Predictive Analytics in Health Care Using Machine Learning Tools and Techniques

B. Nithya
Senior Assistant Professor, Research Scholar
Department of MCA, New Horizon College of Engineering
Bangalore, India
nithya.boopalan@gmail.com

Dr. V. Ilango
Professor, Head of Department
Department of MCA, New Horizon College of Engineering
Bangalore, India
banalysist@yahoo.com

***Abstract* - When we have a huge data set on which we would like to perform predictive analysis or pattern recognition, machine learning is the way to go. Machine Learning (ML) is the fastest rising arena in computer science, and health informatics is of extreme challenge. The aim of Machine Learning is to develop algorithms which can learn and progress over time and can be used for predictions. Machine Learning practices are widely used in various fields and primarily health care industry has been benefitted a lot through machine learning prediction techniques. It offers a variety of alerting and risk management decision support tools, targeted at improving patients' safety and healthcare quality. With the need to reduce healthcare costs and the movement towards personalized healthcare, the healthcare industry faces challenges in the essential areas like, electronic record management, data integration, and computer aided diagnoses and disease predictions. Machine Learning offers a wide range of tools, techniques, and frameworks to address these challenges. This paper depicts the study on various prediction techniques and tools for Machine Learning in practice. A glimpse on the applications of Machine Learning in various domains are also discussed here by highlighting on its prominence role in health care industry.**

Keywords - Predictive Analytics; Machine Learning; Health Care; Prediction Algorithms and Techniques

I. INTRODUCTION

The Industrial Organizations are working with huge amounts of data that have to be recognized with Machine Learning. By gathering insights from these data, organizations are able to work more efficiently as well as gain advantage over their competitors. Innovative predictive models have been applied successfully with Machine Learning algorithms in several domains. Machine learning techniques and its applications are in usage in day to day activities, such us searching, advertisements', YouTube. Health Care informatics, a multi-disciplinary field has become synonymous with the technological advancements and data handling challenges. Medical or Health Informatics is a scientific field that deals with the storage, retrieval, and optimal use of medical information, data, and provides knowledge for problem solving and decision making. Technology in Health has an immense development over the years like advances in information gathering, treatments, communications and research.

II. PREDICTIVE ANALYTICS

Predictive Analytics is the branch of the advanced analytics which is used to make predictions about unknown future events. Predictive analytics applies many techniques from data mining, statistics, modelling, machine learning, and artificial intelligence to investigate current findings to make predictions about future. The predictor is the main entity in predictive analytics which is defined as a variable used for the measurement of future behaviour. With the help of the predictors upcoming probabilities are forecasted with highly reliable results. The approaches used to conduct predictive analytics can be classified into machine learning techniques and regression techniques. Machine learning techniques have become progressively popular in conducting predictive analytics due to their outstanding performance in manage large scale datasets with uniform characteristics and noisy data. Observational studies show that machine learning is appropriate to build predictive models by extracting patterns from large datasets. These models are widely used in predictive data analytics applications like price prediction, risk assessment, predicting customer behaviour, and document classification.

III. PREDICTIVE ANALYTICS IN HEALTH CARE

Predictive Analytics is supporting different segments of health care life sciences and providers. It aims in diagnosing the diseases accurately, enhancement of patient care, resource optimization and also improves clinical outcomes. Predictive Analytics helps organizations to prepare for the health care by optimizing the cost [5]. The accomplishment of predictive analytics in this industry is likely to provide proficient outcome by improving the service quality. Predictive Analytics have the future to transform the health care industry.

IV. MACHINE LEARNING

Machine Learning: the classic definition is - A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E [27]. Machine Learning is a division of artificial intelligence that practices a variety of statistical, probabilistic and optimization techniques that allows computers to learn from prior examples and to detect hard-to-discern patterns

from huge, noisy or complex data sets. Machine Learning is a method of data analysis that automates analytical model building. Through procedures that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look.

A. Machine Learning's Role in Predictions

Enterprises are motivated to find greater meaning in the substantial amounts of data they generate and save every day. Machine learning is provided with effective algorithms, applications, and frameworks to attain greater predictive accuracy and value to enterprises' data sets and contributing to diverse strategies succeeding. Machine learning methodologies are intended to find out the chance to optimize the decisions, based on the predictive value of large-scale data sets. It is evidencing to be effective at handling predictive tasks including defining which behaviors have the maximum tendency to drive preferred outcomes.

B. Steps to apply machine learning to data

Machine learning task can be broken into below steps.

