accuracy the most accurate result found by SVM classifier (accuracy 97.59%) with low error rate2.41%.

Ganggayah et.al.,[10]presented analysis of prognostic factors of breast cancer survival using machine learning techniques. prediction models were built using decision tree, random forest, neural networks, extreme boost, logistic regression, and support vector machine Model evaluation using random forest algorithm yielded slightly better accuracy (82.7%) when compared to other algorithms. The important variables identified in this study were cancer stage classification, tumour size, number of total axillary lymph nodes removed, number of positive lymph nodes, types of primary treatment, and methods of diagnosis.

A. LG and E. AT[11] used three machine learning algorithms for predicting breast cancer recurrence. They implemented machine learning techniques, i.e., Decision Tree (C4.5), Support Vector Machine (SVM), and Artificial Neural Network (ANN) to develop the predictive models. The patients were registered in the Iranian Center for Breast Cancer (ICBC) program from 1997 to 2008. The dataset contained 1189 records.The SVM classification model predicts breast cancer recurrence with least error rate and highest accuracy (95.7%). The predicted accuracy of the DT model is the lowest of all. The results are achieved using 10-fold cross-validation for measuring the unbiased prediction accuracy of each model.

From above literature SVM is the best classifier with high accuracy for predicting breast cancer.

1.3HeartDisease

Lakshmanarao et al.,[12] has built a prediction model for heart disease using ML technique. They used Decision tree, Support Vector Machine, K Nearest Neighbour and Random Forest algorithms are used in the prediction of heart disease and hybrid of these algorithms provides 94 % accuracy.This method proved the accuracy of highest prediction rate.

Lutimathet al.,[13]they built up a productive strategy to procure the clinical and ECG information, in order to group the information in unsupervised way utilizing k-means clustering technique to precisely analyze the heart and anticipate irregularities assuming any. The general procedure can be ordered into three stages. At last they utilizedthis two information's i.e. ECG and clinical information for grouping by k-means approach for ordering the coronary illness and to anticipate variations from the norm in the heart or it's working by contrasting the class and the real class. They at that point locate the most imperative highlights of the dataset utilizing molecule swarm enhancement and bolster vector machines coupled together and after that apply a similar calculation on the decreased dataset. They looked at theoutcomes acquired from the original dataset with the reduced dataset.

Mohan et al.,[14] proposed a novel method that aims at finding significant features by applying machine learning techniques resulting in improving the accuracy in the prediction of cardiovascular disease. The prediction model is introduced with different combinations of features and several known classification techniques. They produced an enhanced performance level with an accuracy level of 88.7% through the prediction model for heart disease with the hybrid random forest with a linear model (HRFLM).

From above literature, hybrid machine learning is very useful for prediction of heart diseases.

1.4. Diabetes Disease

Fiarni et al.,[15] constructed prediction model of diabetes complication disease using data mining algorithm in Indonesia.In this research, the diabetes risk factor narrowed into seven features, which are Age, Gender, BMI, Family history of diabetes, Blood pressure, duration of diabetes suffers and Blood glucose level. Naïve Bayes and C4.5 decision tree-based classification techniques and k means clustering techniques were used to analysed the dataset.accuracy of the proposed model is 68%, with the higest accuracy on Retinophaty prediction model. The model generates rule from diabetic medical data into four groups, which are nephropathy, retinopathy, neuropathy and mixed complications

Sisodia et al.,[16] designed a model for prediction of diabetes using classification algorithms.Three machine learning classification algorithms namely Decision Tree, SVM and Naive Bayes are used in this experiment to detect diabetes at an early stage. Experiments are performed on Pima Indians Diabetes Database (PIDD). Results obtained show Naive Bayes outperforms with the highest accuracy of 76.30% comparatively other algorithms.The performances of all the three algorithms are evaluated on various measures like Precision, Accuracy, F-Measure, and Recall.

K. and Chitra[17]used SVM for classification of diabetes disease.The Pima Indian diabetic database at the UCI machine learning laboratory has become a standard for testing data mining algorithms to see their prediction accuracy in diabetes data classification. The proposed method uses Support Vector Machine (SVM), a machine learning method as the classifier for diagnosis of diabetes. The machine learning method focus on classifying diabetes disease from high dimensional medical dataset. The experimental results obtained show that support vector machine can be successfully used for diagnosing diabetes disease.accuracy of SVM is 78%.

