

Survey on Machine Learning Methods of Data mining for Medical Dataset

Khushbu Verma (Research Scholar)¹,
Narender Kumar (Assistant Professor)²

¹Department of Computer Science and Engineering, H.N.B. Garhwal University, Srinagar Garhwal, Uttrakhand, India

²Department of Computer Science and Engineering, H.N.B. Garhwal University, Srinagar Garhwal, Uttrakhand, India
(E-mail: khushbuverma422@gmail.com)

Abstract- Machine Learning plays an important role in Computer Technology and Artificial Intelligence. Machine Learning algorithms of Data Mining can reduce the human effort in recognizing, learning, predictions and many more areas. In the Dataset of Medical Diseases, there are challenge to prediction, classification and find the certain pattern to identify the growth and spread of medical diseases on the basis of certain criteria. Data mining can be used to analysis the Dataset of Medical Disease. Different methods of Machine Learning to analysis the Dataset of different Medical Diseases are compared in this paper.

Keyword- data mining; medical disease; artificial intelligence; supervised; unsupervised; machine learning

I. INTRODUCTION

Computer is become more understanding through Artificial Intelligence. Machine Learning is a part of AI. According to multiple researcher intelligence may not improved without learn process. In medical sector are generated large amount of multiple disease data from daily treatment. Here is requiring on cleaning these data so that we can retrieve important & useful knowledge from this data. The goal of data mining is to retrieve important knowledge from large amount of heterogeneous database. Machine learning algorithm of data mining applied in various diseases of medical data which play an important role in diagnosis & prediction of the multiple diseases. Database is a aggregation of multiple meaningful data which put into organized form & in such a way that a computer program immediately & easily pick need part of the data. A traditional database is well structured form into field, record & file. The field show single part of information, record is show a whole set of fields & file is show a group of records. A database management system should be capability to handle data. Retrieve some significant knowledge from huge amount of various resource of database that process is call data mining technique. In other word Data mining is a process of mining knowledge from large amount of various heterogeneous databases.

II. MACHINE LEARNING METHODS

Machine Learning method of Data mining: - How to maintain the data on maximum efficiency by train machine

learning algorithms. Few times after looking the data we can't retrieve knowledge from that data so here we apply machine learning algorithms to mining similar information [27]. Machine learning algorithm has following types: -

1. Supervised Machine Learning: In supervised learning here use an algorithm on available input variable (A) & output variable (B) and classify data using learn by example. This input data behaves as a teacher which supervising these iterative learning process until not reach suitable performance [1] [2]. Ex. Classification, Regression

2. Unsupervised Machine Learning: In unsupervised learning only available input variable here not provide output variable or target variable. Unsupervised learning technique explore the similarities between input variable & based on this similarities of input variable classify the data [15] [29] [30]. Ex. Clustering, Association

3. Semi - supervised Machine Learning: Semi - supervised Machine Learning technique have used the combination of Supervised & Unsupervised Machine Learning technique. Here available on less amount of input variable and large amount of output variable. This technique applies an algorithm on these available data variable to classify the data [18] [19].

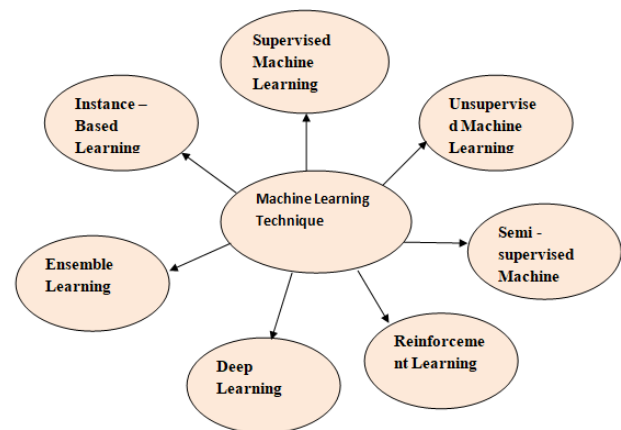


Figure 1. Machine Learning method

4. Reinforcement Learning: Reinforcement learning has not exact input & output data variable. It focuses

on current live interaction of environment. This algorithm is aware when the answer is false, but don't indicate that how to truth it. It has to discover and experiment several cause until it find out the truth answer. Reinforcement learning based on two things first is trial and second is find out error until not get result is maximum positive [4] [5].