1) Collecting data: Whether the data is written on paper, recorded in text files and spreadsheets, or stored in an SQL database, the data need to be gathered in an electronic format suitable for analysis. This data will serve as the learning material an algorithm uses to generate actionable information.

2) Exploring and preparing the data: The quality of any machine learning project is based largely on the quality of data it uses. This stage in the machine learning process tends to require a great deal of human intervention. An often-cited statistic suggests that 80 percent of the effort in machine learning is dedicated to data. Much of this time is spent learning more about the data and its nuances throughout a practice called data exploration.

3) Training a model on the data: The specific machine learning task will inform the selection of an appropriate algorithm, and the algorithm will represent the data in the form of a model.

4) Evaluating model performance: It is very important to estimate how well the algorithm learned from its past experience, since each machine learning model results in a biased solution to the learning problem. The accuracy of the model can be evaluated using a test dataset, depending on the type of model used

5) Improving model performance: It is necessary to utilize the advanced strategies to augment the performance of the model, if better performance is needed. Every now and then, it may be required to change to a different type of model overall.

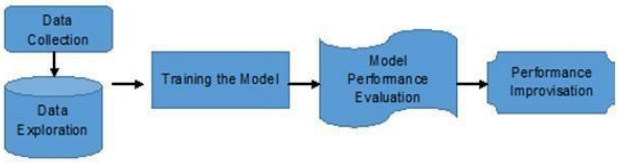

Fig.1. Machine Learning Process

After these steps have been accomplished, if the model appears to be performing acceptably, it can be deployed for its intended task. The model may be applied to provide score data for predictions, for projections of financial data, to generate suitable insight for marketing or research, or to automate tasks. The successes and failures of the deployed model might even provide additional data to train the next generation model.

V. MACHINE LEARNING ALGORITHMS

There are many algorithms available in machine learning and they are classified into three broad categories, depending on the nature of the learning.

The process of selecting a machine learning algorithm involves matching the features of the data to be learned to the biases of the existing approaches. Machine learning algorithms can be divided into two key groups: supervised learners which are used to construct predictive models, and unsupervised learners that are used to build descriptive models.

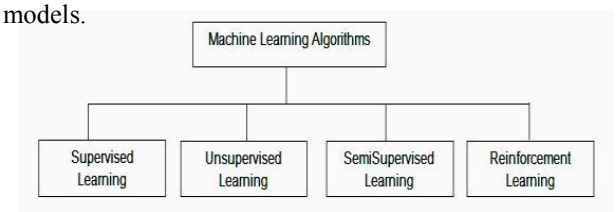

Fig. 2. Types of Machine Learning Process

A. The Supervised Learning/Predictive Models

A predictive model is used for tasks that include the prediction of one value using other values in the dataset. As predictive models are given clear instruction on what they need to learn and how they are intended to learn it, the process of training a predictive model is identified as supervised learning. The aim of supervised, machine learning is to build a model that makes predictions based on evidence in the existence of uncertainty. Exactly, a supervised learning algorithm takes a known set of input data and known responses to the data (output), and trains a model to generate realistic predictions for the response to new data.

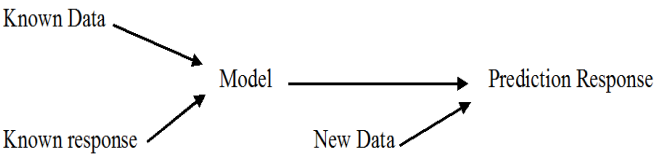

Fig. 3. Supervised Learning

TABLE I. SUPERVISED AND UNSUPERVISED LEARNING ALGORITHMS

S. No.	Type of Learning	Model / Method	Extensively Used Algorithms
1	Supervised Learning	Decision Tree Technique	Classification and Regression Tree (CART)
			Iterative Dichotomiser 3 (ID3)
			C4.5 and C5.0
			Chi-squared Automatic Interaction Detection (CHAID)
			Decision Stump
			M5
			Conditional Decision Trees
		Bayesian Methods	Naive Bayes (NB)
			Gaussian Naive Bayes
			Multinomial Naive Bayes
			Averaged One-dependence Estimators (AODE)
			Bayesian Belief Network (BBN)
			Bayesian Network (BN)
		Artificial Neural Network	Perceptron
			Back-Propagation
			Hopfield Network
			Radial Basis Function Network (RBFN)
		Instance Based Lerning	K - Nearest Neighbour (KNN)
			Learning Vector Quantization (LVQ)
			Self-Organizing Map (SOM)
			Locally Weighted Learning (LWL)
		Ensemble Methods	Boosting
			Bootstrapped Aggregation (Bagging)
			AdaBoost
			Stacked Generalization
			Gradient Boosting Machines (GBM)
			Gradient Boosted Regression Trees (GBRT)
			Random Forest
2	Unsupervised Learning	Clustering Methods	k-Means
			k-Medians
			Expectation Maximization (EM)
			Hierarchical Clustering
3	Supervised / Unsupervised Learning	Regression Algorithms	Ordinary Least Squares Regression (OLSR)
			Linear Regression
			Logistic Regression
			Stepwise Regression
			Multivariate Adaptive Regression Splines (MARS)
			Locally Estimated Scatterplot Smoothing (LOESS)

