

A REVIEW ON MACHINE LEARNING TECHNIQUES INVOLVED IN HEALTHCARE DOMAIN

Akshay Chamoli¹, Hemant Kumar², Arun Agarwal³, and Neha⁴

¹ Jamia Hamdard, New Delhi akachamoli@gmail.com

² Jamia Hamdard, New Delhi hemant.sscbs@gmail.com

³ Ramanujan College, New Delhi arun.agarwal@ramanujan.du.ac.in

⁴ Senior Developer, Alankit Limited, New Delhi nehagodiyal11@gmail.com

Abstract. Machine learning is the most efficient technology and played a major role in the industry domain. It has various applications and especially in healthcare domain machine learning has improved a lot as it has been used for discovering various patterns form diverse sources to provide efficient methods for the prediction of disease. This paper describes various algorithm of machine learning which deals with the development of an effective and efficient decision support system pertaining to applications related to healthcare. This paper has overcome the research gap for the development of an effective decision support system for the application of healthcare.

Keywords: *Healthcare Domain · Decision Support System · Machine Learning Techniques · Disease Prediction.*

Introduction

The technology Machine learning is a multi-disciplinary tool due to its capability of extracting knowledge, predicting future and designing an efficient system. It is being applied in almost every research fields like neuro-science, communication networks, and business analytic [16]. Machine learning is the child domain of artificial intelligence technology as it has the feature to collect information for available data which is further divided into training and testing data set. The approach is to collect the data and let machine decide by itself about the searching and its representation. The data is diagrammatically described using nodes or sub nodes. The categorization of ML algorithms can be done in following ways as given in Fig. 1 [14].

Supervised technique enabling Machine Learning

The logic for this learning deals with training the data set and developing an efficient algorithm which easily maps input to related output. The classification technique is a part and has major role in supervised learning [1].

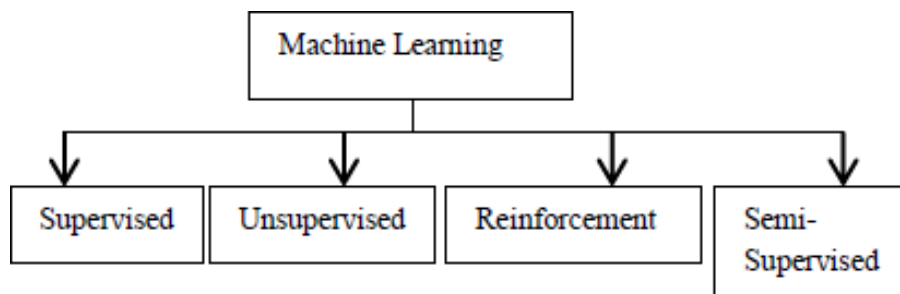


Fig. 1. Classification Techniques featuring machine learning

Semi-supervised technique enabling Machine Learning

This method deals with identifying the optimal classifier from the information which consists of both unlabeled and labelled feature. The efficiency of this procedure is dependent on few of the pre-defined assumptions.

Unsupervised technique enabling Machine Learning

In the process of unsupervised learning, the logic is to defer the training mechanism and giving the input to obtain the required results. The various basic examples for unsupervised techniques are association rule mining, K means, fuzzy and hierarchical clustering [15]. The unsupervised learning deals with the formation of clusters means similar type data grouping with unknown labels on trained data. These algorithm helps in the usage for developing framework for knowledge extraction.

Reinforcement technique enabling Machine Learning

The learning with reinforcement technique, computer program use to give access for dynamic environment for performing a goal oriented job. The focus is to provide feedback in relation with rewards or punishment so as the program learns by itself what is efficient and what is not [8].

Healthcare sector with proficient DSS

Many of the Indians people die each year with the negligence of mistakes produced by hospitals and their doctors. There are a lot of problems and issue with our health care system. The Health Information Technology Framework (HITF)

[28] initiates few effective measures on the involvement of combining, getting the consumer selection of healthcare organizations, and usage of technology. Decision support

This is an effective system working on the concept of machine based learning having health related protection system as the main key and improvises on the large computation ability. Both the machine as well as doctors use to search for various patterns but practically the its really a difficult task for a doctor to examine and analyze the dataset which includes heart rate for patient database or must produce results for the disease in terms of percentage of occurrence for the given disease every time but as compared with machines, they can perform any task optimally

and will provide efficient results for the same.

