A

Seminar Report

On

Machine Learning in Medical Science

By

Abhishek R. Ranjane

Roll No: 305002

Under the Guidance of

Dr. J. R. Prasad

Department of Computer Engineering

Sinhgad College of Engineering, Pune-41

Accredited by NBA

SAVITRIBAI PHULE PUNE UNIVERSITY

2018-19

Sinhgad Technical Education Society,

Department of Computer Engineering

Sinhgad College of Engineering, Pune - 41

Sinhgad Institutes

Date:

Certificate

This is to certify that Mr. Abhishek R. Ranjane has successfully completed her seminar work on " Machine Learning in Medical Science " at Sinhgad College of Engineering , Pune in the partial fulfillment of the Graduate Degree course in T.E. at the department of Computer Engineering, in the academic Year 2018-2019 Semester VI as prescribed by the Savitribai Phule Pune University.

Dr. J. R. Prasad
Guide
.

Prof. M. P. Wankhade
Head of Department
Department of Computer Engineering

Dr. S. D. Lokhande
Principal
Sinhgad College of Engineering

Acknowledgements

Every work is source which requires support from many people and areas. It gives me proud privilege to complete the seminar on "Machine Learning in Medical Science " under valuable guidance and encouragement of my guide Dr. J. R. Prasad. I am also extremely grateful to our respected H.O.D.(Computer Dept.) Prof. M. P. Wankhade for providing all facilities and every help for smooth progress of seminar. I would also like to thank all the Staff Member of Computer Engineering Department for timely help and inspiration for completion of the seminar. At last I would like to thank all the unseen authors of various articles on the Internet, helping me become aware of the research currently ongoing in this field and all my colleagues for providing help and support in my work.

Abhishek R. Ranjane

Roll No.-305002

Abstract

Machine Learning, a form of Artificial Intelligence (AI) is infiltrating medical field and it is an era where machines can play an important role in health improvement. Various systems from machine learning can be applied to medical data repositories that are too large for human brain to parse. The patterns build from large clinical data warehouse, by the machine learning algorithms, can help researchers to draw conclusions and predict an event. Various systems and algorithms are used in medical sector, this seminar focuses on Bayesian classifier, neural networks and decision trees. This seminar presents a comprehensive survey on Machine Learning in medical domain, which identifies the contributions of different methods and applications. In addition current issues and challenges are also discussed to identify promising areas of research.

List of Figures

2.1 use of ML algorithms in medical literature. 4

3.1 building process of Decision trees. 12

3.2 Structure of a neural network. 13

Acronyms

ML	Machine Learning
ULM	Unsupervised Learning Methods
SLM	Supervised Learning Methods
ANN	Artificial Neural Networks
SVM	Support Vector Machine
PCA	Principal Component Analysis

Contents

Certificate **i**

Acknowledgements **ii**

Abstract **iii**

List Of Figures **iii**

Acronyms **v**

1 INTRODUCTION **1**

1.1 Motivation 1

1.2 Timeline/Evolution 2

1.3 Organization Of Report 2

2 Literature Survey **3**

2.1 Fundamentals 3

2.1.1 Unsupervised Learning Methods (ULM) in Medical Diagnosis . . 3

2.1.2 Supervised Learning Methods (SLM) in Medical Diagnosis 4

2.2 Related Works 5

2.2.1 Behavioral modification 5

2.2.2 Drug Discovery and Development 6

2.2.3 Clinical Trial Research 6

2.2.4 Epidemic Outbreak Prediction 6

2.2.5 Radiotherapy 7

2.2.6 Disease Identification 7

2.3 Survey Conclusion 7

3 METHODOLOGY **8**
3.1 Specific requirements for machine learning systems 8
3.1.1 Good performance 8
3.1.2 Dealing with missing data 9
3.1.3 Dealing with noisy data 9
3.1.4 Transparency of diagnostic knowledge 9
3.1.5 Explanation ability 9
3.1.6 Reduction of the number of tests 9
3.2 Method 1: Naive Bayes 10
3.3 Method 2: Decision trees 11
3.4 Method 3: Neural Networks 13
4 Discussion **15**
4.1 Issues and Key Challenges 15
4.1.1 Datasets 15
4.1.2 Privacy and Confidentiality 15
4.1.3 Data Inaccuracies 16
4.1.4 Accountability 16
4.1.5 System Implementation 16
4.1.6 Data Security 17
4.1.7 Threat to Jobs 17
4.2 My views regarding the future role of machine learning in medical diagnosis 17
5 CONCLUSION **19**
References **20**