B. Unsupervised Learning / Descriptive Models

This model is used for tasks that would benefit from the insight gained from summarizing data in new and interesting ways. As contrasting to predictive models that predict a target of interest; in a descriptive model, no single feature is more important than any other. In fact, because there is no goal to learn, the process of training a descriptive model is called unsupervised learning. The aim is to discern the data and get few structures within. Unsupervised learning works fine on transactional data. The methodologies include self-organizing maps, nearest-neighbor mapping, k-means clustering and singular value decomposition are very popular. These algorithms are used to segment text topics, propose items and identify data outliers.

C. Semi Supervised Learning

It is used in all the applications as similar as supervised learning. However, it uses both labelled and unlabeled data for training. Methods such as classification, regression and prediction are using this type of learning. Semi supervised learning is very useful when the cost associated with labelling is too high to allow for a fully labelled training process. Recognizing a person's face on a web cam is one of the examples.

D. Reinforcement Learning (RL)

In this type of learning, machine is trained to take exact decisions based on the business requirement with the objective to maximize the efficiency (performance). This repetitive learning process ensures less participation of human expertise and saves more time. In robotics, gaming and navigation the reinforcement learning is often used. With the help of reinforcement learning, the algorithm discovers through trial and error which actions yield the best rewards.

VI. MACHINE LEARNING TOOLS

Machine learning offers a collection of tools that use computers to convert data into actionable information. Tools are a huge part of machine learning and selecting the right tool can be as important as working with the best algorithms. Machine learning tools make applied machine learning faster, easier. Excellent tools can automate each step in the applied machine learning process by reducing the time.

Three ways to think about machine learning tools are described here.

- Platforms versus Libraries
- Graphical User Interfaces versus Command-Line Interface versus Application Programming Interfaces
- Local versus Remote

A. Platforms versus Libraries:

1) Platform Tools:

- Provide facilities required at each step in a machine learning project.
- The interface may be graphical or command line.
- They afford a loose coupling of features.

- They are provided for general purpose use and exploration rather than speed, scalability or accuracy.

2) Library Tools:

- Provides capabilities for completing part of a machine learning project.
- Provide a specific capability for one or more steps in a machine learning project.
- The interface is normally an application programming interface requiring programming.
- They are customized for a specific use case, problem type or environment.

B. Graphical User Interfaces (GUI) versus Command-Line Interface (CLI) versus Application Programming Interfaces (API):

1) Graphical User Interfaces - GUI:

- Allocate less-technical users to work through machine learning.
- Focus on process and how to get the majority from machine learning techniques.
- Structured process imposed on the user by the interface.
- Major focus on graphical presentations of information such as visualization.

2) Command-Line Interface - CLI:

- Permits technical users who are not programmers to work through machine learning projects.
- Presents many less focused programs or program modes for specific sub-tasks of a machine learning project.
- Frames machine learning tasks in terms of the input required and output to be generated.
- Promotes reproducible results by recording or scripting commands and command line arguments.

3) Application Programming Interfaces - API:

- Used to integrate machine learning into our own software projects.
- Allows creating our own machine learning tools.
- Gives the flexibility to apply our own processes and automations on machine learning projects.
- Allows combining our own methods with those supplied by the library as well as extending provided methods.

C. Local versus Remote:

1) Local Tools: Can be downloaded, installed and run on local environment.

- Tailored for in-memory data and algorithms.
- Control over configuration and parameterization.
- Integrated into our own systems to meet up all the needs.

2) Remote Tools: They can be hosted on a server and called from local environment. These tools are referred to as Machine Learning as a Service (MLaaS).