Healthcare sector in support with decision system

The medical support system which has the inbuilt feature for decision capability and is also convenient with the finance section of the medical domain so for keeping accounts receivable, outlay and accounts payable. This process also involves in keeping the insurance policy for patients with various type of refund choices. This is a specific part of organization having the role to provide various module types for Decision support systems in healthcare domain [21].

The investigation of diseases with the help of Decision support system by arrangement of health related issues for doctors or resolving and exposing with the help of background knowledge for the individual patients. Also it provides the help in rectifying patient's condition and situation at a given instance with providing specific drug for the specific time which reduces the rate of casualty. This can also be seen as a web based scenario that can be utilized to transfer information between patient records and its healthcare attributes.

Applicability of ML in healthcare

Healthcare and IOT

The technological advancement in healthcare and IoT opens up a new research area for designing better solutions for patients suffering with serious diseases. The use of ML, AI techniques in healthcare will have a great impact on providing smart solutions for healthcare monitoring and disease rectification so that efficiency maintained. In the field of medical services with ML feature enabled has improved results as a result, and the various applications involve picture registration, picture description, clinical presentation, etc.

The technology having machine learning enabled feature has limited social impacts on the field of healthcare [2]. The technique related to Machine learning scenario has provided the solution for the increment in health related diagnostic price by maintaining and serving an improved model for patient-clinician scenario.

The solutions provided by machine learning for healthcare in terms of fast disease rectification, health guidance online and it has a great help especially for elderly people.

The current scenario is healthcare involves a very large amount of data that must be accessible and properly maintained. It contains EMRs that has a large amount of data which has the property of both structured or unstructured [12]. The difference between structured and unstructured is that structured data is fully organized and useful as no extra or raw data is present in it while the unstructured data deals with raw and improper data which is not in organized manner.

The structured data related to healthcare is easy to analyze in the database and use to have features of statistics and classes, and is also not restricted of weight of patients and also include generic symptoms like headache, stomach pain etc. [13]. The huge amount of medical knowledge with unstructured information having a variety of data such as images, audio, video, reports and discharging summaries. The procedure that is followed to understand and act upon the conversation that takes place between the supplier and the patient, is a very hectic task. Conversation with patients varies from one to another with different requirements that must be satisfied by the supplier.

Applications of machine learning with the involvement in Healthcare

The algorithm in relation with machine learning technology has the feature to identify complex patterns with huge data. These feature are well suited with clinical related applications and especially for people dependent on genomics and measurements with respect to proteomics. These are mostly used for various detection as well as diagnosis of any disease. The Applications involved with medical domain with machine learning techniques helps in effective decision making so that new treatment plans can be provided for patients so that a new and efficient healthcare system can be provided for the users [4].

The efficient models for predicting any disease in healthcare sector uses factors like patient data, uneasiness levels, exigency office outlines, What's more, even the format of the medical clinic room itself to close stand by times. Utilizing the prognostic model, centers will think medical clinic room affirmations. So AI application could benefit patients by diminishing value, rising exactness, or diffusing experience that is to sum things up offer.

Possible Solution Techniques using Machine Learning

Various solutions have been given in literature that combines machine learning in healthcare as given in Fig. 2.

Supervised technique: Support Vector Machine The design for Support vector machine was in 1990's. To achieve optimal results with machine learning techniques SVM can play a vital role and it's a simple and easy method. In this technique a collection of training samples are used and these are also divided into different categories. The SVM technique basically deals with problems related to classification and regression [17].