Chapter 1

INTRODUCTION

Machine Learning is a part of computer science, whose focus is to make machines analyze a large amount of data, use algorithms and techniques to train models that can predict the outcome of a certain event.Machine learning algorithms were from the very beginning designed and used to analyze medical datasets. Today, machine learning provides several indispensable tools for intelligent data analysis. Especially in the last few years, the digital revolution provided relatively inexpensive and available means to collect and store the data.

Modern hospitals are well equipped with monitoring and other data collection devices, and data is gathered and shared in large information systems. Machine learning technology is currently well suited for analyzing medical data, and in particular there is a lot of work done in medical diagnosis in small specialized diagnostic problems.

Data about correct diagnoses are often available in the form of medical records in specialized hospitals or their departments. All that has to be done is to input the patient records with known correct diagnosis into a computer program to run a learning algorithm. This is of course an oversimplification, but in principle, the medical diagnostic knowledge can be automatically derived from the description of cases solved in the past. The derived classifier can then be used either to assist the physician when diagnosing new patients in order to improve the diagnostic speed, accuracy and/or reliability, or to train students or physicians non-specialists to diagnose patients in a special diagnostic problem.

1.1 Motivation

The historical development of machine learning and its applications in medical diagnosis shows that from simple and straightforward to use algorithms, systems and methodology have emerged that enable advanced and sophisticated data analysis. In the future, intelligent data analysis will play even a more important role due to the huge amount of information produced and stored by modern technology.

Current machine learning algorithms provide tools that can significantly help medical practitioners to reveal interesting relationships in their data. Experiments show that in medical domains, various classifiers perform roughly the same. So one of the important factors when choosing which classifier to apply is its explanation ability. Experiments show that the physicians prefer explanations as provided by the Bayesian classifiers and decision tree classifiers: Assistant-R and LFC. However, instead of selecting a single best classifier, it seems that the best solution is to use all of them and combine their decision when solving new problems. The physicians found that the combination of classifiers was the appropriate way of improving the reliability and comprehensibility of diagnostic systems. The combination should be done in an appropriate way and the reliability of each classifier on the given new case should be taken into account.

1.2 Timeline/Evolution

As soon as electronic computers came into use in the 1950s and 1960s, the algorithms were developed that enabled modeling and analyzing large sets of data. From the very beginning, three major branches of machine learning emerged. Classical work in symbolic learning is described by Hunt et al., in statistical methods by Nilsson, and in neural networks by Rosenblatt. Through the years, all three branches developed advanced methods [2]: statistical or pattern recognition methods, such as the k-nearest neighbors, discriminant analysis, and Bayesian classifiers, inductive learning of symbolic rules, such as top down induction of decision trees, decision rules and induction of logic programs, and artificial neural networks, such as the multilayered feedforward neural network with backpropagation learning, the Kohonen's self-organizing network, and the Hopfield's associative memory.

1.3 Organization Of Report

The Report is divided into Five Chapters. In Chapter 2, it is all discussion about Literature Review that covers fundamental of topic, the different types of machine learning algorithms and its pros and cons. In Chapter 3, it is all discussion about Methodology and the construction and working of the models. In Chapter 4, it is all discussion about Result and the applications of ML in medical field. In Chapter 5, it is all discussion about Conclusion and Future work.

Chapter 2

Literature Survey

Machine learning in general consists of two types:

1. Unsupervised Learning Methods(ULM) and
2. Supervised Learning Methods(SLM).

Unsupervised techniques include Principal Component analysis, Partitioning Clustering, Hierarchical Clustering and Density Based Clustering. Supervised Learning methods include Neural Networks, Support Vector Machines, Dicision Trees, Naive Bayes, Random Forest, Nearest Neighbours Linear Regression.

2.1 Fundamentals

2.1.1 Unsupervised Learning Methods (ULM) in Medical Diagnosis

ULM is well known for feature extraction.Principal Component Analysis (PCA) and Clustering algorithms are the two main categories of unsupervised learning methods. Clustering deals with exploring an elite structure from raw database. The prominent clustering algorithms include partitioning, hierarchical and density based clustering methods. PCA is mainly applied for dimensionality reduction, when the trait is recorded in a large number of dimensions.