- Adapted for scale to be run on larger datasets.
- Run across multiple systems, multiple cores and shared memory.
- Fewer algorithms because of the modifications essential for running at scale.
- Simpler interfaces providing less control over run configuration and algorithm parameterization.
- Integrated into our local environment through remote procedure calls.

D. Machine Learning Using R

R is the foremost choice among data professionals who want to recognize and explore data, using statistical methods and graphs. It has several machine learning packages and their advanced implementations for the top machine learning algorithms. R is an open source language and functions in R are grouped into packages which are automatically loaded when R starts. The process of a machine learning project involves the following steps:

- Defining the Problem
- Preparing the Data
- Evaluating the Algorithms
- Improving the Results
- Presenting the Results

Testing and evaluation of any Machine Learning algorithm require variety of datasets to be used. R comes with its own datasets. Various functions are available in R for analyzing and plotting the data. Several Machine Learning algorithms can be implemented in R for the chosen datasets and their results can be compared for improvisations.

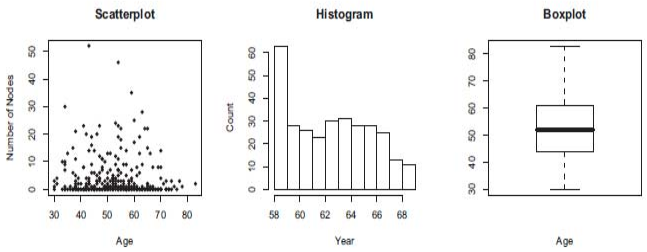

Fig. 4. Plotting Examples in R

In R language, various types of plotting styles are available. The plotting styles like Scatter-Plot, Histogram and Box-Plot are shown in the above diagram as an example. Various features of the data sets can be evaluated and distinguished through these styles.

Construction of predictive models in R can be designed as follows.

- Model creation using basic function.
- Evaluation of the model properties using print, plot and other methods.
- Prediction of outcomes for samples using predict method.

TABLE II. MACHINE LEARNING TOOLS

Category	Tools	Features
Platforms Vs. Libraries	WEKA - Waikato Environment for Knowledge Analysis.	● A modern platform for applied machine learning. ● Allows completing machine learning projects without programming. ● Comprises tools for data pre-processing, classification, regression, clustering, association rules, and visualization.
	R	➢ One of the most efficient and most widespread platforms for statistical programming and applied machine learning. ➢ It is a variant of Lisp ➢ Can parse and execute R scripts that are typed in directly or loaded from a file with .R extension. ➢ Can create graphics to be displayed on the screen or saved to file, also prepare models that can be queried and updated.
	Scikit-learn	● A Python module for machine learning built on top of SciPy. ● Used for math and science work. ● The libraries can be used for interactive "workbench" / embedded into added software and reused.
	Accord	○ The Accord.NET Framework delivers machine learning, mathematics, statistics, computer vision, computer audition, and several scientific computing related methods and techniques to .NET. ○ Contains libraries that provide a more conventional range of machine learning functions.
	JSAT	❖ Java Statistical Analysis Tool with a Java library for Machine Learning.
GUI Vs. CLI Vs. API	Orange	● An appropriate tool for quick comparisons across estimation techniques. ● A slick desktop app where data can be processed through a number of steps and estimation techniques. ● Uses a widget and connector construction on the GUI.
	Waffles	➢ Waffles apps are thin wrappers around functionality in a well-documented C++ class library. ➢ Script-friendly, and are designed to be used with automated processes. ➢ Waffles algorithms can automatically tune their own parameters.
	Deeplearning4j	❖ The main commercial-grade, open-source, distributed deep-learning library written for Java and Scala. ❖ Integrated with Hadoop and Spark. ❖ DL4J is used in business environments on distributed GPUs and CPUs.
Local Vs. Remote	Shogun	● Implemented in C++ and interfaces to Matlab(tm), R, Octave, Java, C#, Ruby, Lua and Python. ● Planned for unified large-scale learning for a wide range of feature types and learning settings.
	GoLearn	✓ Google's Go language has wide collection of libraries. ✓ GoLearn is a 'batteries included' machine learning library for Go. ✓ Created to address the lack of an all-in-one machine learning library for Go.
	Google Prediction API	● Provides pattern-matching and machine learning capabilities. ● After learning from training data, Prediction API can predict a numeric value or choose a category that describes a new piece of data. ● Examples: Guessing what movies or products a user might like, labeling emails as spam or non-spam / guessing how much a user might spend on a given day.
	AWS	Amazon Machine Learning offers visualization tools and wizards that direct through the process of creating ML models without having to learn complex ML algorithms and technology.
	Microsoft Azure	✓ A fully managed cloud service that enables to easily build, deploy and share predictive analytics solutions.

