JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Study on Machine Learning in area of Healthcare

1st Shalinee Pareek
Computer science engineering
Sangam university
Bhilwara,India
shalinee.pareek@sangamuniversity.ac.in

2nd Ajay Kumar Suwalka
Computer science engineering
Sangam university
Bhilwara,India
ajay.suwalka@sangamuniversity.ac.in

3rd Himanshu Kothari
Computer science engineering
Sangam university
Bhilwara,India
himanshu.kothari@sangamuniversity.ac.in

Abstract-In today scenario machine learning is contemporary and highly enlightened technology which becomes immense and trendy in the industrial area. There are various applications in which machine learning is used widely. Machine learning is playing crucial part in various fields' such as health and medical area, finance and the area of security. In health area machine leaning technology serves as important technology by using and discovering different patterns from various data and produce capabilities to discover the critical diseases. In this paper, we review different algorithms of machine learning that are used for developing decision support in efficient manner for application in the area of healthcare. This paper helps to reduce the gap of research for developing efficient decision support system for applications in medical area.

Keywords- Machine Learning; Medical area; Disease Discovery; Healthcare; Decision Support System; Heart Disease.

I. INTRODUCTION

Machine learning has the statistics, algebra, knowledge analytics and data processing and can be considered as extensive area of multidisciplinary that make the same difficult to define. Machine Learning is the methodology of artificial intelligence that retrieves the meaningful information from training data. Machine learning is divided into classes are as follow[2].

A. Supervised Learning.

B. Unsupervised Learning

C. Reinforcement Learning

A. Supervised Machine Learning

A. Supervised Machine Learning

This learning is taken under training and algorithm develops an exercise that matches inputs to related outputs. One common establishment of the supervised learning task is the classification issue.

B. Semi-supervised Learning

It is a method of determining the best classifier from each unlabeled and labeled information. By using untagged information, it provides high classification efficiency. The success of this method depends entirely on a few basic assumptions.

C. Unsupervised M L

In unsupervised learning, shortcuts do not learn. The technique used for unsupervised learning is clustering, fuzzy clustering, hierarchical clustering, K stands for clustering, associative rule analysis of the algorithm that is included in this training. Here the cluster is formed by the presence of trained data with unknown labels. These algorithms are used to develop a framework using sample data.

D. Reinforcement Learning

In this learning, access is given by the computer program to the dynamic environment to achieve a specific goal. Feedback on rewards and punishments is provided to the program as it navigates to its disadvantage.

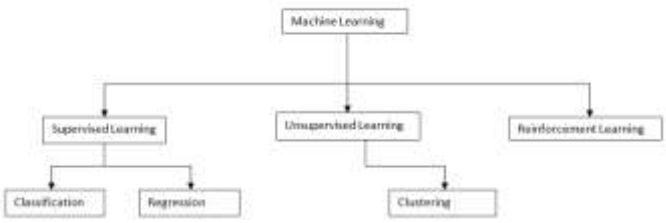

FIG 1: CLASSIFICATION TECHNIQUES IN M L

II. THE NEED OF THE DECISION MAKING SUPPORT SYSTEM FOR HEALTHCARE

Most Americans die each year from health care errors and thousands of people are damaged by non-fatal burns due to the constant cause. The Health Information Technology (IT) Framework [6] has recommended a number of strategies, such as collaboration, knowledge of consumer selection of doctors and organizations, and adoption of IT.

A. *Decision support*

An effective health care system based on machine learning is based on the great computing power of a computer, as well as the ability of a doctor to reason. Both the apparatus and the doctor are looking for patterns, but the doctor cannot assess the heartbeat of each patient or know all the nuances of the disease.

B. *Decision Support System in Healthcare*

The decision support system helps the clinic's financial branch keep records of fees, receivables, expenses and accounts payable. This method also helps to keep the patient's insurance policy at various reimbursement options. It is an organization that offers completely different modules for DSS in the field of healthcare [7].

The decision support system helps in investigating the disease by arranging a perception of health problems to doctors or otherwise by exposing basic knowledge about individual patients. It also provides help in identifying a patient's situation and a guide suggests to the patient when to use the appropriate medication at a certain time, a web-based structure that has been involved with the patient's computer / therapeutic registration.

III. HEALTHCARE USING MACHINE LEARNING

A. *Introduction*

Hazardous improvements, given new opportunities for health-related information for the development of a patient's well being, show a significant performance in machine learning health care and are mostly applicable to healthcare, including computer aided diagnosis, image registration Includes image annotation, image healing, and image database retrieval, multimodal image fusion, medical image segmentation, where the deficiency may be incurable.