5. Deep Learning: Deep Learning is an extended version of neural network which has contains more algorithms. This learning algorithms model gives high – level outline idea to data. These models create deep graph with several processing layer, which make up of various linear and nonlinear variation [22] [23] [24].

6. Ensemble Learning: This type of learning combining in several single learners into one forms then that specific type of learning is known Ensemble Learning. This single learner may be support vector machine, K-NN, C4.5, etc. We can identify that group of learner do better performance in specific work compare to single learner [31] [32]. Ex. Boosting, Bagging

7. Instance-Based Learning: these types of learning learner learn a specific type of model & try to apply new data into this model. If size of the data is upgrade than complexity increase of this learning algorithm. Ex. K – Nearest Neighbor

III. Machine Learning Methods for Diagnosis of Medical Diseases

Many researchers have study on various machine learning methods for Diagnosis of medical disease data. Machine Learning methods of Data Mining have been worked easily in diagnosis of various medical diseases. Here study of different review paper for diagnosis of medical diseases by data mining learning algorithms are Breast Cancer, Brain Tumor, Heart diseases [6] [7] [10].

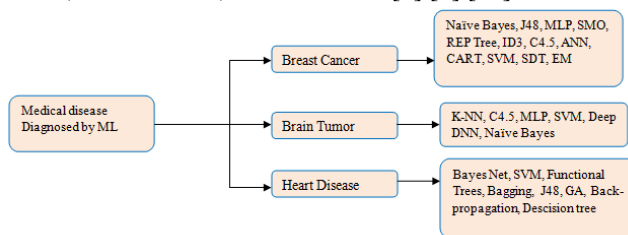


Figure 2. Disease analysis by Machine Learning method

A. Breast Cancer Disease

R. Nithya & B. Santhi [28] in their study they was developed CAD system for the classification of mammogram into normal and cancer pattern. Their proposed algorithms trained with 200 mammograms (100 normal, 100 cancers) and tested 50 mammograms. In this paper CAD (Computer Aided Diagnosis) system developed using GLCM feature and used neural network ML method. This paper was examined on DDSM (Digital Database for Screening Mammography) database was used classification of w. r. to accuracy, sensitivity and specificity. The maximal accuracy

percent for normal and cancer classification was 96% obtained. Vikas Chaurasia and Saurabh Pal [39] in their study they was investigate the breast Cancer data with a 683 instance and 10 attributes take from the Wisconsin dataset from UCI machine learning. They discard the 16 instances with missing values from the dataset. After this build a new dataset with 683 instances. Benign: 458 (65.5%) Malignant: 241 (34.5%) was take for Analysis by three Machine Learning algorithms w. r. to BF Tree, IBK, SMO of data mining classification technique on WEKA software. They obtained that accuracy of ML method of BF Tree, IBK, SMO (Sequential Minimal Optimization) have w. r. to 95.46%, 95.90%, 96.19%. We have easily seen that SMO ML method has high accurate 96.19% rate among of them. G. Sujatha and K. Usha Rani [14] they have used two data set primary tumor and colon tumor. The primary tumor data set retrieved from UCI machine learning Repository and the colon tumor data set retrieved from Bioinformatics Group Seville. Both Data set analyzed by WEKA tool. They have used Data Mining classification Machine learning algorithms ID3, C4.5, CART. After performed Machine learning algorithms in ID3, C4.5, CART they have obtained accuracy of primary tumor w. r. to 34.22, 40.12, 39.82. As well as after performed Machine learning algorithms in ID3, C4.5, CART they have obtained accuracy of colon tumor w. r. to 59.68, 82.26, 75.81. We have optimized that high accuracy of primary tumor data set is 40.12 w.r.to ID3 ML methods and high accuracy of colon tumor data set is 82.26 w.r.to C4.5 ML methods [13]. Keerti Yeulkar, Dr. Rahila Sheikh [20] in their research they have investigated data mining classification of two machine learning algorithms was C4.5 and Naive Bayes. Their research work attempts to analyze the data mining classification techniques for the SEER (Surveillance, Epidemiology and End Results). They have retrieved 700 samples Record from SEER dataset. The SEER breast cancer Dataset classify into “malignant” and “benign” cases and analyzed by R programming language. They have obtained Compare various Data mining method for analysis of Breast Cancer accuracy of the C4.5, Naive Bayes, Naive Bayes (Laplace) and Naive Bayes (Metric Prediction) ML algorithm w.r.to 98.09, 95.85, 95.85, 50.5. We have analyzed that C4.5 ML algorithm is high accuracy 98.09% among of all given ML algorithms. Dursun Delen, Glenn Walker, Amit Kadam [11] their study they have take the SEER Breast cancer dataset which contained of 433,272 records/cases and 72 variables and used this dataset on the SPSS statistical analysis tool for abstained result. They have analyzed dataset of data mining classification technique of machine learning methods. Their research they take three ML algorithms ANN, decision trees and statistics (logistic regression) and get accuracy w. r. to 91.2%, 93.6%, 89.2%. After performed their research of three methods they obtained that the decision tree method (C5) comes the high accuracy 93.6% among of ANN, logistic regression. Sudhir D. Sawarkar, Ashok A. Ghatol, Amol P. Pande [33] in their research they have retrieved data set from UCI WBCD. The dataset have 699 instance 10 attributes. Data set applied on