VII. Aid of Machine Learning In varied fields

Machine learning played excessive part in recent years as significant development happened in various fields using it. ML allows machines to make decisions from massive data. Companies such as Google, Amazon, Accenture, Toyota, Hitachi, Tesla, Johnson & Johnson have embraced machine learning at immense scale and improved their products and services. Small companies also developed innovative applications using machine learning. Amazon launched machine learning platform in 2015 and exhibited more supportive reviews to customers, Google used the ML to translate text in 27 languages. Tesla adopted ML in Auto pilot technology [20]. Machine learning is being used in a wide range of application domains and few foretastes are listed here.

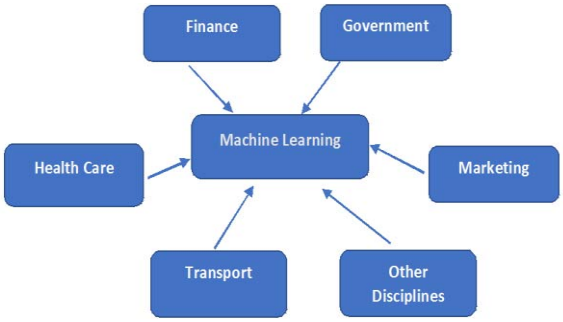

Fig. 5. Machine Learning in Various Disciplines

A. Financial Services and Economics

In financial industries such as banks and other businesses use machine learning methodologies to identify significant insights in data, and to prevent fraud. A group of researchers in the Machine Learning Technologies group at IBM Research - Haifa are taking fraud prevention and detection to a new level. The recommended solution analyzes historical transaction data to build a model that can detect fake patterns. This model is then used to process and examine a large amount of financial transactions as they occur in real time, also known as stream computing.

B. Administration in Government Organizations

Government agencies such as public safety and utilities have a specific need for machine learning since they have multiple sources of data that can be mined for insights. Machine learning can also help to detect fraud and minimize identity theft. In U.K the Government Digital Service (GDS) has been experimenting with special applications such as predicting page views to do anomaly detection and so far, is focusing on signifying the competences of machine learning algorithms on a number of products and prototype services.

C. Health Care Industry

Machine learning is a fast-growing approach in the health care industry. This ML technology can also help medical experts analyze data to identify risks that may lead to improved diagnoses and treatment. IBM research group - The Machine Learning for Healthcare and Life Sciences is evolving and applying machine learning and data mining tools to a group of diverse challenging problems from clinical genomic analysis, through designing clinical decision support systems, to investigate real world evidence for personalized medicine.

D. Promotion, Sales and Marketing

Websites recommend few items to purchase based on the customers' previous purchases. It is using machine learning to analyze our buying history, and promote other items the person would be interested in. This facility to capture data, analyze it and use it to personalize a shopping experience is the future of retail. Microsoft's new cloud based predictive analytics tool called Azure Machine Learning for sales forecasting is in progression and vibrant in this province.

E. Shipping and Transportation

Analyzing data to recognize patterns and trends is a key to the transportation industry, which relies on making routes more efficient and predicting possible problems to increase profitability. Traffic has been rising in major cities around the world given the increase in masses of cars on roads and the slow development of road infrastructure. Research scientist and developer teams at Microsoft Research pioneered the use of machine learning methods to build predictive models for traffic. The work focused early on to prototypes that can infer and predict the flow of traffic at different times into the future based on the scrutiny of large amounts of data on traffic over months and years.

VIII. Machine Learning In Health Care Predictions

Machine learning algorithms are effective in spotting complex patterns in massive data. This competence is well-suited to medical applications, especially those that depend on complex proteomic and genomic measurements. Hence machine learning is frequently used in various disease diagnosis and detection. In medical applications machine learning algorithms can produce better decisions about treatment plans for patients by means of providing effective healthcare system. Machine learning plays a key role in many radiology applications [32]. It also identifies complex patterns automatically and helps radiologists make intelligent decisions on radiology data such as conventional radiographs, CT, MRI, and PET images and radiology reports. Observational studies have shown that various types of disease predictions were carried out using different types of machine learning prediction techniques. In recent times diagnosis and prediction on various diseases like Cardiovascular diseases, cancers, Diabetes, Hepatitis Asthma, Tuberculosis (TB) and Blood Pressure Monitoring have been carried out using various machine learning predictions techniques. Few glimpses on prediction models applied in major diseases through the practice of machine learning are discussed here.