Machine learning is likely to have limited social effects in the field of health [8]. Machine learning provides the solution to lower the rising cost of healthcare and to create improved patient-clinician communication. ML solutions will be used for an excess of health-relevant uses; some include serving clinicians to identify a lot of prescriptions and personalized therapy for patients and, in addition, serving patients for identification only once and if they need to record subsequent appointments.

Currently, in the field of health, a huge amount of information has become accessible. It contains EMRs that consist of information that can be either unstructured or structured [9]. Structured health information is simple data analyze in a database and will have a set of statistics and classes, but still are not limited to patient weights and even generic symptoms such as stomach pain, headache, etc. [10]. Most medical knowledge is unstructured information in numerous notes, images, audio and video recordings, completely different reports and download summaries. It is terribly tiring to quantify and analyze a conversation between provider and patient; the conversation is incredibly personalized and could take many alternative directions [11].

B. *Applications of ML in Healthcare*

Machine learning algorithms are useful in identifying complicated patterns in huge and prosperous data. It is often used in numerous diagnoses and disease detection. In medical applications, machine learning algorithms will produce greater decisions about treatment plans for patients through suggestions for implementing a useful health care system. [12].

Healthcare management uses this method to forecast waiting times for patients in the emergency department who are waiting for places. These models use factors such as patient information, discomfort levels, demand department charts, and even the appearance of the hospital room itself to end waiting times.

Using the prognostic model, clinics will consider hospitalizations. Thus, the machine learning application could benefit patients by lowering the price, increasing the accuracy or disseminating the experience offered briefly.

C. *Different Techniques Used by ML*

1. Support Vector Machine

Support Vector Machine (SVM), designed in the 1990s. To perform machine learning tasks (ML) the Support Vector Machine (SVM) is used and is a simple and prominent process. During this technique, a collection of training samples is given each sample is divided into different categories. Support vector machine (SVM) used mainly for classification and regression problems [13].

2. Naive Bayes classification

Statistical classifiers are an example of Bayesian classifiers. Naive Bayes identifies the probabilities of belonging to the class based on the given class label [14]. It performs a data scan and therefore the classification is easy.

3. Decision Tree

Choice Tree (DT) is generally utilized method for grouping containing inward hub and one leaf hub with a class name. The top hubs of the choice tree (DT) are called as root hubs. The choice tree is exceptionally famous in light of the fact that the development is extremely straightforward, which won't need any boundaries[15].

4. K-nearestneighbor

K-closest neighbor is oftentimes utilized methodology for arrangement of tests. By utilizing this strategy we can compute distance measure from N number of preparing tests [16].

5. FuzzyLogic

Fuzzy Logic which is evolved from the theory of Fuzzy sets. These values are between 0 and 1. It is a very popular method that is used in engineering applications [17].

6. CART

The classification and regression tree methodology is called CART. In the classification and regression trees, the target variable is represented as categorical and continuous. These variables are used to predict values in the tree[18].

IV. LITERATURE REVIEW

Janee et al [6]; proposed “multi-stage lung cancer detection and prediction using multi-class SVM classifier” describes an effective algorithm using SVM for lung cancer detection, diagnosis and it's also capable to anticipate the possibility of lung cancer. An algorithm is developed using MATLAB and employing image processing techniques like image enhancement, detection and segmentation, detection and feature extraction is done by adopting the Gray Level Co- occurrence Method (GLCM) method. SVM is applied for classification purposes. Prediction is done through the binarization technique. UCI ML database is acquired, which includes 500 infected and 500 non-infected CT images. The proposed method detected 126 images as infected out of 130

and predicted 87 images as cancerous out of 100 formerly specified images. The experimental analysis shows 97% precision in the identification and 87% for prediction.

M.Gomathi et al [7]; defined "A Computer-Aided Diagnosis System for Detection of Lung Cancer Nodules using Extreme Learning Machine" In this paper CAD model is developed for the analysis of cancer in Ct images. The fundamental phase of CAD is to detect the region of interest in input CT images. The lung region extraction is preceded with lung region segmentation, detection of cancer nodules is accomplished with Fuzzy Possibility CMean (FPCM) clustering algorithm. A maximum Drawable Circle intensity value is used for the formulating the diagnostic rules. Then these rules are implemented to learn along with the assist of the ELM (Extreme Learning Machine).

