**HEART DISEASE DETECTION USING DEEP LEARNING**

**ABSTRACT**

Heart related diseases or Cardiovascular Diseases (CVDs) are the main reason for a huge number of death in the world over the last few decades and has emerged as the most life-threatening disease, not only in India but in the whole world. So, there is a need of reliable, accurate and feasible system to diagnose such diseases in time for proper treatment. Machine Learning algorithms and techniques have been applied to various medical datasets to automate the analysis of large and complex data. Many researchers, in recent times, have been using several machine learning techniques to help the health care industry and the professionals in the diagnosis of heart related diseases.

Heart is the next major organ comparing to brain which has more priority in Human body. It pumps the blood and supplies to all organs of the whole body. Prediction of occurrences of heart diseases in medical field is significant work. Data analytics is useful for prediction from more information and it helps medical centre to predict of various disease. Huge amount of patient related data is maintained on monthly basis. The stored data can be useful for source of predicting the occurrence of future disease. Some of the data mining and machine learning techniques are used to predict the heart disease, such as Artificial Neural Network (ANN), Random Forest, and Support Vector Machine (SVM).

Prediction and diagnosing of heart disease become a challenging factor faced by doctors and hospitals both in India and abroad. In order to reduce the large scale of deaths from heart diseases, a quick and efficient detection technique is to be discovered. Data mining techniques and machine learning algorithms play a very important role in this area. The researchers accelerating their research works to develop a software with the help machine learning algorithm which can help doctors to take decision regarding both prediction and diagnosing of heart disease. The main objective of this research project is predicting the heart disease of a patient using machine learning algorithms.

**INTRODUCTION**

Heart is an important organ of the human body. It pumps blood to every part of our anatomy. If it fails to function correctly, then the brain and various other organs will stop working, and within few minutes, the person will die. Change in lifestyle, work related stress and bad food habits contribute to the increase in rate of several heart related diseases.

Heart diseases have emerged as one of the most prominent cause of death all around the world. According to World Health Organisation, heart related diseases are responsible for the taking 17.7 million lives every year, 31% of all global deaths. In India too, heart related diseases have become the leading cause of mortality [1]. Heart diseases have killed 1.7 million Indians in 2016, according to the 2016 Global Burden of Disease Report, released on September 15,2017. Heart related diseases increase the spending on health care and also reduce the productivity of an individual. Estimates made by the World Health Organisation (WHO), suggest that India have lost up to $237 billion, from 2005-2015, due to heart related or Cardiovascular diseases [2]. Thus, feasible and accurate prediction of heart related diseases is very important.

Medical organisations, all around the world, collect data on various health related issues. These data can be exploited using various machine learning techniques to gain useful insights. But the data collected is very massive and, many a times, this data can be very noisy. These datasets, which are too overwhelming for human minds to comprehend, can be easily explored using various machine learning techniques. Thus, these algorithms have become very useful, in recent times, to predict the presence or absence of heart related diseases accurately

The usage of information technology in health care industry is increasing day by day to aid doctors in decision making activities. It helps doctors and physicians in disease management, medications and discovery of patterns and relationships among diagnosis data. Current approaches to predict cardiovascular risk fail to identify many people who would benefit from preventive treatment, while others receive unnecessary intervention. Machine-learning offers opportunity to improve accuracy by exploiting complex interactions between risk factors. We assessed whether machine-learning can improve cardiovascular risk prediction.

“Cardiovascular disease is the leading cause of illness and death worldwide,” said Dr. Stephen Weng, of Nottingham University’s National Institute for Health Research School [1]. “Our study shows that artificial intelligence could significantly help in the fight against it by improving the number of patients accurately identified as being at high risk and allowing for early intervention by doctors to prevent serious events like cardiac arrest and stroke.” Based on their results, it is clear that artificial intelligence and machine learning techniques have a key role in fine-tuning risk management strategies for individual patients.

In today's opportunity at numerous spots clinical test outcomes are regularly made in light of specialists' instinct and experience as opposed to on the rich data accessible in numerous expansive databases. Numerous a times this procedure prompts inadvertent predispositions, lapses and a tremendous medicinal expense which influences the nature of administration gave to patients.

Today numerous doctor's facilities introduced some kind of quiet's data frameworks to man-age their social insurance or patient information. These data frameworks commonly produce a lot of information which can be in distinctive organization like numbers, content, diagrams and pictures yet sadly, this database that contains rich data is once in a while utilized for clinical choice making. Like business knowledge and examination, the term information mining can mean diverse things to distinctive individuals. In exceptionally straightforward way we can characterize information mining as this is the investigation of substantial information sets to discover examples and utilize those examples to foresee or fore-cast the probability of future occasions. The motivation to do this problem comes from World Health Organization estimation.

According to the World Health Organization estimation till 2030, very nearly 23.6 million individuals will pass on because of Heart malady. So to minimize the danger, expectation of coronary illness ought to be finished. Analysis of coronary illness is typically in view of signs, manifestations and physical examination of a patient. The most troublesome and complex assignment in medicinal services area is finding of right ailment. This colossal entirety huge of rough data is the rule resource that can be capably pre-taken care of and inspected for key information extraction that direct or by suggestion influences the remedial society for cost sufficiency and reinforce decision making. Authentic determination of coronary sickness can't be possible by using simply human understanding. There are heaps of parameters that can impacts the accurate conclusion like less exact results, less experience, time subordinate execution, data up degree and whatnot. Packs of headway and examination happened in this field using multi-parametric qualities with nonlinear and direct parts of Heart Rate Variability (HRV).A novel framework was proposed by Heon Gyu Lee et al. To fulfill this, various experts have used various classifiers e.g. CMAR (Classification Multiple Association Rules), SVM (Support Vector Machine), Bayesian Classifiers and C4.5). A latest's rate techniques in this field depicted. Some plausible strategies and technique we recommended incorporates the clinical information institutionalization, examination and the information sharing over the related industries to improve the precision & viability of information mining applications in social insurance. It is likewise prudent to investigate the utilization of content digging and picture digging for extension the nature and extent of information mining applications in medicinal services part. Information mining application can likewise be investigated on computerized indicative pictures for application viability. Some advancement has been made in these areas.

There is a lot of data put away in stores that can be utilized viably to guide a medical practitioners in decision making in human services. This brings up an essential issue:

"By what means would we be able to transform information into helpful data that can empower medicinal services practitioners to settle on viable clinical decision?"