SVM and ANN on the WBC data. They have achieved accuracy of ANN and SVM (ANN) w.r.to 85%, 97%. Uma Ojha & Savita Goel [37] their research paper they have used WPBC dataset from UCI machine learning. Their dataset sample contained 194 record and 35 attributes. data mining

Author	Year	Disease	Resource of data set	Tool	Data mining technique	Machine learning methods	Accuracy
Keerti Yeulkar, Rahila Sheikh	2017	Breast Cancer	SEER	R programming language	Classification	C4.5 Naive Bayes NB (Laplace) NB(Metric Prediction)	98.0% 95.8% 95.8% 50.5%
Uma Ojha & Savita Goel	2017	Breast Cancer	UCI	R programming language tool	Classification & Clustering	C5.0 SVM EM	81% 81% 68%
Vikas Chaurasia and Saurabh Pal	2014	Breast Cancer	UCI	WEKA	Classification	BF Tree IBK SMO	95.4% 95.9% 96.1%
G. Sujatha and K. Usha Rani	2013	Primary tumor	UCI	WEKA	Classification	ID3 C4.5 CART	34.2% 40.1% 39.8%
R. Nithya & B. Santhi	2011	Breast Cancer	DDSM	CAD System	Classification	Neural network	96%
Sudhir D. et al.	2006	Breast Cancer	UCI	WEKA	Classification	ANN SVM	85% 97%
Dursun Delen et al.	2004	Breast Cancer	SEER	SPSS	Classification	ANN Decision trees Statistics (logistic regression)	91.2% 93.6% 89.2%

Table 1. Compare various Data mining method for analysis of Breast Cancer

methods They have used clustering & classification data mining techniques of eight popular ML methods four from clustering (K-means, EM, PAM and Fuzzy c-means) and four from classification (SVM, C5.0, KNN and Naive Bayes). They have performed analysis of dataset on R programming language tool. After analysis tasks they obtained accuracy of ML classification algorithm C5.0 and SVM w. r. to 81% and clustering of ML algorithm EM was obtained 68% accuracy among of all ML algorithms.

Analysis: In study of literature paper, SVM give maximum accuracy of 97% in table 1. In various medical diseases, SVM play well for performance of result. Uma Ojha & Savita Goel in 2017 are correctly show of Classification & Clustering data mining techniques which classification of ML SVM give accurate 81% but it is comparatively minimum as in 2006. For preprocessing data are used different tools.

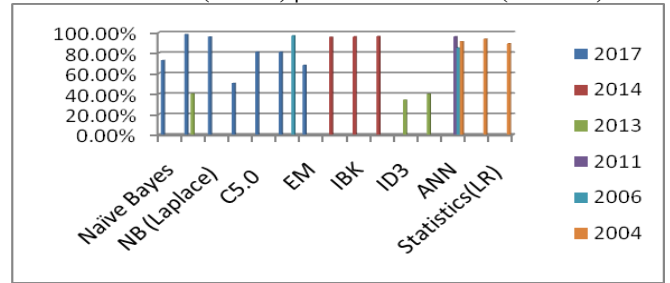


Figure 3. Detect Breast Cancer accuracy by Machine Learning method

Merits and Demerits of SVM:

Merits: Build accurate classification and avoiding over fitting, powerful to noise.

Demerits: SVM has binary classifier. Its behavior is pair wise classification of multiple class. Computational price is expensive, so it works slowly.