Shingo kakeda et al [8]; proposed a commercial CAD model for detecting lung nodules on Chest films. This paper presents the CAD model, which comprises an image server and EpiSight/XR software. This method is carried out in four fundamental steps. The first complex automatic structure is reduced to produce different images. Nodule candidates are detected by employing the Multiple Gray Level thresholding methods. To differentiate the true nodule with false-positive nodule features are input and difference images are used to extract features. Formerly extracted features are utilized for the reduction of false-positive nodules; a rule-based analysis and ANN are employed for this purpose. For experimentation, the model developed was analyzed on a database that consists of 274 radiographs with 323 lung nodules. Out of 315, 235(75%) false-positive images were detected as a normal automatic structure and 155(49%) are detected as pulmonary vessels.

Metin N et al [9]; proposed "Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system" The proposed method is carried out in five steps, first regions are segmented by applying k-means clustering method. After segmenting the lung curves, by employing pre-processing algorithms suspicious regions are segmented from that of lung regions that produce a binary image which consists of holes due to the segmentation process, flood –the filled algorithm is used to fill this hole as nodule-candidate are considered as solid objects. This system may contain general regions and lung nodules comprising of blood vessels. To differentiate this nodule, Rule-based classifiers with 2D and 3D features are used. At last, the false positive objects are detected using Linear Discrimination Analysis (LDA). The proposed method was analyzed on a dataset that includes 1454CT images gathered from 34 diagnosed patients with 63 lung nodules.

Kazak Awai et al [10]; described a system for evaluating the CAD effect on radiologist's pulmonary nodules detection. The proposed method used image processing techniques for lung and intrapulmonary structure segmentation gray-level threshold, 3D labeling techniques and mathematical morphological techniques are used for lung segmentation. For the segmentation of intrapulmonary structures Top-hat transformation method is employed, on an input image to detect the smoothed image. Sieve filter is used for the identification of primary potential nodules, then features of these pulmonary nodules are extracted to differentiate true nodules from that of false-positive nodules, ANN is adapted to decide the probability of region of interest based on an image feature. Adaption of this system enhanced pulmonary nodule resident's detection of CT scans.

Cheran et al [11]; proposed "Computer-aided diagnosis for lung CT using artificial life models" This CAD model is developed by deploying various algorithms. First, the ribcage region is determined by employing a 3D region growing algorithm, and then the active contour technique is implemented to develop a specific area for the approaching ants which are redistributed to develop a specific and precise rebuilding of the vascular tree and pleura. To regenerate the bronchial and the vascular trees artificial life models are used. By utilizing active shape models, it is determined whether the previously constructed branches contain nodules and also to detect whether the nodules are connected to the pleura. By employing snakes and dot enhancement algorithm cleaner algorithm is produced to confine the nodules.

Lakshmanaprabu et al [12]; developed the Optimal Deep Neural Network (ODNN) and Linear Discriminate Analysis (LDA) based classification model for CT images. Lung nodule classification is done by applying LDR and optimization is done by applying the Modified Gravitational Search Algorithm to predict lung cancer. Standard CT database is used for experimental analysis which comprises 50-low dosage lung cancer CT images. This model is correlated with existing models such as KNN, NN, DNN SVM, and so on., and the experimental analysis shows best results for the developed model with 94.56% accuracy, sensitivity and specificity 96.2% and 94.2% respectively.

Worawate et al [13]; developed a method "Automatic Lung Cancer Prediction from Chest X-ray Images using Deep Learning Approach", Authors used DensNet-121 (121 layers Convolutional neural network) in conjunction with transfer learning for classifying using chest images. Model is trained on two datasets i.e. Chest X-ray 14 and JSRT to identify the nodules. The model obtained an accuracy of about 74.43±6.01%, sensitivity, and specificity of about 74.68 ± 15.33% and 74.96 ± 9.85% respectively.

Christoph et al [14]; proposed a prediction model based on tomography lung cancer images. A CNN is utilized for feature extraction by fine-tuning pre-trained ResNet18 and multimodal features CNN is trained by the Cox model for hazard prediction. Lung1 dataset is used for experimental analysis which can be accessed from "The Cancer Imaging Archive" (TCIA) that comprises 422 NSCLC (Non-Small Cell-Lung Carcinoma) images for 318 of 422 patients.

Jason et al [15]; developed a Deep Screener algorithm which is a form of different deep learning approaches. The Deep Screener is an end-to-end automated screening of lung cancer on low dose CT images. TCIA dataset is used for experimental analysis which comprises of 1449 low dose CT images. The model developed was correlated with the grt123 algorithm of Data Science Bowl 2017 for lung cancer analysis and the result was too close wining algorithm grt123.

The proposed model predicted about 1359 out of 1449 CT scans about 82% accuracy, AUC about 0.885, AUPRC about 0.837.