**REVIEW OF LITERATURE**

In late time, numerous associations in human services division utilizes data mining applications seriously and broadly on substantial scale. In information mining we can utilize diverse master cess and innovation to change this colossal measures of information into helpful data for solid and exact choice making. Another reason is that the social insurance exchanges created by this part are excessively voluminous and perplexing, making it impossible to be broke down and prepared by customary systems. Choice using so as to make can be enhanced majorly by using

mining applications in finding patterns and examples in substantial volumes of ordinary data. In late patterns investigation on these extensive dataset has gotten to be fundamental because of monetary weights on medicinal services commercial enterprises. This separated data can be utilized for choices making taking into account the relapse examination of restorative and money related information. Learning extraction can impact industry working proficiency, income and expense maintaining so as to utilize learning disclosure from database with at most care. Research demonstrates that on the off chance that we utilizes information mining applications as a part of social insurance organizations then these associations would be in better position to meet their fleeting objectives and long haul needs, Benko and Wilson argue. We can get extremely valuable results from human services crude. information by changing crude information into helpful data. An extraordinary reason that empowers analysts in this

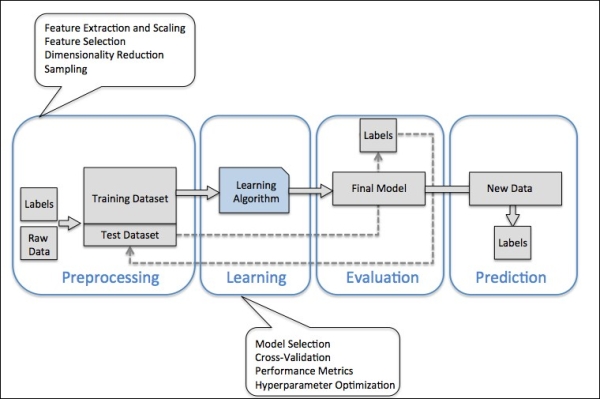
field is that this is exceptionally helpful for all partner included in the human services division. Like, in the event that we consider Insurance supplier, they can identify misuse and extortion, expert in human services can pick up help with choices making, similar to in client relationship administration. Social insurance suppliers (doctor's facilities, doctor, test research centres and patient and so forth.) can likewise utilize information mining applications in their separate master zone for master choice finding so as to make for instance, best practices and right & compelling medicines.

Heart Disease risk level prediction

The Heart disease database contains the screening clinical information of heart patients. At first, the database preprocessed to make the mining handle more able.

**What is Machine Learning:**

Machine Learning is a branch of [**artificial intelligence**](https://www.techsparks.co.in/artificial-intelligence-as-an-m-tech-thesis-topic-for-cse/) that gives systems the ability to learn automatically and improve themselves from the experience without being explicitly programmed or without the intervention of human. Its main aim is to make computers learn automatically from the experience.

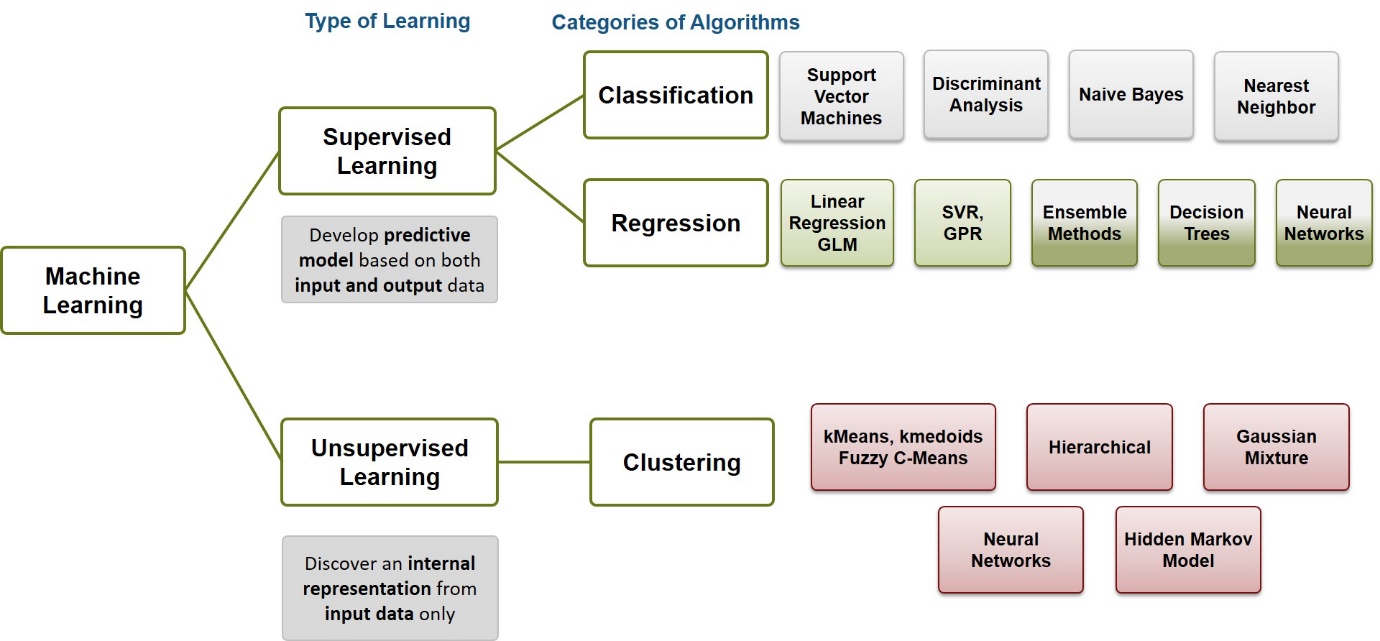


**Requirements of creating good machine learning systems**

So what is required for creating such machine learning systems? Following are the things required in creating such machine learning systems:

* **Data –**Input data is required for predicting the output.
* **Algorithms –**Machine Learning is dependent on certain statistical algorithms to determine data patterns.
* **Automation –**It is the ability to make systems operate automatically.
* **Iteration –**The complete process is iterative i.e. repetition of process.
* **Scalability –**The capacity of the machine can be increased or decreased in size and scale.
* **Modeling –**The models are created according to the demand by the process of modeling.

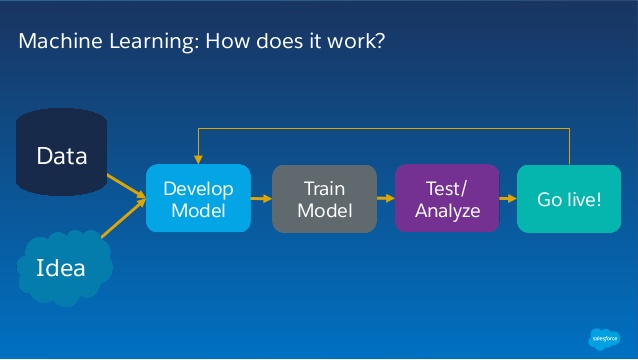
**Methods of Machine Learning**



Machine Learning methods are classified into certain categories. These are:

* **Supervised Learning –**In this method, input and output is provided to the computer along with feedback during the training. The accuracy of predictions by the computer during training is also analyzed. The main goal of this training is to make computers learn how to map input to the output.
* **Unsupervised Learning –**In this case, no such training is provided leaving computers to find the output on its own. Unsupervised learning is mostly applied on transactional data. It is used in more complex tasks. It uses another approach of iteration known as deep learning to arrive at some conclusions.
* **Reinforcement Learning –**This type of learning uses three components namely – agent, environment, action. An agent is the one that perceives its surroundings, an environment is the one with which an agent interacts and acts in that environment. The main goal in reinforcement learning is to find the best possible policy.