B. Heart Disease

Otoom et al [26] in their study they have investigated coronary artery heart disease. They have taken heart dataset from UCI. Their data set included of 303 instances & 76 rows. Their proposed system used Bayes Net, Support vector machine and Functional Trees ML methods on WEKA tool for explore the disease. After examination they obtained accuracy of Bayes Net, SVM and Functional Trees w. r. to 84.5%, 85.1%, 84.5%. Vikas Chaurasia & Saurabh Pal [40] they have study of find out heart disease. In their research Naive Bayes, bagging and J48 machine learning algorithms was applied on WEKA tool for explorer heart disease. They have taken UCI data set which contained 76 cases. Their accuracy rate of machine learning algorithms Naive Bayes, bagging and J48 are w.r.to 82.31%, 85.03%, 84.35%. we observed that bagging machine learning algorithms is maximum 85.03% accuracy among of all. Vembandasamy et al [38] in their research paper they have obtained data-set

Table 2. Compare various Data mining method for analysis of Heart Disease

which contained of 500 patients record from research organization in Chennai. For data analysis they have used Naïve Bayes ML data mining classification technique on WEKA tool. Naïve Bayes Machine learning algorithm achieves 86.42% accuracy performed by WEKA tool. Parthiban and Srivatsa [12] In their study they provide diagnosis of heart problem for diabetic patients used by Naïve Bayes & SVM machine learning algorithm. They have taken 500 data set from organization of research in chennai. This data set applied on WEKA tool. Naïve Bayes method 74% and SVM method 94.61% of accuracy was achieved. We have observed that SVM method gave highest correct accurate result compare than Naïve Bayes method. Tan K.C et [36] al. in their study they have retrieved five different data set also included by heart disease from UCI for analysis purpose. They have used two ML method which was Genetic method and SVM method. Their data set was applied on WEKA tool and LIBSVM tool and after analysis they have achieved 84.10% accuracy for heart disease. We also know that G.A. and SVM are hybrid technique of machine learning algorithm. Ashish Kumar Sen et al [3] in their study they have described a two level views for determine the heart disease probability. They have retrieved coronary heart disease data set from UCI ML data set with 303 cases & 9 attributes repository to test the back-propagation and neuro-fuzzy machine learning algorithm was applied to cure the coronary heart disease problem. They have used MATLAB tool for analysis of data set. The data set was split into two parts first part of 80% data set was selected for training and second part of 20% data set was selected for testing. This data set was employed on MATLAB tool and after analysis they have achieved work percent was very high and error percent was very less. Syed Umar Amin et al. [35] in this paper they have investigated prediction of large risk factor for heart disease. They have used data set from American heart institute of 50 people with risk factor 12. Analysis of this data set through neural network and genetic algorithm method on MATLAB tool and obtained accuracy of 89% for predicted risk of heart problem. Jyoti Soni et al. [17] in their study they have used dataset which contained 909 instance and 15 factors from Cleveland Heart database. In their research work they have applied data set on Tanagra tool for evaluated result. After evaluated on the data set they had achieved accuracy of Naïve Bayes and Decision tree machine learning algorithm w. r. to 96.5%, 99.2%.

Analysis: Founded on Naïve Bayes system is good for Heart Disease. Naïve Bayes give highest accuracy of 96.5% in 2011. The outcome display that system may predict analysis with minimum error and also this system is useful for diagnosis of heart disease. But in 2012, accuracy obtained by Naïve Bayes is very less. It give 74% accuracy. In 2013 & 2015 it presents optimum 82.32% & 86.42% accuracy. Their

Author	Year	Disease	Resource of data set	Tool	Data mining technique	Machine learning methods	Accuracy
Otoom et al.	2015	(CAD) Heart Disease	UCI ML repository	WEKA	classification	Bayes Net SVM Functional Trees	84.5% 85.1% 84.5%
Vembandamy et al.	2015	Heart Disease	research organization in Chennai	WEKA	classification	Naïve Bayes	86.4%
Vikas Chaurasia & Saurabh Pal	2013	Heart Disease	UCI ML laboratory	WEKA	classification	Naïve Bayes bagging J48	82.3% 85.0% 84.3%
Ashish Kumar Sen et al	2013	(CAD) Heart Disease	UCI ML repository	MATLAB	classification	back-propagation	90%
Syed Umar Amin et al.	2013	Heart Disease (risk factor)	American heart institute	MATLAB	classification	GA + neural network	89%
Parthiban and Srivatsa	2012	Heart Disease	research organization in Chennai	WEKA	classification	Naïve Bayes SVM	74% 94.6%
Jyoti Soni et al.	2011	Heart Disease	Cleveland Heart DB	Tanagra	Classification & clustering	Naïve Bayes Decision tree	96.5% 99.2%
Tan et al.	2009	Heart Disease	research organization in Chennai	WEKA and LIBSVM	classification	GA + SVM	84.1%

large data for training & testing.