A. Predictions on Cardio Vascular Diseases

Cardiovascular disease (CVD) is a class of diseases that involve the heart or blood vessels. Cardiovascular disease comprises coronary artery diseases (CAD) such as angina and myocardial infarction (heart attack), stroke, hypertensive heart

disease, rheumatic, cardiomyopathy, atrial fibrillation, congenital heart disease, endocarditis, aortic aneurysms, peripheral artery disease and venous thrombosis. Diagnosis and prediction of heart diseases is very much essential as it is one of the leading causes of death all over the world [31]. Enormous works on prediction of heart diseases are carried out in health care research. The two most effective tools, neural networks and genetic algorithms presented a technique for prediction of heart disease using major risk factors like age, family history, diabetes, hypertension, cholesterol, smoking, alcohol intake, obesity or physical inactivity [3]. The methods Naive Bayes, Neural network and Decision tree algorithms are used on analysis of medical data sets with respect to heart disease prediction [19]. A prototype Intelligent Heart Disease Prediction System (IHDPS) has been developed using Decision Trees, Naive Bayes and Neural Network. Results exhibited that each technique has its unique strength in realizing the objectives of the defined mining goals [28].

B. Diabetes Predictions

Diabetes mellitus is a chronic disease and a major public health challenge worldwide. It develops when there are high blood sugar levels over a prolonged period. Recently, it has been quoted as a risk factor for developing Alzheimer, and a leading cause for blindness & kidney failure. Many techniques have been discovered to find the causes of diabetes and cure it. Some of the predictions on diabetes is stated here.

The research on diabetes prediction discussed about establishing a relationship between diabetes risk likely to be developed from a person's daily lifestyle activities such as his/her eating habits, sleeping habits, physical activity along with other indicators like BMI (Body Mass Index), waist circumference [4]. A Chi-Squared Test of Independence was performed followed by application of the CART (Classification and Regression Trees) machine learning algorithm on the data. The analysis about the performance of the Naïve Bayes classifier, RBF network and SVM Classifier in the heart, cancer and diabetes datasets shows that SVM classifier produces better percentage of accuracy in classification [21]. The work has been executed in WEKA environment and attained results show that SVM is the most robust and effective classifier for medical data sets.

C. Hepatitis Disease Prediction

Hepatitis means an injury to the liver with inflammation of the liver cells. The illness can be self-limiting or can progress to fibrosis, cirrhosis or liver cancer. Hepatitis viruses are the most common source of hepatitis in the world but other infections and toxic substances can also cause hepatitis. There are five core hepatitis viruses, stated to as types A, B, C, D and E. Diagnosis of hepatitis disease is conducted with a machine learning methods proposed a novel machine learning method that hybridizes support vector machine (SVM) and Simulated Annealing (SA) [22]. Simulated annealing is a stochastic method currently in wide use for optimization.

There was a focus on the aspect of Medical diagnosis by learning pattern through the collected data of hepatitis and to develop intelligent medical decision support systems to help the physicians [30]. This work anticipated the use of decision trees C4.5 algorithm, ID3 algorithm and CART algorithm to categorize the diseases. Few studies proposed the ways to obtain accurate prediction results and more informative knowledge about the Hepatitis C Virus (HCV) protein cleavage sites using Decision tree algorithm [1].

D. Cancer Predictions Using Machine Learning

Cancer is the universal name for a group of more than 100 diseases. While there are many kinds of cancers, all cancers start as abnormal cells grow out of control. Over the past decades, a continuous evolution related to cancer research has occurred concerning the prognosis and prediction of cancers like lung, skin, liver and stomach cancers [2, 24]. Several data mining and machine learning techniques are used for cancers prediction and some of the works are considered at this point.

In recent years of research, a few optimized models on breast cancer diagnosis and recurrence prediction were developed by using various risk prediction algorithms and techniques [17, 7]. There are reviews about various researches made to predict cancer prognosis across cancer types and subtypes [16]. This review suggests to validate a method across multiple sets of patients and conclusion is to understand which data sets are most informative for a particular cancer type/subtype. Various studies described the practical application of data mining and machine learning methods for estimation of survival rate and disease relapse for breast cancer patients [13]. A comparative study of prominent machine learning models was carried out where Naive Bayes classifier is selected as a model for prognosis of cancer survivability on the basis of the 5 years' survival rate, while the Artificial Neural Network has achieved the best performance in prognosis of breast cancer recurrence [18]. A research study implemented machine learning techniques like Decision Tree (C4.5), Support Vector Machine (SVM), and Artificial Neural Network (ANN) to develop the predictive models for recurrence prediction in breast cancer. The main idea of this work was to relate the performance of these three well-known algorithms on the data through sensitivity, specificity, and accuracy. This work concluded that SVM classification model predicted breast cancer recurrence with least error rate and highest accuracy.