**How does machine learning work?**



Machine learning makes use of processes similar to that of data mining. Machine learning algorithms are described in terms of target function(f) that maps input variable (x) to an output variable (y). This can be represented as:

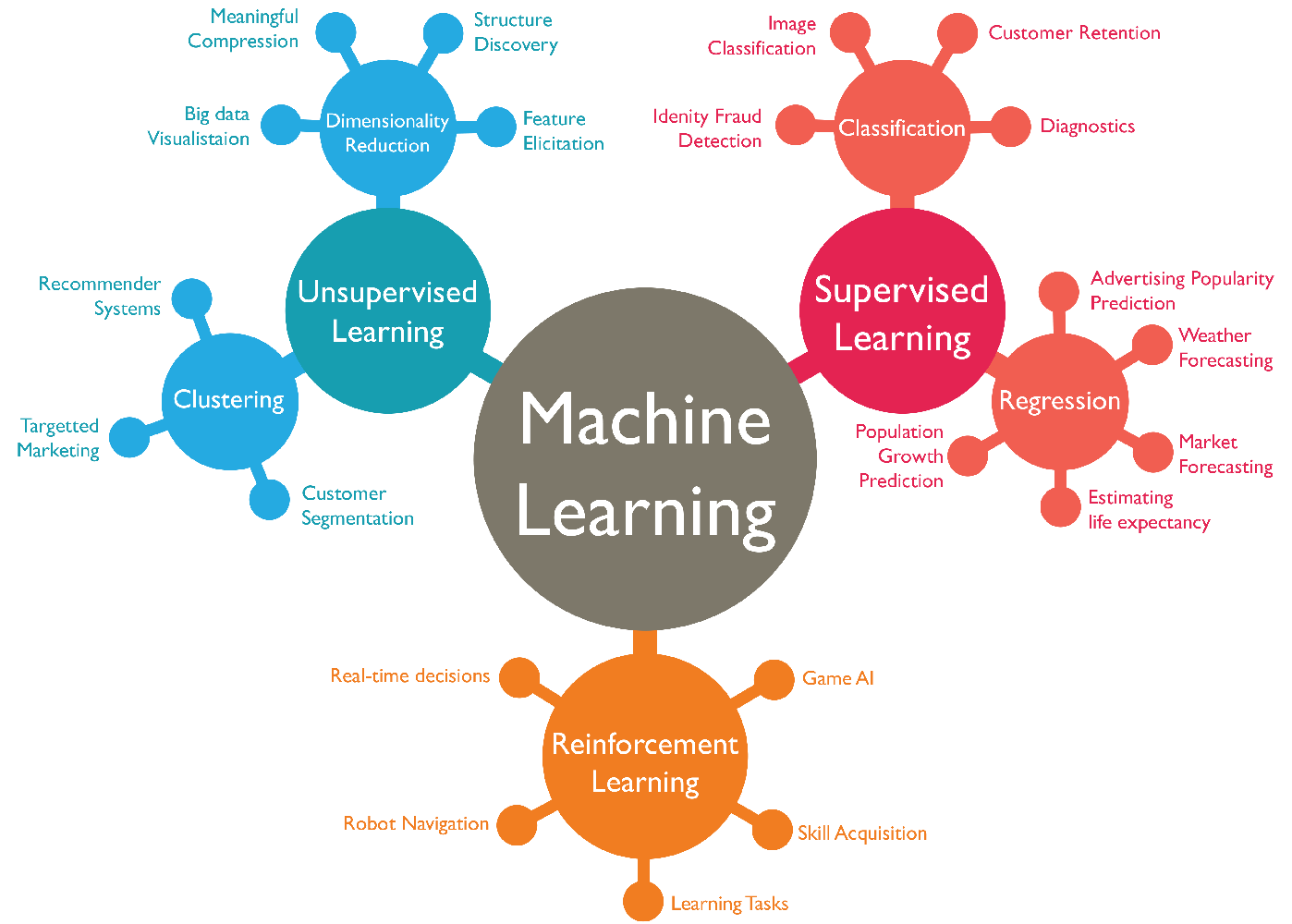
y=f(x)

There is also an error e which is the independent of the input variable x. Thus the more generalized form of the equation is:

y=f(x) + e

In machine the mapping from x to y is done for predictions. This method is known as predictive modeling to make most accurate predictions. There are various assumptions for this function.

**Benefits of Machine Learning**



Everything is dependent on machine learning. Find out what are the benefits of machine learning.

* **Decision making is faster –**Machine learning provides the best possible outcomes by prioritizing the routine decision-making processes.
* **Adaptability –**Machine Learning provides the ability to adapt to new changing environment rapidly. The environment changes rapidly due to the fact that data is being constantly updated.
* **Innovation –**Machine learning uses advanced algorithms that improve the overall decision-making capacity. This helps in developing innovative business services and models.
* **Insight –**Machine learning helps in understanding unique data patterns and based on which specific actions can be taken.
* **Business growth –**With machine learning overall business process and workflow will be faster and hence this would contribute to the overall business growth and acceleration.
* **Outcome will be good –**With machine learning the quality of the outcome will be improved with lesser chances of error.

**Branches of Machine Learning**

**Computational Learning Theory –**Computational learning theory is a subfield of machine learning for studying and analyzing the algorithms of machine learning. It is more or less similar to supervised learning.

**Adversarial Machine Learning –**Adversarial machine learning deals with the interaction of machine learning and computer security. The main aim of this technique is to look for safer methods in machine learning to prevent any form of spam and malware. It works on the following three principles:

* Finding vulnerabilities in machine learning algorithms.
* Devising strategies to check these potential vulnerabilities.
* Implementing these preventive measures to improve the [**security**](https://www.techsparks.co.in/tools-and-technologies/ns2/) of the algorithms.

**Quantum Machine Learning –**This area of machine learning deals with quantum physics. In this algorithm, the classical data set is translated into quantum computer for quantum information processing. It uses Grover’s search algorithm to solve unstructured search problems.

**Predictive Analysis –**Predictive Analysis uses statistical techniques from data modeling, machine learning and data mining to analyze current and historical data to predict the future. It extracts information from the given data. Customer relationship management(CRM) is the common application of predictive analysis.

**Robot Learning –**This area deals with the interaction of machine learning and robotics. It employs certain techniques to make robots to adapt to the surrounding environment through learning algorithms.

**Grammar Induction –**It is a process in machine learning to learn formal grammar from a given set of observations to identify characteristics of the observed model. Grammar induction can be done through genetic algorithms and greedy algorithms.

**Meta-Learning –**In this process learning algorithms are applied on meta-data and mainly deals with automatic learning algorithms.