For prediction of diabetes literature said that SVM is best classifier as a prediction model.

III.CONCLUSION

Machine learning technique is very useful in mathematical, statistical and computational data.A machine learning technique plays a very important role in healthcare. due to this we can predict the diseases in short time.ML algorithms a very useful and shows its accuracy also. According to accuracy of ML SVM algorithms is very useful for prediction of diseases.In thyroid disease prediction decision tree show high accuracy.SVM also show high accuracy in breast cancer disease. In heart disease hybrid technique shoe best accuracy and in diabetes SVM perform very good accuracy.So above technique is very useful for taking quick decision and for predicting any disease.

REFERENCES

[1] T. Oladipupo, “Types of Machine Learning Algorithms,” *New Adv. Mach. Learn.*, no. August, 2010, doi: 10.5772/9385.

[2] M. Fatima and M. Pasha, “Survey of Machine Learning Algorithms for Disease Diagnostic,” *J. Intell. Learn. Syst. Appl.*, vol. 09, no. 01, pp. 1–16, 2017, doi: 10.4236/jilsa.2017.91001.

[3] A. J. Frandsen, “Machine Learning for Disease Prediction,” *Brigham Young Univ. Sch.*, p. Paper 5975,

2016, [Online]. Available: https://scholarsarchive.byu.edu/etd/5975.

[4] S. Barik, S. Mohanty, D. Rout, S. Mohanty, A. K. Patra, and A. K. Mishra, "Heart Disease Prediction Using Machine Learning Techniques," *Lect. Notes Electr. Eng.*, vol. 665, no. 3, pp. 879–888, 2020, doi: 10.1007/978-981-15-5262-5_67.

[5] A. I. Technology, "EFFICIENT THYROID DISEASE CLASSIFICATION USING," vol. 88, no. 3, pp. 410–420, 2016.

[6] I. Ioniță, "Prediction of Thyroid Disease Using Data Mining Techniques," pp. 115–124.

[7] M. T. Raghuraman, E. Sailatha, and S. Gunasekaran, "CIKITUSI JOURNAL FOR MULTIDISCIPLINARY RESEARCH ISSN NO : 0975-6876 Discrimination Efficiency Identification for Human Thyroid Disorder Blood Discovery and PredictionUsing Machine Learning Algorithms," vol. 6, no. 6, pp. 31–37.

[8] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis," *Procedia Comput. Sci.*, vol. 83, no. Fams, pp. 1064–1069, 2016, doi: 10.1016/j.procs.2016.04.224.

[9] G. R. Kumar, G. A. Ramachandra, and K. Nagamani, "An Efficient Prediction of Breast Cancer Data using Data Mining Techniques," *Ijiet*, vol. 2, no. 4, pp. 139–144, 2013.

[10] M. D. Ganggayah, N. A. Taib, Y. C. Har, P. Lio, and S. K. Dhillon, "Predicting factors for survival of breast cancer patients using machine learning techniques," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, pp. 1–17, 2019, doi: 10.1186/s12911-019-0801-4.

[11] A. LG and E. AT, "Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence," *J. Heal. Med. Informatics*, vol. 04, no. 02, pp. 2–4, 2013, doi: 10.4172/2157-7420.1000124.

[12] A. Lakshmanarao, Y. Swathi, and P. Sri Sai Sundareswar, "Machine learning techniques for heart disease prediction," *Int. J. Sci. Technol. Res.*, vol. 8, no. 11, pp. 374–377, 2019.

[13] N. M. Lutimath, C. Chethan, and B. S. Pol, "Prediction of heart disease using machine learning," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2 Special Issue 10, pp. 474–477, 2019, doi: 10.35940/ijrte.B1081.0982S1019.

[14] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[15] C. Fiarni, E. M. Sipayung, and S. Maemunah, "Analysis and prediction of diabetes complication disease using data mining algorithm," in *Procedia Computer Science*, 2019, vol. 161, pp. 449–457, doi: 10.1016/j.procs.2019.11.144.

[16] D. Sisodia and D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 1578–1585, 2018, doi: 10.1016/j.procs.2018.05.122.

[17] K. and Chitra, "Classification Of Diabetes Disease Using Support Vector Machine," vol. 3, no. 2, pp. 1797–1801, 2018, [Online]. Available: https://www.researchgate.net/publication/320395340.