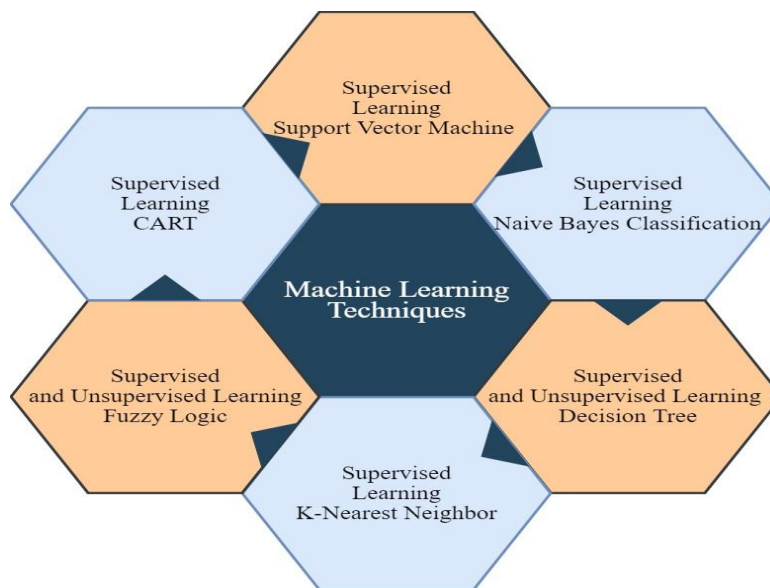


Fig. 2. Machine Learning techniques

Supervised learning: Naive Bayes classification Classifier in relation with statistical analysis is the most appropriate example of Bayesian classifier. The Naïve Bayes technique identifies the probabilities of membership classes based on the class labels provided [7]. It performs various scanning of data so therefore classification becomes easy.

Supervised and unsupervised learning: Decision Tree This classification technique with features of tree type structures consisting of root node, leaf node with class label present within it. The front or top nodes present within decision tree is called root nodes. This scenario is simple and effective technique due to its feature of not consisting any parameters in it [25].

Supervised learning: K-nearest neighbor This is also a common technique which comes under classification. The technique is used for the distance calculation from training samples with N numbers [3].

Supervised and unsupervised learning: Fuzzy Logic This concept is fully based on fuzzy logic. The value comes under 0 to 1. This is a simple process as it is mostly used for the applications related with engineering [31].

Supervised learning: CART The methodology for Classification and Regression techniques which are based on the concept of tree structure is known as CART. In the above mentioned tree methodology the target variable is represented in continuous and categorical state. These variable used for the value prediction in the tree [5].

Disease prediction with machine learning

Machine Learning improvised for the prediction of heart disease

For fast and efficient diagnosis of heart disease these machine learning techniques have proven to provide better results. The various dataset can be provided by UCI machine learning repository. In the paper of Parthiban and Srivatsa have proposed an algorithm based on the properties to detect and analyse heart related diseases with the pre-defined feature of machine learning using Naive Bayes algorithm and Support vector machine techniques [20]. The Naive Bayes technique shows that 74% results for accuracy and SVM shows 94.60% for accuracy. Ootom performed techniques such as Support Vector Machine and Bayes Net for the prediction of coronary heart diseases [18]. SVM attained 88.3% accuracy while Bayes Net attains 84% accuracy.

Machine Learning improvised for the prediction of diabetes

Accuracy related feature for the analysis of diabetic diseases with various machine techniques. Dataset being used in this paper is obtained from UCI Machine Learning Repository. Iyer in his paper has shared the algorithm featuring machine learning for the analysis and prediction of diabetic disorder. The technique being used are Naive Bayes and Decision trees. Naive Bayes has provided accuracy of 79.56% while decision tree has provides accuracy of 76.95% [6]. In the paper of Dash and Sen has provided with techniques related to machine learning and produced the diagnosis of diabetes related disease. Logiboost and CART algorithms results shows that Logiboost has correctness of 77.479% [23].

Machine Learning improvised for the prediction of Breast cancer

This disease commonly happens in females and most commonly found in locations of United States and in Asian nations. A few machine learning techniques are very commonly used for the prediction of breast cancer. The WISCONSIN dataset is being used with UCI machine learning

repository. Williams et al. also used J48 predictive methods for the study of breast cancer. The research simulation is done in WEKA tool and results produced with J48 providing optimal results with accuracy of 94.2% and Naive Bayes on the other hand use to provide accuracy of 82.6% [30].