**Best Machine Learning Tools**

Here is a list of artificial intelligence and machine learning tools for developers:

1. **ai-one –**It is a very good tool that provides software development kit for developers to implement artificial intelligence in an application.
2. **Protege –**It is a free and open-source framework and editor to build intelligent systems with the concept of ontology. It enables developers to create, upload and share applications.
3. **IBM Watson –**It is an open-API question answering system that answers questions asked in natural language. It has a collection of tools which can be used by developers and in business.
4. **DiffBlue –**It is another tool in artificial intelligence whose main objective is to locate bugs, errors and fix weaknesses in the code. All such things are done through automation.
5. **TensorFlow –**It is an open-source software library for machine learning. TensorFlow provides a library of numerical computations along with documentation, tutorials and other resources for support.
6. **Amazon Web Services –**Amazon has launched toolkits for developers along with applications which range from image interpretation to facial recognition.
7. **OpenNN –**It is an open-source, high-performance library for advanced analytics and is written in C++ programming language. It implements neural networks. It has a lot of tutorials and documentation along with an advanced tool known as Neural Designer.
8. **Apache Spark –**It is a framework for large-scale processing of data. It also provides a programming tool for deep learning on various machines.
9. **Caffe –**It is a framework for deep learning and is used in various industrial applications in the area of speech, vision and expression.
10. **Veles –**It is another deep learning platform written in C++ language and make use of python language for interaction between the nodes.

**Machine Learning Applications**

Following are some of the applications of machine learning:

* Cognitive Services
* Medical Services
* Language Processing
* Business Management
* Image Recognition
* Face Detection
* Video Games
* Computer Vision
* Pattern Recognition

**Machine Learning in Bioinformatics**

Bioinformatics term is a combination of two terms bio, informatics. Bio means related to biology and informatics means information. Thus bioinformatics is a field that deals with processing and understanding of biological data using computational and statistical approach. Machine Learning has a number of applications in the area of bioinformatics. Machine Learning find its application in the following subfields of bioinformatics:

**Genomics –**Genomics is the study of DNA of organisms. Machine Learning systems can help in finding the location of protein-encoding genes in a DNA structure. Gene prediction is performed by using two types of searches named as extrinsic and intrinsic. Machine Learning is used in problems related to DNA alignment.

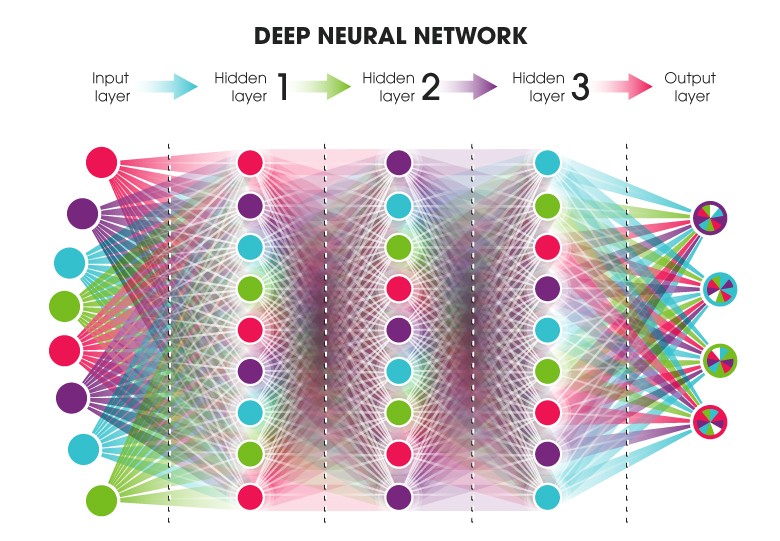
**Proteomics –**Proteomics is the study of proteins and amino acids. Proteomics is applied to problems related to proteins like protein side-chain prediction, protein modeling, and protein map prediction.

**Microarrays –**Microarrays are used to collect data about large biological materials. Machine learning can help in the data analysis, pattern prediction and genetic induction. It can also help in finding different types of cancer in genes.

**System Biology –**It deals with the interaction of biological components in the system. These components can be DNA, RNA, proteins and metabolites. Machine Learning help in modeling these interactions.

**Text mining –**Machine learning help in extraction of knowledge through natural language processing techniques.

**Deep Learning**



Deep Learning is a part of the broader field machine learning and is based on data representation learning. It is based on the interpretation of artificial neural network. Deep Learning algorithm uses many layers of processing. Each layer uses the output of previous layer as an input to itself. The algorithm used can be supervised algorithm or unsupervised algorithm. Deep Learning is mainly developed to handle complex mappings of input and output. It is another hot topic for M.Tech thesis and project along with machine learning.

**Deep Neural Network**

Deep Neural Network is a type of Artificial Neural Network with multiple layers which are hidden between the input layer and the output layer. This concept is known as feature hierarchy and it tends to increase the complexity and abstraction of data. This gives network the ability to handle very large, high-dimensional data sets having millions of parameters. The procedure of deep neural networks is as follows:

1. Consider some examples from a sample dataset.
2. Calculate error for this network.
3. Improve weight of the network to reduce the error.
4. Repeat the procedure.

**Applications of Deep Learning**

Here are some of the applications of Deep Learning:

* Automatic Speech Recognition
* Image Recognition
* Natural Language Processing
* Toxicology
* Customer Relationship Management
* Bioinformatics
* Mobile Advertising

**Advantages of Deep Learning**

Deep Learning helps in solving certain complex problems with high speed which were earlier left unsolved. Deep Learning is very useful in real world applications. Following are some of the main advantages of deep learning:

* **Eliminates unnecessary costs** – Deep Learning helps to eliminate unnecessary costs by detecting defects and errors in the system.
* **Identifies defects which otherwise are difficult to detect –**Deep Learning helps in identifying defects which left untraceable in the system.
* **Can inspect irregular shapes and patterns –**Deep Learning can inspect irregular shapes and patterns which is difficult for machine learning to detect.

From this introduction, you must have known that why this topic is called as hot for your M.Tech thesis and projects. This was just the basic introduction to machine learning and deep learning. There is more to explore in these fields. You will get to know more once you start doing research on this topic for your M.Tech thesis. You can get thesis assistance and guidance on this topic from experts specialized in this field.

Thesis and Research Topics in Machine Learning

Here is the list of current thesis and research topics in Machine Learning:

* Machine Learning Algorithms
* Computer Vision
* Supervised Machine Learning
* Unsupervised Machine Learning
* Deep Learning
* Neural Networks
* Reinforcement Learning
* Predictive Learning
* Bayesian Network
* Data Mining

**Machine Learning Algorithms**

For starting with Machine Learning, you need to know some algorithms. Machine Learning algorithms are classified into three categories which provide the base for machine learning. These categories of algorithms are supervised learning, unsupervised learning, and reinforcement learning. The choice of algorithms depends upon the type of tasks you want to be done along with the type, quality, and nature of data present. The role of input data is crucial in machine learning algorithms.

**Computer Vision**

Computer Vision is a field that deals with making systems that can read and interpret images. In simple terms, computer vision is a method of transmitting human intelligence and vision in machines. In computer vision, data is collected from images which are imparted to systems. The system will take action according to the information it interprets from what it sees.

**Supervised Machine Learning**

It is a good topic for machine learning masters thesis. It is a type of machine learning algorithm in which makes predictions based on known data-sets. Input and output is provided to the system along with feedback. Supervised Learning is further classified into classification and regression problems. In the classification problem, the output is a category while in regression problem the output is a real value.

**Unsupervised Machine Learning**

It is another category of machine learning algorithm in which input is known but the output is not known. Prior training is not provided to the system as in case of supervised learning. The main purpose of unsupervised learning is to model the underlying structure of data. Clustering and Association are the two types of unsupervised learning problems. k-means and Apriori algorithm are the examples of unsupervised learning algorithms.