Bhuvaneswari ng has proposed a medical diagnosis system to predict the risk of heart diseases. The system was built using a combined approach based on PCA and Adaptive NeuroFuzzy Inference System (ANFIS).PCA was used to reduce the dimensionality of heart disease dataset from 13 attributes to 7 attributes. ANFIS was used for classification purpose and accuracy obtained through proposed approach is 93.2%.

Jan L Bruse et al. proposed a model with high F-Score and Matthews correlation coefficient.The input data based on shape was classified automatically using hierarchical clustering. The resulting clusters obtained were with high specificity value and less misclassification rates. Conclusion was clustering results depend on the distance/linkage combination.

Negar Riazifar et al. have analyzed dense based clustering algorithm for retinal vessel segmentation. To identify cluster of random shapes, Density Based Spatial clustering of applications with noise algorithm has been designed. Performance metrics evaluated for the proposed method is 5.36 and 3.82 for specificity and sensitivity respectively.

Siamak et al. have worked on recognizing patterns for visual field loss. Gaussian Mixture model with expectation maximization(GEM) was applied to automatically classify data. Principal Component Analysis was used to subdivide cluster into various axes. The optimal GEM approach classified FDT fields into three clusters. Principal Component Analysis (PCA) was used to subdivide clusters into different axes. PCA detected axes were 2,2 and 5 respectively. Cluster1 has 94 percent specificity and cluster 2,3 combined has 77 percent sensitivity. GEM with PCA successfully separated and detected glaucoma eyes.

2.1.2 Supervised Learning Methods (SLM) in Medical Diagnosis

SLM considers the subjects outcome together with their traits and undergoes a certain training process to find the best possible output associated with the inputs that are closest to the outcomes on average. The training data will be associated with the labels representing classes of observations.

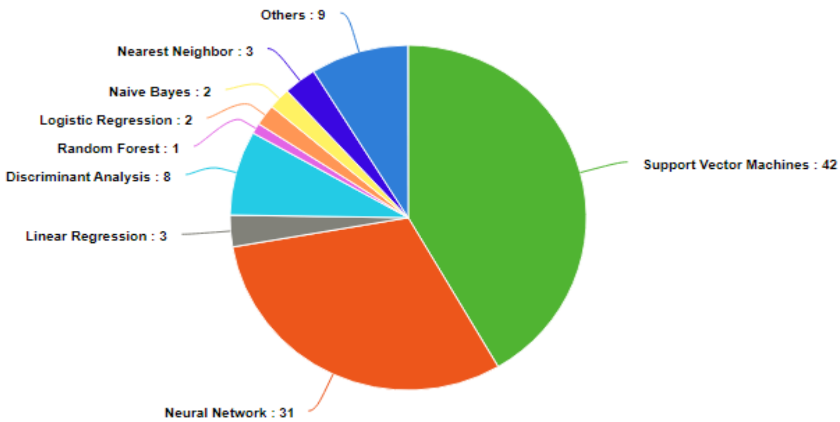

Figure 2.1: use of ML algorithms in medical literature.

Dana Bazazeh et al. have compared three of the most popular ML techniques commonly used for breast cancer detection and diagnosis, namely Support Vector Machine (SVM), Random Forest (RF) and Bayesian Networks (BN). In order to evaluate the performance of three classifiers, the Wisconsin original breast cancer data set was used as a training set. They have evaluated the data on the following metrics accuracy, recall, precision and area of region of convergence.

Allison M Rossetto et al. have presented an ensemble of Convolution Neural Networks (CNN). They have used multiple preprocessing methods to enhance the accuracy of the automated labeling of the scans. Along with a CNN, a voting system is implemented to get the consensus of two networks. The results of proposed method shows both a consistently high accuracy (97.5%) and a low percentage of false positives (10%).

Quanun Liu et al. have proposed an improved method based on C4.5 decision tree and Relief-F attribute weighting techniques. Relief-F algorithm is used to remove irrelevant attributes and C4.5 decision tree is used for classification purpose. Two experiments, one for data on university of California Irvine and other for health examination was conducted.