To predict the complex diseases like breast cancer Senturk et al. in their research paper have used many classification models involving Support vector machine (SVM), Naive Bayes (NB), K-nearest neighbor (K-NN), and Decision tree (DT). Accuracy produced with K-NN algorithm is 95.15% and accuracy shared with SVM algorithm is 96.40% [22].

In this paper Majali et al. shared that detection of breast cancer is a complex task and with the help of learning techniques like frequent patterns and decision making tree the results are obtained to be more efficient. The optimal results has been provided with decision tree technique and shown accuracy results to be more optimal with 94% [15].

Machine Learning improvised for the prediction of Thyroid Disorder

For the prediction of thyroid diseases many machine learning algorithm can get used. The algorithm based on classification technique using support vector machine and Decision tree as given in Table 1.

Table 1. Diagnosis of diseases with a variety of machine learning techniques

S.No.	References	ML Techniques	Disease	Data-set	Accuracy
1	Chaurasia Pal [5]	J48, SVM	Heart, Disease	UCI	84.35%, 85.03%
2	Vembandasamy [29]	Naive Bayes	Heart Disease	DDRI Chennai	86.41%
3	Kumari and Chitra [10]	SVM	Diabetic Disease	UCI	78%
4	Sarwar and Sharma [25]	Naive Bayes	Diabetes type-2	India	95%
5	Shrivastava et al [26]	J48	Breast Cancer	UCI	98.14%
6	Venkatesan et al [11]	Decision Tree	Breast Cancer	SVDC Hospital	97%
7	Sivakami [27]	DT+SVM	Breast Cancer	UCI ML	91%
8	Shajahaan et al [24]	CART	Naive Bayes	BCU Wisconsin	92.42% , 97.42%
9	Kousarrizi et al [9]	SVM	Thyroid Disease	UCI	98.62%

The dataset are taken from UCI repository. Papageorgiou EI, Papandrianos NI have proposed various advanced approaches for thyroid diagnosing diseases using the fuzzy map technique utilizing data mining algorithms [19].

CONCLUSION

The machine learning algorithm are efficient in nature and use to provide better results. Healthcare domain on the other hand is improved with the involvement of techniques having the inbuilt features of machine learning. The paper here deals with a variety of machine learning techniques applied with the healthcare cases and better results have been achieved. The most common health related diseases are heart disease, breast cancer, diabetic disease and thyroid disease. The analysis study on ML techniques showed that Naïve Bayes giving accuracy of 86% in relation for diagnostic with heart disease. SVM on the other hand gives an accuracy of 96.40% for the diagnostic of breast cancer while CART gives an accuracy of 79% for diabetes detection. The future analysis must include improvement in the accurate prediction in any of the cancerous disease with the help of various machine learning techniques.