**Deep Learning**

Deep Learning is a hot topic in Machine Learning. It is already explained above. It is a part of the family of machine learning and deals with the functioning of the artificial neural network. Neural Networks are used to study the functioning of the human brain. It is one of the growing and exciting field. Deep learning has made it possible for the practical implementation of various machine learning applications.

**Neural Networks**

Neural Networks are the systems to study the biological neural networks. It is an important application of machine learning and a good topic for masters thesis and research. The main purpose of Artificial Neural Network is to study how the human brain works. It finds its application in computer vision, speech recognition, machine translation etc. Artificial Neural Network is a collection of nodes which represent neurons.

**Reinforcement Learning**

Reinforcement Learning is a category of machine learning algorithms. Reinforcement Learning deals with software agents to study how these agents take actions in an environment in order to maximize their performance. Reinforcement Learning is different from supervised learning in the sense that correct input and output parameters are not provided.

**Predictive Learning**

Predictive Learning is another good topic for thesis in machine learning. In this technique, a model is built by an agent of its environment in which it performs actions. There is another field known as predictive analytics which is used to make predictions about future events which are unknown. For this, techniques like data mining, statistics, modeling, machine learning, and artificial intelligence are used.

**Bayesian Network**

It is a network that represents probabilistic relationships via Directed Acyclic Graph(DAG). There are algorithms in Bayesian Network for inference and learning. In the network, a probability function is there for each node which takes an input to give probability to the value associated with the node. Bayesian Network finds its application in bioinformatics, image processing, and computational biology.

**Data Mining**

[Data Mining](https://www.techsparks.co.in/data-mining-its-process-and-techniques/) is the process of finding patterns from large data-sets to extract valuable information to make better decisions. It is a hot area of research. This technology use method from machine learning, statistics, and database systems for processing. There exist data mining techniques like clustering, association, decision trees, classification for the data mining process.

**DATASET**

We performed computer simulation on one dataset. Dataset is a Heart dataset. The dataset is available in UCI Machine Learning Repository [10]. Dataset contains 303 samples and 14 input features as well as 1 output feature. The features describe financial, personal, and social feature of loan applicants. The output feature is the decision class which has value 1 for Good credit and 2 for Bad credit. The dataset-1 contains 700 instances shown as Good credit while 300 instances as bad credit. The dataset contains features expressed on nominal, ordinal, or interval scales. A list of all those features is given in Table

**Database Details**

**(a) Database Creators:**

V.A. Therapeutic Center, Long Beach and Cleveland Clinic Foundation: Robert Detrano,

M.D., Ph.D.

**(b) Database Donor:**

David W. Aha (aha@ics.uci.edu) (714) 856-8779

The registry contains a database related with coronary disease. Data can be collected from uci.

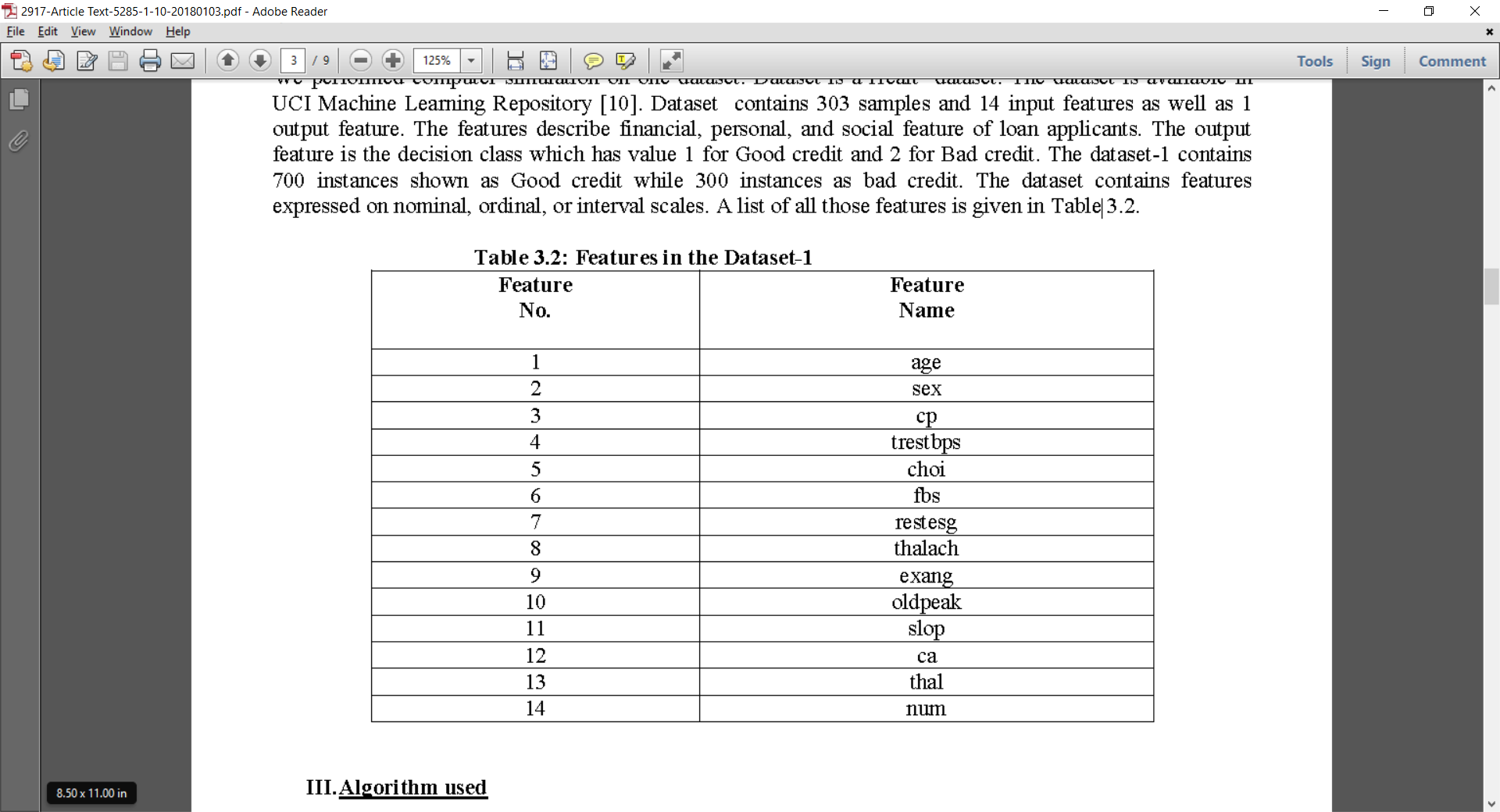
Cleveland Clinic Foundation (Cleveland. Data) [12].