The study which applied three machine learning approaches support vector machine, C5.0 and extreme learning machine to find important risk factors to predict the recurrence-proneness for cervical cancer in small number of patients illustrated that C5.0 model is the most useful approach to the discovery of recurrence proneness factors [12]. To diagnose recurrence in cervical cancer, probabilistic model method has been used. In cervical cancer recurrence prediction research, MARS and C5.0 classifier techniques were used, and the results showed that decision tree model is better than MARS [6, 11].

In recent times diagnosis and prediction on further diseases like Asthma, Tuberculosis (TB) and Blood Pressure Monitoring have been carried out using various data mining and machine learning predictions techniques [25, 5, 23].

IX. CONCLUSION

In the past two decades Machine Learning has been transformed from an academic curiosity to a multi-billion-dollar industry, and a point of focus for our economic, social, scientific, and security infrastructure. Abundant work in machine learning has drawn interest on research in optimization, motivated by large-scale applications requiring analysis of massive high-dimensional data.

This study shows that Machine Learning has given medical providers new tools to work with, novel ways to practice medicine. It also confirms that machine learning tools and techniques are decisive in health care province and exclusively used in the diagnosis and predictions of various types of cancers. There are a lot of open problems and future challenges in dealing with massive amounts of heterogeneous, distributed, diverse, highly dynamic data sets and increasingly large amounts of unstructured and non-standardized information with respect to varied types of cancers. Some of the most important challenges in clinical practice and biomedical research include the need to develop and apply novel tools for the effective integration, analysis and interpretation of complex biomedical data with the aim to identify testable hypothesis, and build accurate models to diagnose and predict various types of cancers and their recurrences. The algorithms must also handle incomplete, noisy, even contradictory/ambiguous information. Therefore, efficient machine learning approaches becomes essential in the health care industry to address these challenges. Machine Learning and predictive analytics techniques could revolutionize the entire healthcare industry by providing accurate insights and predictions related to symptoms, diagnoses, procedures and medications.

REFERENCES

[1] Ahmed Mohamed Samir, ―A Data Mining Approach for the Prediction of Hepatitis C Virus Protease Cleavage Sites, International Journal of Advanced Computer Science and Applications, Vol. 2, No. 12, December 2011.

[2] Akutekwe, Arinze, Huseyin Seker, and Sunday Iliya. "An optimized hybrid dynamic Bayesian network approach using differential evolution algorithm for the diagnosis of Hepatocellular Carcinoma." Adaptive Science & Technology (ICAST), 2014 IEEE 6th International Conference, IEEE, 2014.

[3] Amin, S.U.; Agarwal, K.; Beg, R., "Genetic neural network based data mining in prediction of heart disease using risk factors", IEEE Conference on Information & Communication Technologies (ICT), 2013, vol., no.,pp.1227-31,11-12 April 2013.

[4] Anand, Ayush, Shakti, Divya, "Prediction of diabetes based on personal lifestyle indicators," 1st International Conference on Next Generation Computing Technologies (NGCT), vol., no., pp.673-676, 4-5 Sept. 2015.

[5] B Nithya, "Study on Predictive Analytics Practices in Health Care System", IJETTCS, Vol. 5, Feb. 2016.

[6] Bahl, Ravinder, S. K. Spolia, and Chandra Mauli Sharma. "Predicting recurrence in cervical cancer patients using clinical feature analysis." British Journal of Medicine and Medical Research 6.9 (2015): 908-917.

[7] Bhardwaj, Arpit, and Aruna Tiwari. ―Breast cancer diagnosis using genetically optimized neural network model, Expert Systems with Applications 42.10 (2015): 4611-4620.

[8] Brett Lantz, ―Machine Learning with R, Packt Publishing Ltd, 2013, ISBN 978-1-78216-214-8.

[9] Cancer Informatics 2014:13(S2) 19–28 doi: 10.4137/CIN.S13788.