In table 4: We have reviewed different lung cancer prediction & detection research papers and listed out the algorithms and Datasets and measures used in those papers.

V. ML TECHNIQUES USED FOR PREDICTION OF VARIOUSDISEASES

A. Heart disease detection

ML techniques are often used to improve the accuracy of diagnosing heart disease. This dataset is viewed by the UCI Machine Learning Repository.

Parthiban and Srivatsa proposed a ML algorithm for detecting and analyzing heart disease using a naive Bayesian algorithm, a vector support machine [19]. Using Naive Bayes algorithm, it achieves 74% accuracy and SVM achieves 94.60% accuracy.

Otoom used a support vector machine, a Bayesian network, to predict coronary heart disease [20]. SVM accuracy - 88.3%, Bayesian network - 84%.

B. Diabetes disease analysis

A variety of machine learning techniques are used to increase the accuracy of scrutiny for diabetic diseases. This is a dataset considered by the UCI Machine Learning Repository.

Iyer proposed a ML algorithm to predict diabetic disorder by using Naive Bayes and Decision trees. Naive Bayes gives 79.56% accuracy and decision tree provides 76.95% accuracy.

Dash and Sen performed ML algorithms for diagnosing diabetes disease. Logiboost, CART algorithms are used and Logiboost provides the correctness of 77.48%.

C. Machine Learning in Disclosure of Breast cancer

It is one of the top cancer that occurs in a woman and it is the second main leading reason for woman in the United States and in Asia countries. Some machine learning algorithms are used to predict breast cancer. The data considered from WISCONSIN dataset UCI machine learning repository.

Williams et al. used a j48, Naive Bayes to identify breast cancer risks in the United States. The experiment is performed using the WEKA tool. They conclude that j48 is the best algorithm for predicting breast cancer, it offers an accuracy of 94.2%, and Naive Bayes offers 82.6% [23].

To predict breast cancer Senturk et al. used several classification models, such as Support Vector Machine (SVM), Naive Bayes (NB), nearest neighbor K, and decision tree (DT). K-NN offers an accuracy of 95.15% and SVM an accuracy of 96.40%[24].

Majali et al. used decision tree and Frequent Pattern in data mining to predict the breast cancer. They conclude decision tree gives 94% accuracy [25].

D. Diagnosis of Thyroid Disorder

Machine learning techniques are used to predict thyroid disease. Classification algorithms that support vector machines and the decision tree are used and the data set in the ICU repository is taken into account.

Papageorgiou EI, Papandrianos NI has proposed advanced approaches for diagnosing thyroid disease using a fuzzy map using data extraction algorithms [26]. The table below summarizes the different ML techniques used to diagnose various diseases.

VI CONCLUSION

As today scenario businesses are growing in different field machine learning (ML) techniques are decisive in these fields. In the field of healthcare, there are different types of exorbitant. There is a number machine learning techniques are adopted to amend them. In this, here is presenting several machine learning techniques to predict the different diseasessuch as disease of heart, breast cancer, diabetic disease and disease of thyroid. In future, we are trying to improve the accuracy of cancer prediction at initial stage by adopting distinct machine learning algorithms.

REFERENCE

[1] T. Mitchell, " Machine Learning", McGraw Hill. P. 2,1997.

[2] Apoorva Mishra and AnupamShukla, "From Machine Learning to Deep Learning Trends and Challenges", CSI Communications, December 2018.

[3] Allix, N. M. "Epistemology And Knowledge Management Concepts And Practices". Journal of Knowledge ManagementPractice.

[4] Anil Mathur, G. P. "Socialization influences on preparation for later life". Journal of Marketing Practice: Applied Marketing Science, vol.5, 163 – 176,2007.

[5] LP. Kaelbling and Aw. Moore, "Reinforcement Learning A Survey", Journal of Artificial Intelligence Research, vol-4, pp.237-285,1996.

[6] T.G. Thompson and D.J. Brailer, "The decade of health information technology: Delivering consumer-centric and information-rich health care: Framework for strategic action",2007.

[7]

[8] K. Rajalakshmi, S. Chandra Mohan, S. DhineshBabu, "Decision Support System in Healthcare Industry", International Journal of Computer Applications, vol.9,2013.

[9]

[10] RohanBhardwaj, Ankitha R. Nambiar, "A Study of Machine Learning in Healthcare", IEEE 41st Annual Computer Software and Applications Conference, 2017.

[11] M. Hauskrecht ,Visweswaran S., Cooper G., Clermont G.: Clinical Alerting of Unusual Care that Is Based on Machine Learning ,2015.