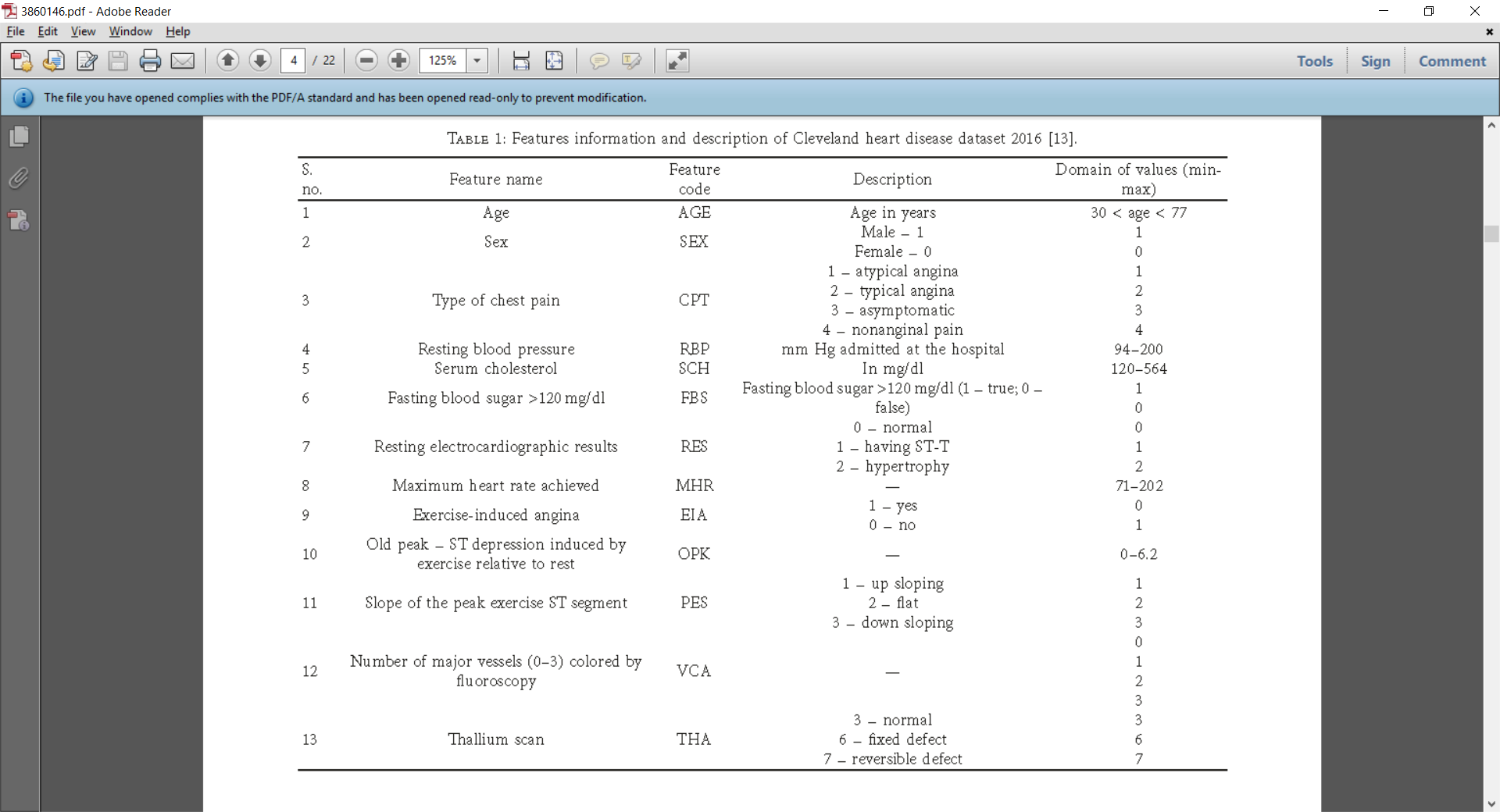
**Inputs attributes**

Age, Sex, Chest Pain, Resting blood pressure, Serum cholesterol, Fasting blood sugar, Resting electrocardiographic results, Maximum heart rate achieved, Exercise induced angina, ST depression, Slope of the peak exercise ST segment, Number of major vessels colored by fluoroscopy and thal.

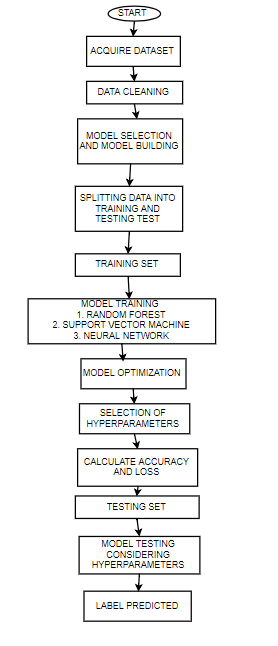
**Outputs class attribute**

num (the predicted attribute)





**PROPOSED METHODOLOGY**

****

**Random Forest**

Random forest is a concept falling under the general technique of random decision. This

algorithm operates by creating a group of decision trees at training time and outputting the

class that represents the mode of classes or the mean prediction of the individual trees.

Individual decision trees are generated using a random selection of attributes at each node to

determine split. During classification, each tree casts a vote and the most popular class is

returned. Using the Random forests, the variance can be reduced by averaging the deep

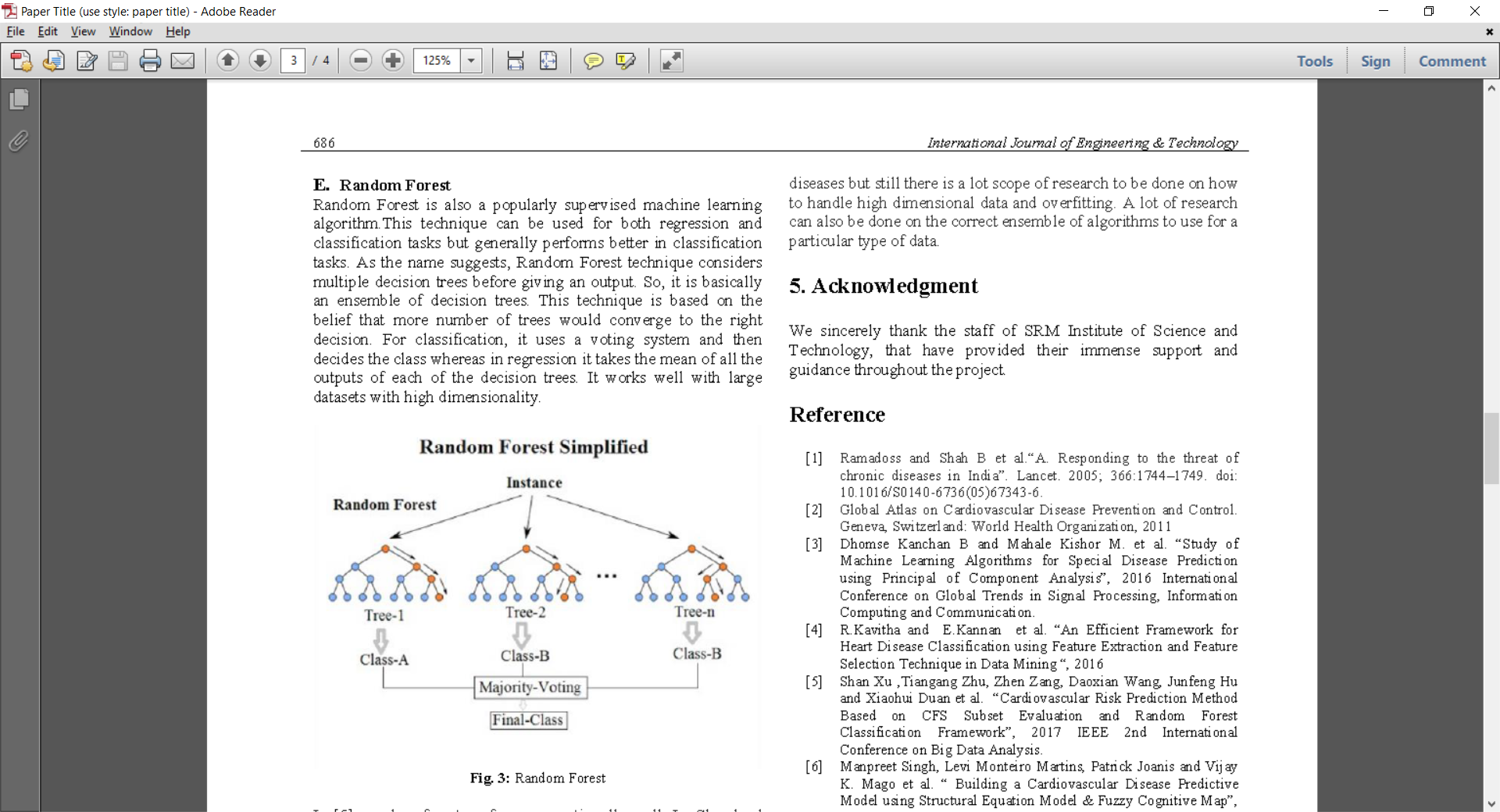
decision trees trained with different parts of the training set. To form Random forests, tree