[10] Charles R. Farrar, Keith Worden, ―Structural Health Monitoring, A Machine Learning Perspective‖, John Wiley & Sons, Ltd., Publication, ISBN: 978-1-119-99433-6.

[11] Chi-Chang Chang, Sun-Long Cheng, Chi-Jie Lu and Kuo-Hsiung Liao, ―Prediction of Recurrence in Patients with Cervical Cancer Using MARS and Classification‖, International Journal of Machine Learning and Computing, Vol. 3, No. 1, February 2013.

[12] Chih-Jen Tseng, Chi-Jie Lu, Chi-Chang Chang, Gin-Den Chen, - Application of machine learning to predict the recurrence-proneness for cervical cancer‖, Neural Computing and Applications, Springer, May 2014, Vol.24, Issue 6, pp 1311-1316.

[13] Cirkovic, Bojana R.Andjelkovic; Cvetkovic, Aleksandar M, Ninkovic, Srdjan M, Filipovic, Nenad D., "Prediction models for estimation of survival rate and relapse for breast cancer patients", IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE), vol., no., pp.1-6, 2-4 Nov. 2015.

[14] Daniel Combs, Safal Shetty, Sairam Parthasarathy, - Big-Data or Slim-Data: Predictive Analytics Will Rule with World, J Clin Sleep Med 16.

[15] Daniel Pop, Gabriel Iuhasz, - Overview of Machine Learning Tools and Libraries.

[16] Das, Jishnu, Kaitlyn M. Gayvert, and Haiyuan Yu. "Predicting cancer prognosis using functional genomics data sets." Cancer informatics 13. Suppl. 5 (2014): 85.

[17] Drukker, Caroline A, "Optimized outcome prediction in breast cancer by combining the 70-gene signature with clinical risk prediction algorithms", Breast cancer research and treatment 145.3(2014): 697-705.

[18] Eshlaghy et al., ―Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence‖, J Health Med Inform 2013, 4:2.

[19] Gandhi M, Singh, S.N, "Predictions in heart disease using techniques of data mining", 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE), vol., no., pp.520-525, 25-27 Feb. 2015.

[20] https://www.research.ibm.com/

[21] Janardhanan, Padmavathi L, Heena, Sabika, Fathima, - Effectiveness of Support Vector Machines in Medical Data mining, Journal of Communications Software & Systems, Mar2015, Vol. 11 Issue 1, p25-30. 6p.

[22] Javad Salimi Sartakhti, Mohammad Hossein Zangooei, Kourosh Mozafari, - Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA), Computer methods and programs in biomedicine, Volume 108, Issue 2, November 2012, Pages 570–579, Elsevier.

[23] Kaur et al., Using hybrid models to predict blood pressure reactivity to unsupported back based on anthropometric characteristics, Front Inform. Technology Electron. Eng. 2015.

[24] Konstantina Kourou, Themis P. Exarchos, Konstantinos P. Exarchos, Michalis V. Karamouzis, Dimitrios I. Fotiadis, - Machine learning applications in cancer prognosis and prediction, Elsevier, Computational and Structural Biotechnology Journal 13 (2015) 8–17.

[25] Luo et al., ―Predicting asthma control deterioration in children, BMC Medical Informatics and Decision Making (2015) 15:84.

[26] M. I. Jordan, T. M. Mitchell, ―Machine learning: Trends, perspectives, and prospects, Science 349, 255 (2015).

[27] Machine Learning, Tom Mitchell, McGraw Hill, 1997.

[28] Palaniappan S, Awang R, "Intelligent heart disease prediction System using data mining techniques," IEEE/ACS International Conference on Computer Systems and Applications, AICCSA 2008., vol., no., pp.108-115, March 31 2008-April 4 2008.

[29] S. Muthukaruppan, ―A hybrid particle swarm optimization based fuzzy expert system for the diagnosis of coronary artery disease‖, Expert Systems with Applications, (2012) 11657–1166.

[30] Sathyadevi G, "Application of CART algorithm in hepatitis disease diagnosis", International Conference on Recent Trends in Information Technology (ICRTIT), vol., no., pp.1283-1287, 3-5 June 2011.

[31] Shanthi Mendis, Pekka Puska, Bo Norrving, World Health Organization (2011), Global Atlas on Cardiovascular Disease Prevention and Control, PP. 3–18. ISBN 978-92-4-156437-3.

[32] Shijun Wang, Ronald M. Summers, ―Machine Learning and Radiology, Med Image Anal. 2012 July; 16(5): 933–951.