Mustafa Turan Arslan et al. have worked on Electroencephalography (EEG) data based on silent reading task and mental arithmetic operation. Feature vectors were obtained using discrete wavelet transform. SVM is used to classify the obtained feature vectors. Correlation based algorithm was used for selecting the most constructive channels. The proposed algorithm obtained an accuracy 0f 90.71%, F-measure rate of 90.3% and a precision rate of 91.03% on a mean of 26 channels.

Felipe M Pages et al. have proposed Model Observer based on Naive Bayes classification method. Single-photon emission computed tomography polar map was used to extract features. Then Naive Bayes method was applied over extracted feature set to evaluate perfusion scores. Computation based on Average Model observer achieved good results.

2.2 Related Works

2.2.1 Behavioral modification

It is a predictive analytics approach of study based on person unique genetic profile, that makes them liable to diseases. This domain is currently predominated bysupervised learning. Four forces impacting precision medicine are increasing incidence of cancer, reduced sequencing cost, new cancer drugs and need for faster and more accurate data. Precision medicine is helping to transform healthcare by offering physical therapies in the area of oncology.

Laura Maciejko et al. have proposed a new technique for cancer immunotherapy by considering new technologies to predict clinical efficacy and construct fully humanized platform with immune checkpoint.

2.2.2 Drug Discovery and Development

It includes detection of disease related genomics, identification of targets and validation, lead discovery and lead optimization, preclinical tests.

Lima A N et al. presents an overview of ML applications in design of drug. Major observations are: The ML techniques can be employed in structure based and ligand based drug design studies. The techniques can also be applied for similarity searches, prediction of secondary structures and biological activity.

2.2.3 Clinical Trial Research

It involves devices or procedures to find out how safe the new drugs by testing them in a well-defined plan and protocols. The research also includes the study and evaluation of side-effects and medical benefits.

Diana de et al. have proposed a supervised learning method for classifying Nano or non-Nano categories of clinical trial summaries. The study involves the case data within the registries of the clinicaltrials.gov database. The performance result matched area under curve over 0.95 with the best automated classifier. They also suggested that the presence of nanotechnology products in clinical trial summaries can be easily identified with higher efficiency.

2.2.4 Epidemic Outbreak Prediction

It is the retrospective analysis of infectious diseases. The four main stages involved are to predict, prevent, detect and intervene. Prediction involves determining baseline disease incidence, identifying main drivers and their predictive potential.

Mary Bates et al have discussed about how the advances in machine learning and data processing allow epidemiologists to carefully sift through the millions of digital traces. The machine learning techniques as of now cannot completely replace traditional epidemiological techniques but they are helping to fill critical gaps. ML techniques were used to trace the outbreak and try to cut- off the stem.

2.2.5 Radiotherapy

It is a treatment that uses high-energy rays to treat disease. It can kill cancer cells and shrink tumors. It can be external treatment by using rays or internal treatment by placing radio-active material near cancer cells.

Tong Lin et al. have proposed a novel framework for targeting tumor in lung cancer radiotherapy. A method called gating based on ML algorithm was considered for two-class classification problem. Artificial Neural Network along with Principal Component analysis is used to solve the identified problem. The tracking problem was considered as a regression task. Four regression techniques were evaluated: 1-degree and 2-degree linear regression, SVM, Artificial Neural Network. The proposed algorithm was applied on 10 fluoroscopic image sequences considering 9 patients cases. The precision value ranged from 90% to 99% with mean value of 96.5% for gating algorithm. The localization error obtained was about 2.1 pixels.

2.2.6 Disease Identification

It is the leading most machine learning research in medicine. Development of computers with the capability of identifying and diagnosing diseases has been the biggest breakthrough in medical field.

Yanjia Wang et al. have proposed a statistical method to optimize individualized treatment rules. The approach merges medical domain knowledge with statistical modeling technique, making use of electronic health record. The experiment was conducted on Columbia university clinical dataset. In order to treat type 2 diabetic patients, a decision tree was constructed for best second line therapy.

2.3 Survey Conclusion

Thus the report concludes that machine learning and artificial intelligence has been used extensively in the medicine and healthcare sectors for years. The ML algorithms have a very diverse application in the field and will certainly play a huge role in further development.