predictors should be integrated in a way that each tree be dependent on the values of a random

vector sampled independently and uniformly from all trees in the forest. We use this

approach to predict flight delays in our database.

Random Forest is also a popularly supervised machine learning algorithm.This technique can be used for both regression and classification tasks but generally performs better in classification tasks. As the name suggests, Random Forest technique considers multiple decision trees before giving an output. So, it is basically an ensemble of decision trees. This technique is based on the belief that more number of trees would converge to the right decision. For classification, it uses a voting system and then decides the class whereas in regression it takes the mean of all the outputs of each of the decision trees. It works well with large datasets with high dimensionality



**Support Vector Machines (SVMs)**

Support vector machines exist in different forms, linear and non-linear. A support vector machine is a supervised classifier. What is usual in this context, two different datasets are involved with SVM, training and a test set. In the ideal situation the classes are linearly separable. In such situation a line can be found, which splits the two classes perfectly. However not only one line splits the dataset perfectly, but a whole bunch of lines do. From these lines the best is selected as the "separating line".

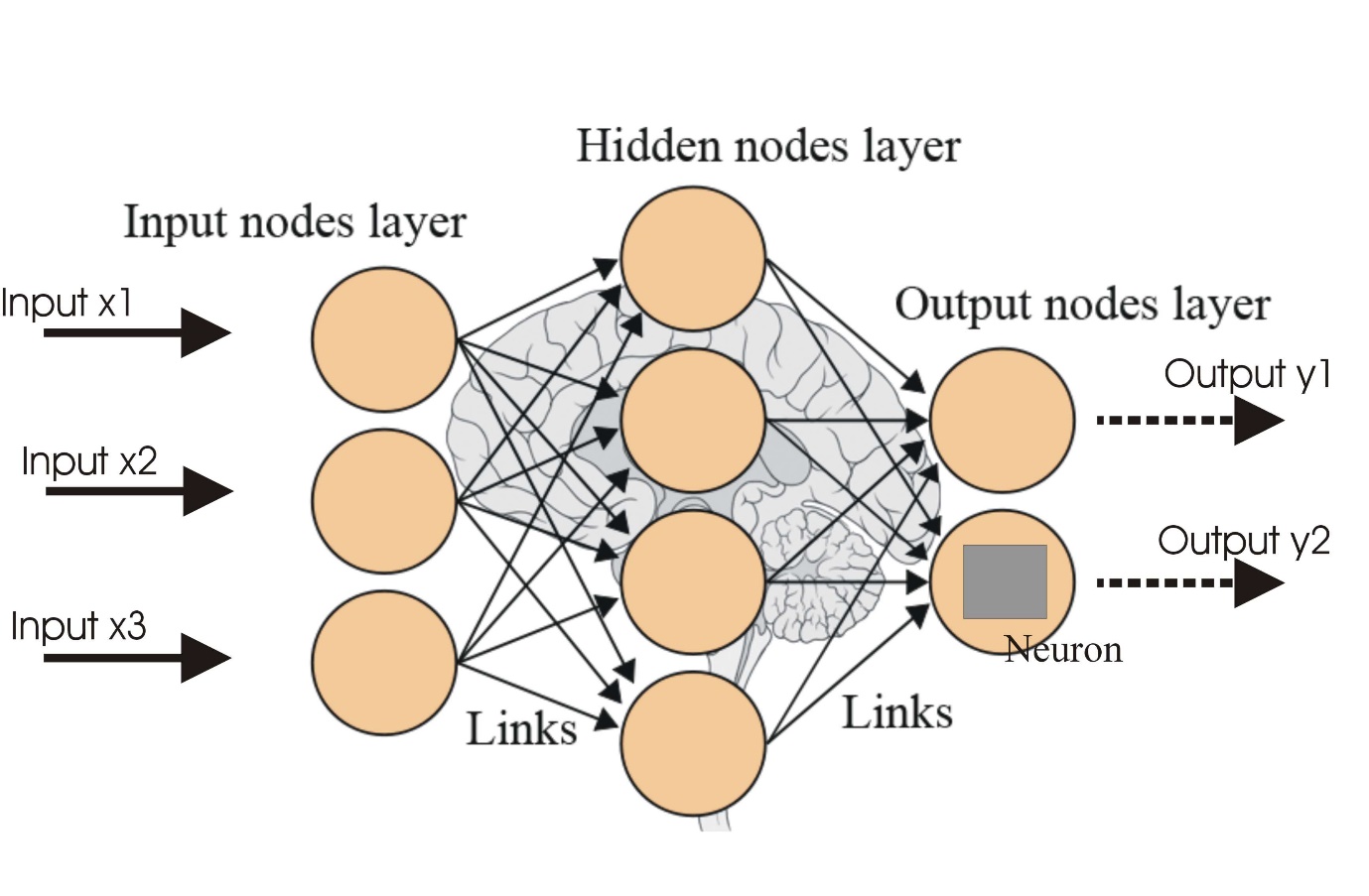
The best line is found by maximizing the distance to the nearest points of both classes in the training set. The maximization of this distance can be converted to an equivalent minimization problem, which is easier to solve. The data points on the maximal margin lines are called the support vectors. Most often datasets are not nicely distributed such that the classes can be separated by a line or higher order function. Real datasets contain random errors or noise which reates a less clean dataset. Although it is possible to create a model that perfectly separates the data, it is not desirable, because such models are over-fitting on the training data.