[12] M. Kohn, "Real World Data and Clinical Decision Support", ELSEVIER, 2009.

[13] D. Maddux, "The Human Condition in Structured and Unstructured Data".

[14] Mrs.J. Sukanya, "Applications of Big Data Analytics and Machine Learning Techniques in Health Care Sectors", International Journal of Engineering and Computer Science, vol.6, pp.21963-21967,2017.

[15] K .P Murphy, Machine Learning: A Probabilistic Perspective, The MIT Press,2012.

[16] Animesh et al., "Study and analysis of Breast cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms", International Journal of Computer Applications, vol.2,2016

[17] P. Sharma and A. P. R. Bhartiya, "Implementation of Decision Tree Algorithm to Analysis the Performance", vol. 10, pp. 861–864,2012.

[18] C.M.Bishop(1995)," Neural networks for pattern recognition", England: Oxford University.

[19] H. Zimmermann, Fuzzy set theory and itsapplications. Kluwer Academic Publishers.2001.

[20] Jalpaiguri et al.," Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques: A Review", Advances in Computational Sciences and Technology, vol.10, pp. 2137-2159,2017.

[21] Parthiban, G. and Srivatsa, S.K., "Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients",

International Journal of Applied Information Systems, vol.3, pp.25-30,2012.

[22] Otoom A.F., Abdallah E.E., Kilani Y., Kefaye A., "Effective Diagnosis and Monitoring of Heart Disease", International Journal of Software Engineering and Its Applications, vol.9, pp.143-156,2015.

[23] Iyer A., Jeyalatha S. and Sumbaly R, "Diagnosis of Diabetes Using Classification Mining Techniques", International Journal of Data Mining & Knowledge Management Process (IJDKP), vol.5, pp.1-14,2015.

[24] Sen S.K. and Dash S, "Application of Meta Learning Algorithms for the Prediction of Diabetes Disease", International Journal of Advance Research in Computer Science and Management Studies, vol.2, pp.396- 401,2014.

[25] Williams et al., "Breast cancer risk prediction using data mining classification techniques", Transactions on Networks and Communicatons,vol.2,pp.1-11,2015.

[26] Senturk et al., "Breast cancer Diagnosis via Data Mining: Performance Analysis of Seven Different Algorithms ", Computer Science & Engineering", vol.1, pp.1-10,2014.

[27] Majali et al., "Data Mining Techniques for Diagnosis and Prognosis of Cancer", International Journal of Advanced Research in Computer and Communication Engineering, vol.3, pp. 613-616,2015.

[28] Papageorgiou EI, Papandrianos NI, Apostolopoulos DJ, Vassilakos PJ., " Fuzzy cognitive map based decision support system for thyroid diagnosis management", International Conference on Fuzzy Systems pp. 1204-1211,2008.

[29] Chaurasia, V. and S. Pal," Data Mining Approach to Detect Heart Disease", International Journal of Advanced Computer Science and Information Technology, vol.2,pp.56-66,2018.

[30] Vembandasamy et al., "Heart Diseases Detection Using Naive Bayes Algorithm", vol.2, pp. 441-444,2015.

[31] Kumari, V.A. and R. Chitra, "Classification of Diabetes Disease Using Support Vector Machine", International Journal of Engineering Research and Applications, vol.3, pp. 1797-1801,2013.

[32]] Sarwar, A. and Sharma, V. , "Intelligent Naïve Bayes Approach to Diagnose Diabetes Type-2. Special Issue", International Journal of Computer Applications and Challenges in Networking, Intelligence and Computing Technologies, vol. 3, pp.14-16,2012.

[33] Shrivastava et al., "An Overview on Data Mining Approach on Breast Cancer data", International Journal of Advanced Computer Research, vol.3, pp. 256-262,2013.

[34] E. Venkatesan et al., "Performance analysis of decision tree algorithms for breast cancer classification", Indian Journal of Science and Technology, vol.8, pp.1-8,2015.

[35] K. Sivakami, "Mining Big Data: Breast Cancer Prediction using DT-SVM Hybrid Model", International Journal of Scientific Engineering and Applied Science vol.1, pp.418-429,2015.

[36] S. Shajahaan et al., "Application of Data Mining techniques to model breast cancer data" International Journal of Emerging Technology and Advanced Engineering, vol.3, pp.362-369,2013.

[37] M. R. NazariKousarrizi et al., "An Experimental Comparative Study on Thyroid Disease Diagnosis Based on Feature Subset Selection and classification", International Journal of Electrical& Computer Sciences, vol.1,pp.13-19,2012.