Chapter 3

METHODOLOGY

There are various methods used in Machine Learning for medical science. First the report discusses the requirements for an algorithm to perform well in this field and then about each individual method with examples.

3.1 Specific requirements for machine learning systems

For a machine learning (ML) system to be useful in solving medical diagnostic tasks, the following features are desired: good performance, the ability to appropriately deal with missing data and with noisy data (errors in data), the transparency of diagnostic knowledge, the ability to explain decisions, and the ability of the algorithm to reduce the number of tests necessary to obtain reliable diagnosis. In this section, first, there is a discussion of these requirements. Then an overview of the comparison study of six representative machine learning algorithms to illustrate more concretely the points made.

3.1.1 Good performance

he algorithm has to be able to extract significant information from the available data. The diagnostic accuracy on new cases has to be as high as possible. Typically, most of the algorithms perform at least as well as the physicians, and often the classification accuracy of machine classifiers is better than that of physicians when using the same description of the patients. Therefore, if there is a possibility to measure the accuracy of physicians, their performance can be used as the lower bound on the required accuracy of a ML system in the given problem.

In the majority of learning problems, various approaches typically achieve similar performance in terms of the classification accuracy, although in some cases some algorithms may perform significantly better than the others . Therefore, a prior almost none of the algorithms can be excluded with respect to the performance criterion. Rather, several learning approaches should be tested on the available data and the one or few with best estimated performance should be considered for the development of the application.

3.1.2 Dealing with missing data

In medical diagnosis, very often the description of patients in patient records lacks certain data. ML algorithms have to be able to appropriately deal with such incomplete descriptions of patients.

3.1.3 Dealing with noisy data

Medical data typically suffer from uncertainty and errors. Therefore, machine learning algorithms appropriate for medical applications have to have effective means for handling noisy data.

3.1.4 Transparency of diagnostic knowledge

The generated knowledge and the explanation of decisions should be transparent to the physician. She should be able to analyze and understand the generated knowledge. Ideally, the automatically generated knowledge will provide to the physician, a novel point of view on the given problem, and may reveal new interrelations and regularities that physicians did not see before in an explicit form.

3.1.5 Explanation ability

The system must be able to explain decisions when diagnosing new patients. When faced with an unexpected solution to a new problem, the physician shall require further explanation, otherwise she will not seriously consider the system's suggestions. The only possibility for physicians to accept a "black box" classifier is in the situation where such a classifier outperforms by a very large margin all other classifiers, including the physicians themselves in terms of the classification accuracy. However, such situation is typically highly improbable.

3.1.6 Reduction of the number of tests

In medical practice, the collection of patient data is often expensive, time consuming, and harmful for the patients. Therefore, it is desirable to have a classifier that is able to reliably diagnose with a small amount of data about the patients. This can be verified by providing all candidate algorithms with a limited amount of data. However, the process of determining the right subset of data may be time consuming, as it is essentially a combinatorial problem. Some ML systems are themselves able to select an appropriate subset of attributes, i.e. the selection is done during the learning process and may be more appropriate than others that lack this facility.

3.2 Method 1: Naive Bayes

A classier that uses the naive Bayesian formula to calculate the probability of each class C given the values V i of all the attributes for an instance to be classified, assuming the conditional independence of the attributes given the class:

P(C$|V_1,...,V_n) = P(C)\prod_i \frac{P(C|V_i)}{P(C)}$

A new instance is classified into the class with maximal calculated probability. The m-estimate of probabilities makes the naive Bayesian classifier more roboust.

The algorithm is extremely simple but very powerful, and it can provide also comprehensive explanations which was confirmed in long discussions with physicians.It has the efficiency and ability to outperform most advanced and sophisticated algorithms in many medical and also non-medical diagnostic problems. The naive Bayesian classifier became for me a benchmark algorithm that in any medical domain has to be tried before any other advanced method. Other researcher had similar experience. For example, Spiegelhalter et al. were for several man-months, developing an expert system based on Bayesian belief networks for diagnosing the heart disease for newborn babies. The final classification accuracy of the system was 65.5%. When they tried the naive Bayesian classifier, they obtained 67.3%.The theoretical basis for the successful applications of the naive Bayesian classifier (also called simple Bayes) and its variants was developed by 'Good'. We demonstrated the efficiency of this approach in medical diagnosis and other applications. But only inthe early 1990s, the issue of the transparency (in terms of the sum of information gains in favor or against a given decision) of this approach was also addressed and shown successful in the applications in medical diagnosis.