Overfitting is caused by incorporating the random errors or noise in the model. Therefore the model is not generic, and makes significantly more errors on other datasets. Creating simpler models keeps the model from over-fitting. The complexity of the model has to be balanced between fitting on the training data and being generic. This can be achieved by allowing models which can make errors. A SVM can make some errors to avoid over-fitting. It tries to minimize the number of errors that will be made. Support vector machines classifiers are applied in many applications. They are very popular in recent research. This popularity is due to the good overall empirical performance. Comparing the naive Bayes and the SVM classifier, the SVM has been applied the most

**Artificial** **Neural Network**

These are used to model/simulate the distribution, functions or mappings among variables as modules of a dynamic system associated with a learning rule or a learning algorithm. The modules here simulate neurons in nervous system and hence ANN collectively refers to the neuron simulators and their synapsis simulating interconnections between these modules in different layers [6]. The defining aspect of an ANN is the function implemented at each neuron and the learning algorithm for the dynamic weights assigned to the interconnections among neurons. What makes ANN stand apart is its ability to simulate human thought process coupled with continuous learning, growth and evolution. Also it is capable of handling large number of parameters and large set of data with noise and yet achieves high accuracy.

Neural Network is built by stacking together multiple neurons in layers to produce a final output. First layer is the input layer and the last is the output layer. All the layers in between is called hidden layers. Each neuron has an activation function. Some of the popular Activation functions are Sigmoid, ReLU, tanh etc. The parameters of the network are the weights and biases of each layer. The goal of the neural network is to learn the network parameters such that the predicted outcome is the same as the ground truth. Back-propagation along loss-function is used to learn the network parameters.

 Figure shows the neural network we used in this project. As shown in the figure, our neural network consists of a single hidden layer with

four neurons. We used Sigmoid activation function for neurons in both hidden and output layer, and a binary cross-entropy loss function.

**SOFTWARE USED:-**

**Python**

To collect data a web scraper programmed in Python was used. According to Wikipedia Python’s syntax allows programmers to express concepts in fewer lines of codes.

Guido van Rossum at CWI in the Netherlands started Python’s implementation in December 1989. Python 2.0 was released on October 16th 2000 and Python 3.0 was released December 3rd 2008.

Python is intended to be a highly-readable language. It is designed to have an uncluttered visual layout, frequently using English keywords where other languages use punctuation. Furthermore, Python has a smaller number of syntactic exceptions and special cases than C or Pascal. Python uses whitespace indentation.

Why use Python for web scraping and not another thing? Python offers a module called ‘urllib2’, which has suitable functions to open websites and extract information easily.

Python is used to program the web scraper that is in charge of collecting the weather data for the model.

**MS Excel**

Microsoft Excel is a spreadsheet application developed by Microsoft for Windows and Mac OS X. It features calculation, graphing tools, pivot tables and a macro-programming language. The first version was released in 1987.

Why choose MS Excel versus another similar type of software? MS Excel is a very complete spreadsheet application tool, which supports almost any kind of file extension, and it has a lot of features. Its user-friendly interface helps you most of the time. However, if this doesn’t seem enough, I will say that, apart from the typical things a normal user would do in Excel (Charts, Calculation…), it enables you to use the VBA language to create functions to use on the spreadsheets you’ve created. Excel can also be used as if it were an SQL database as was explained in a previous chapter. Having said this, for me it is the perfect program.

MS Excel is used a lot throughout the project, to visualize the data and perform cleaning tasks on it.

**Conclusion**

This project provides the deep insight into machine learning techniques for classification of heart diseases. The role of classifier is crucial in healthcare industry so that the results can be used for predicting the treatment which can be provided to patients. The existing techniques are studied and compared for finding the efficient and accurate systems. Machine learning techniques significantly improves accuracy of cardiovascular risk prediction through which patients can be identified during an early stage of disease and can be benefitted by preventive treatment.

It can be concluded that there is a huge scope for machine learning algorithms in predicting cardiovascular diseases or heart related diseases. Each of the above-mentioned algorithms have performed extremely well in some cases but poorly in some other cases.

**REFERENCES**

[1] Prerana T H M1, Shivaprakash N C2 , Swetha N3 ”Prediction of Heart Disease Using Machine Learning

Algorithms- Naïve Bayes,Introduction to PAC Algorithm, Comparison of Algorithms and HDPS”

International Journal of Science and Engineering Volume 3, Number 2 – 2015 PP: 90-99 ©IJSE Available at

www.ijse.org ISSN: 2347-2200

[2] B.L Deekshatulua Priti Chandra “Classification of Heart Disease Using K- Nearest Neighbor and

Genetic Algorithm “ M.Akhil jabbar\* International Conference on Computational Intelligence: Modeling

Techniques and Applications (CIMTA) 2013.

[3] Michael W.Berry et.al,”Lecture notes in data mining”,World Scientific(2006)

[4] S. Shilaskar and A. Ghatol, “Feature selection for medical diagnosis : Evaluation for cardiovascular

diseases,” Expert Syst. Appl., vol. 40, no. 10, pp. 4146–4153, Aug. 2013.

[5] C.-L. Chang and C.-H. Chen, “Applying decision tree and neural network to increase quality of

dermatologic diagnosis,” Expert Syst. Appl., vol. 36, no. 2, Part 2, pp. 4035–4041, Mar. 2009.

[6] A. T. Azar and S. M. El-Metwally, “Decision tree classifiers for automated medical diagnosis,” Neural

Comput. Appl., vol. 23, no. 7–8, pp. 2387–2403, Dec. 2013. [10] Y. C. T. Bo Jin, “Support vector machines

with genetic fuzzy feature transformation for biomedical data classification.,” Inf Sci, vol. 177, no. 2, pp.

476–489, 2007.

[7] N. Esfandiari, M. R. Babavalian, A.-M. E. Moghadam, and V. K. Tabar, “Knowledge discovery in

medicine: Current issue and future trend,” Expert Syst. Appl., vol. 41, no. 9, pp. 4434–4463, Jul. 2014.

[8] A. E. Hassanien and T. Kim, “Breast cancer MRI diagnosis approach using support vector machine and

pulse coupled neural networks,” J. Appl. Log., vol. 10, no. 4, pp. 277–284, Dec. 2012.

[9] Sanjay Kumar Sen 1, Dr. Sujata Dash 21Asst. Professor, Orissa Engineering College, Bhubaneswar,

Odisha – India.

[10]. UCI Machine Learning Repository, Available at http://archive.ics.uci.edu/ml/machinelearningdatabases/

statlog/german/

[11] Domingos P and Pazzani M. “Beyond Independence: Conditions for the Optimality of the Simple

Bayesian Classifier”, in Proceedings of the 13th Conference on Machine Learning, Bari, Italy, pp105-112,

1996.

[12] Elkan C. “Naive Bayesian Learning, Technical Report CS97-557”, Department of Computer Science

and Engineering, University of California, San Diego, USA, 1997.

[13] B.L Deekshatulua Priti Chandra “Reader, PG Dept. Of Computer Application North Orissa University,

Baripada, Odisha – India. “Empirical Evaluation of Classifiers’ Performance Using Data Mining Algorithm”