References

- Allix, N.M., et al.: Epistemology and knowledge management concepts and practices. *Journal of Knowledge Management Practice* 4(1) (2003)
- Bhardwaj, R., Nambiar, A.R., Dutta, D.: A study of machine learning in healthcare 2, 236–241 (2017)
- Bishop, C.M., et al.: *Neural networks for pattern recognition*. Oxford university press (1995)
- Callahan, A., Shah, N.H.: *Machine learning in healthcare* pp. 279–291 (2017)
- Chaurasia, V., Pal, S.: Data mining approach to detect heart diseases. *International Journal of Advanced Computer Science and Information Technology (IJACSIT)* Vol 2, 56–66 (2014)
- Iyer, A., Jeyalatha, S., Sumbaly, R.: Diagnosis of diabetes using classification mining techniques. *arXiv preprint arXiv:1502.03774* (2015)
- Jin, B., Che, C., Liu, Z., Zhang, S., Yin, X., Wei, X.: Predicting the risk of heart failure with ehr sequential data modeling. *Ieee Access* 6, 9256–9261 (2018)
- Kaelbling, L.P., Littman, M.L., Moore, A.W.: Reinforcement learning: A survey. *Journal of artificial intelligence research* 4, 237–285 (1996)
- Khennou, F., Khamlichi, Y.I., Chaoui, N.E.H.: Improving the use of big data analytics within electronic health records: a case study based openehr. *Procedia Computer Science* 127, 60–68 (2018)
- Kumari, V.A., Chitra, R.: Classification of diabetes disease using support vector machine. *International Journal of Engineering Research and Applications* 3(2), 1797–1801 (2013)
- Learning, M.: Heart disease diagnosis and prediction using machine learning and data mining techniques: A review. *Advances in Computational Sciences and Technology* 10(7), 2137–2159 (2017)
- Liu, Z., Hauskrecht, M.: Clinical time series prediction: Toward a hierarchical dynamical system framework. *Artificial intelligence in medicine* 65(1), 5–18 (2015)
- Maddux, D.: *The human condition in structured and unstructured data*. Acumen Physician Solutions
- Manogaran, G., Lopez, D., Thota, C., Abbas, K.M., Pyne, S., Sundarasekar, R.: Big data analytics in healthcare internet of things pp. 263–284 (2017)
- Mathur, A., Moschis, G.P.: Socialization influences on preparation for later life. *Journal of Marketing Practice: Applied Marketing Science* (1999)
- Mitchell, T.: *Machine learning*, mcgraw-hill higher education. New York (1997)
- Murphy, K.P.: *Machine learning: a probabilistic perspective*. MIT press (2012)

Otoom, A.F., Abdallah, E.E., Kilani, Y., Kefaye, A., Ashour, M.: Effective diagnosis and monitoring of heart disease. *International Journal of Software Engineering and Its Applications* 9(1), 143–156 (2015)

Papageorgiou, E.I., Papandrianos, N.I., Apostolopoulos, D.J., Vassilakos, P.J.: Fuzzy cognitive map based decision support system for thyroid diagnosis management pp. 1204–1211 (2008)

Parthiban, G., Srivatsa, S.: Applying machine learning methods in diagnosing heart disease for diabetic patients. *International Journal of Applied Information Systems (IJ AIS)* 3(7), 25–30 (2012)

- Rajalakshmi, K., Mohan, S.C., Babu, S.D.: Decision support system in healthcare industry. *International Journal of computer applications* 26(9), 42–44 (2011)
- Senturk, Z.K., Kara, R.: Breast cancer diagnosis via data mining: performance analysis of seven different algorithms. *Computer Science & Engineering* 4(1), 35 (2014)
- Shailaja, K., Seetharamulu, B., Jabbar, M.: Machine learning in healthcare: A review pp. 910–914 (2018)
- Shajahaan, S.S., Shanthi, S., ManoChitra, V.: Application of data mining techniques to model breast cancer data. *International Journal of Emerging Technology and Advanced Engineering* 3(11), 362–369 (2013)
- Sharma, P., Bhatia, A.P.R.: Implementation of decision tree algorithm to analysis the performance (2012)
- Shrivastavat, S.S., Sant, A., Aharwal, R.P.: An overview on data mining approach on breast cancer data. *International Journal of Advanced Computer Research* 3(4), 256 (2013)
- Sivakami, K., Saraswathi, N.: Mining big data: breast cancer prediction using dt- svm hybrid model. *International Journal of Scientific Engineering and Applied Science (IJSEAS)* 1(5), 418–429 (2015)
- Thompson, T.G., Brailer, D.J.: *The decade of health information technology: delivering consumer-centric and information-rich health care*. Washington, DC: US Department of Health and Human Services (2004)
- Vembandasamy, K., Sasipriya, R., Deepa, E.: Heart diseases detection using naive bayes algorithm. *International Journal of Innovative Science, Engineering & Technology* 2(9), 441–444 (2015)
- Williams, K., Idowu, P.A., Balogun, J.A., Oluwaranti, A.I.: Breast cancer risk prediction using data mining classification techniques. *Transactions on Networks and Communications* 3(2), 01 (2015)
- Zimmermann, H.J.: *Fuzzy control*. Springer (2001)