Anjali Gautama et al. have proposed a simple thresholding technique for leukocytes segmentation. Naive Bayes classifier was used for classification purpose. The classification results on training data of 20 images and test images of 68 is about 80.88%.

Advantages of Naive Bayes Classifier:
1.Fast to implement
2.Less model dependency complexity

Disadvantages of Naive Bayes Classifier:
1.No variable
2.Ignores underlying geometry of data

3.3 Method 2: Decision trees

in many other areas, decisions play an important role also in medicine, especially in medical diagnostic processes. Decision support systems helping physicians are becoming a very important part in medical decision making, particularly in those situations where decision must be made effectively and reliably. Since conceptual simple decision making models with the possibility of automatic learning should be considered for performing such tasks, decision trees are a very suitable candidate.They have been already successfully used for many decision making purposes.

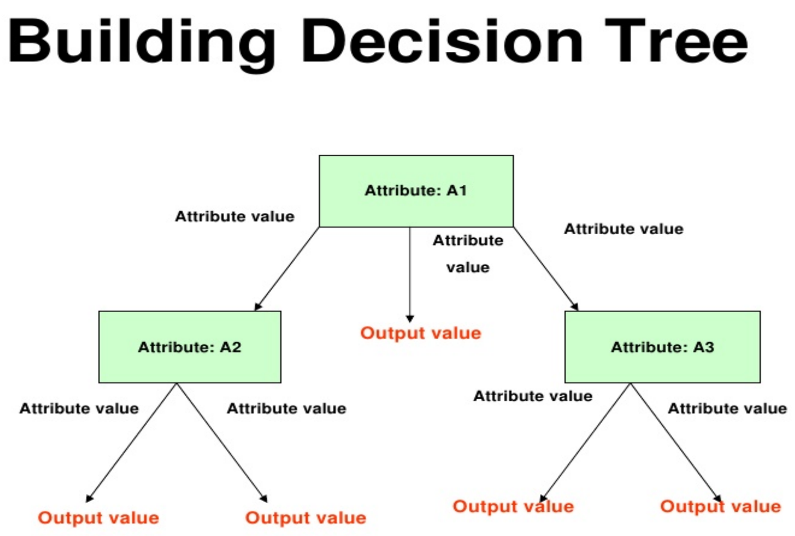

Figure 3.1: building process of Decision trees.

Quanun Liu et al. have proposed an improved method based on C4.5 decision tree and ReliefF attribute weighting techniques. ReliefF algorithm is used to remove irrelevant attributes and C4.5 decision tree is used for classification purpose. Two experiments, one for data on university of California Irvine and other for health examination was conducted.

Advantages of Decision Tree Classifier:

1.Handle large number of features.

2.Easily interpretable.

Disadvantages of Decision Tree Classifier:

1.Tend to over fit.

2.Little training data for lower nodes.

3.4 Method 3: Neural Networks

An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurones) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurones. This is true of ANNs as well.

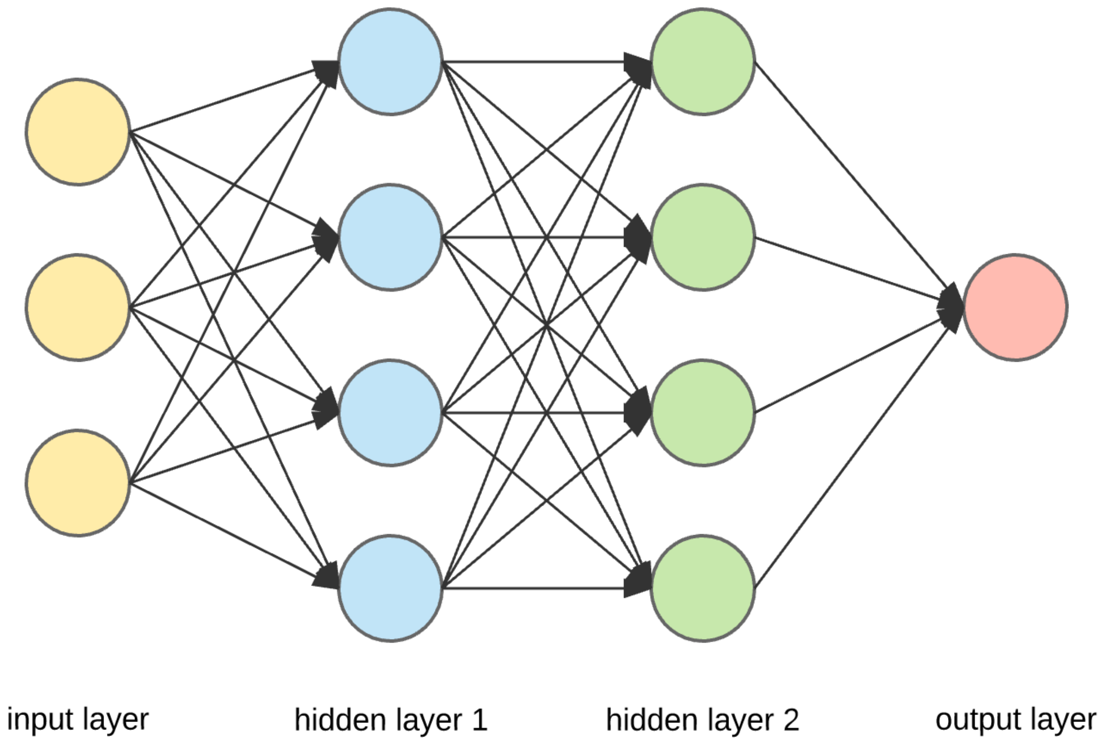

Figure 3.2: Structure of a neural network.

A study by Francis et al. (2015) analyzing post-op data from colorectal cancer surgery patients with a neural network yielded results compliant with standard regression methods and identified two factors behind readmission within the first thirty days after surgery. Another recent study using ANNs to model treatment outcome data for radiotherapy in cancer patients displayed computation advantages for ANN time-series modeling over traditional techniques with applications in quality assurance research (Li et al., 2016). One possible benefit of ANNs over regressions in this area is that while complex statistical knowledge is necessary to run a proper regression, ANNs are easy to operate with little-to-no knowledge of their function, thus making it so more hospitals could take advantage of such technologies (Francis et al., 2015). Clearly, as research continues to progress in the intersection of medicine and technology, there is room for AI in predicting disease progression both before and after treatment - an initiative which has undeniably great clinical utility for cancer patients.

Advantages of Decision Tree Classifier:

1.Handle Noisy data.

2.Detect non-linear time.

Disadvantages of Decision Tree Classifier:

1.Slow training relation.

2.Hard to interpret.

Chapter 4

Discussion

4.1 Issues and Key Challenges

4.1.1 Datasets

Medical datasets fetched directly from electronic health record are extremely complex. Data is highly diverse and integrating data from various databases is a main bottleneck. Availability of large labeled medical dataset is a major issue. Data about rare diseases are depreciated in the data sets, which lead to difficulty in prediction of labels. Training data is expensive because annotation of images is resource intensive in terms of manpower, cost and time. Training data may be uninformative and imperfect because of missed labels and wrongly labeled data. In medical vision, collection of training data based on images helps in better diagnosis, but it may be too expensive.

4.1.2 Privacy and Confidentiality

Major issue in handling medical data is privacy. Data may be misused since it will be in the hands of third parties. Assessing privacy in the realm of ML will be massively challenging because of black-box approach followed. Connecting personal health information electronically poses data to hostile attacks and threats. Information shared under clinical interaction is confidential and must be protected. Access permission to be assigned to users based on role based privileges. Controlling access to health records is a major aspect of medical record security

4.1.3 Data Inaccuracies

Major concerns have been raised about the reliability and accuracy of data maintained in EHR. Incorrect representation of patients present condition and treatment is intolerable. It increases the legal responsibility for clinicians,organizations and high risk for patients. Loss or deletion of data that may occur during data transfer raises another concern about the data accuracy. An increasing problem nowadays is medical identity theft. This kind of theft results in the input of inappropriate information into the victims record.

4.1.4 Accountability

Addressing the accountability, the major question is about "who owns the data?" If we rely on machines to make significant decisions, we require a mechanism to rectify the results that turn out to be unsuitable and incomprehensible. It is important to understand how liability is assigned in case of improper, incorrect medical diagnosis. "Who takes the responsibility for the medical diagnosis directed to a patient by a system employing Machine Learning?"

4.1.5 System Implementation

Health care organizations face major challenges in the phase of implementation of electronic health records. Improper plan of implementation may lead to resource wastage, loss of confidence by patients, financial loss. One of the objectives for the adoption of ML is long term cost saving. But it requires large upfront investments in infrastructures. The planning, implementing and maintenance of EHRs involves sufficient funds, resources and participation of clinicians, technologists, consultants. Maintenance and testing of systems on a routine basis is mandatory for regulating major risk. Training users and physician on new technology is an exorbitant process.

4.1.6 Data Security

With millions of patients personal data storing in electronic systems, data security is utmost importance. Data violation in healthcare has reportedly affected millions of people leading to identity theft, monetary losses and leakage of sensitive personal data to third parties. There are several instances of security violations in global medical industry. Heritage Valley Health systems became the victim of Ransomware attack . Another instance a Bayer MedRad device was attacked by WannaCry ransomware virus.

4.1.7 Threat to Jobs

While emerging ML technologies can improve speed, cost and quality of medical diagnosis, it may also supersede large number of healthcare workers. Surgeons already use automated systems to aid with minimally invasive methods. IBMs Watson manifested that it can diagnose lung cancer from analyzing MRI scans much more efficiently and accurately than real people.

4.2 My views regarding the future role of machine learning in medical diagnosis

- Machine learning technology has not been accepted in the practice of medical diagnosis to an extent that the clearly demonstrated technical possibilities indicate. However, it is hard to expect that this disproportion between the technical possibilities and practical exploitation will remain for very much longer.

- Among the reasons for slow acceptance, perhaps the most reasonable one is that the introduction of machine learning technology will further increase the abundance of tools and instrumentation available to physicians. Any new tool has the undesirable side effect of further increasing the complexity of the physician's work which is already sufficiently complicated. Therefore, machine learning technology will have to be integrated into the existing instrumentation that makes its use as simple and natural as possible.

- Machine learning-based diagnostic programs will be used as any other instrument available to physicians: as just another source of possibly useful information that helps to improve diagnostic accuracy. The final responsibility and judgment whether to accept or reject this information will, as usual, remain with the physician.

- Complementary medicine is becoming in recent years more and more important, which can be seen also by the amount of money people spend on various complementary medicine treatments. Physicians are becoming aware of the efficiency and the benefits of complementary medicine and they need verification procedures in order to acknowledge the benefits and issue licences for the use of complementary approaches. Machine learning can play an important role in this process in particular due to the transparency. of data analysis.

Chapter 5

CONCLUSION

Machine Learning will play a major role in automating tasks that are currently handled by human decisions.ML allows building models by analyzing historical, real time data and delivering better results. ML in healthcare could organize patients data, plan for better treatment and help physicians in decision making. Medical information is sensitive and organizations that handle it need to take care of storage, sharing and access polices. A need of creation of processed and empowered protocols applicable to and mandatory for the whole medical sector.ML enabled systems must be governable, understandable and transparent. Health data requires high level of adaptability because of inherent complication of human body, the receptiveness and congruity of data sources and combining into existing healthcare systems. Diagnosis, Therapy and Prevention are all major issues that depend on large amount of data and their advancement depicts immeasurable value.

References

[1] Rashmi Mothkur Poornima K M, Department of CSE, CMR Institute of Technology, Bangalore, India, 2018 IEEE International Conference on Current Trends toward Converging Technologies "Machine Learning will Transfigure Medical Sector: A Survey."

[2] Igor Kononenko,Faculty of Computer and Information Science, University of Ljubljana " Machine learning for medical diagnosis: history, state of the art and perspective"

[3] Trevor Hastie Robert Tibshirani Jerome Friedman " The Elements of Statistical Learning Data Mining, Inference, and Prediction "

[4] K. Shailaja, M. Tech Scholar, Centre for Data Science, Vardhaman College of Engineering " Machine Learning in Healthcare: A Review"

[5] https://www.kaggle.com/tags/medicine

[6] https://www.kaggle.com/tags/healthcare

[7] https://en.wikipedia.org "Artificial_intelligence_in_healthcare "