Merits and Demerits of Naïve Bayes:

Merits: The Naïve Bayes algorithm is increased execution of classification. It also remove the un-matching symptom. Its execution is good. It takes minimum computation time.

Demerits: Naïve Bayes algorithm require maximum item of data to obtain good result. It is slow to store the whole testing & training examples.

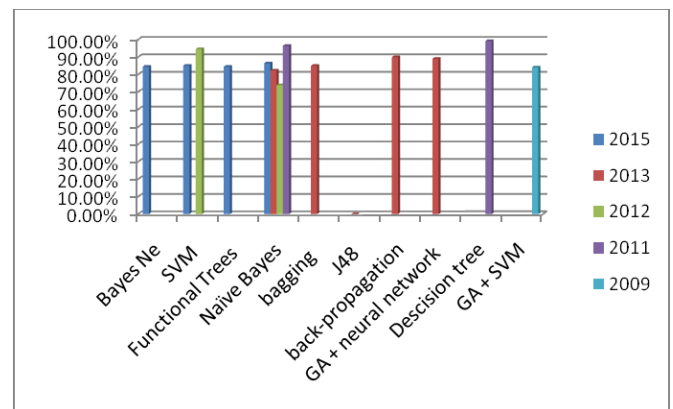


Figure 4. Detect Heart Disease accuracy by Machine Learning methods

Author	Year	Disease	Resource of dataset	Tool	Data mining technique	Machine learning methods	Accuracy
Heba Mohsen et al	2018	Brain tumor	Harvard Medical School	MATLAB & WEKA	classification	Deep DNN	96.97%
B. Balakumar et al.	2017	Brain tumor	Brainix Dataset	MATLAB	classification	SVM	89.5%
Sunita Singh	2015	Brain tumor	Brain MRI	MATLAB	classification	K-NN	84%
Dena Nadir et al.	2015	Brain tumor	Brain MRI	MATLAB & WEKA	classification	C4.5 MLP	91% 95%
N. Subash And J. Rajeesh	2015	Brain tumor	Brain MRI	WEKA	classification	SVM	95%
Komal Sharma et al.	2014	Brain tumor	Brain MRI	MATLAB & WEKA	classification	MLP & Naive Bayes	98.6% 97.6%

Learning method

C. Brain Tumor

Sunita Singh [34] in their research they have used GLCM method for retrieve textural form to the image. In their research paper, they have developed a method which automatic find out brain tumors with more speed of MRI image. They have used Jolliffe's B4 method for choose textural form. In their research paper they proposed K-NN machine learning algorithm of classification Technique. Their research paper they have used MRI images, MRI may display variation of abnormal brain. They have achieved 84% accurate result by classification of K-NN method. Dena Nadir et al. [8] in their research paper investigation of find out abnormal brain by MRI image. This MRI has complex and variant tasks. The proposed method was used two predictive algorithm first is C4.5 decision tree and second is Multi Layer Perceptron (MLP) method. MATLAB and WEKA tool was used for analysis of their result. In their paper they have acquired 174 specimen of brain MRI and this sample classify by on C4.5 decision tree & MLP method. After computation they have achieved precision of brain C4.5 was 91% and MLP was 95%. B. Balakumar et al. [9] Their proposed model well identify system for classify of brain tumor. Their work has two steps was pre-processed the Brain MRI & second was divided for retrieved ROI. They have collected 60 sample included normal and abnormal brain tumor from Brainix Dataset. SVM ML method used for classification of the data and after computation obtained 89.5% accurate result. MATLAB tool was used for image processing and analysis the proposed system. N. Subash and J. Rajeesh [25] in their research work used SVM algorithm based on classification technique for detect brain tumor from MRI images. The SVM acquired more accuracy and less error rate with high 95% of well classification. Heba Mohsen et al. [16] in their investigation they have used classification of MRI for find out brain tumor. They have used Deep DNN ML algorithm to separate the normal and abnormal brain

tumor. The dataset contained 66 sample of brain MRI with 22 non cancerous & 44 malignant retrieved from Harvard Medical School. The data set applied on MATLAB for preprocessing of the data & WEKA tool was used for analysis of the ML method. The Deep DNN ML algorithm was presented 96.97% accurate result. Komal Sharma et al. [21] their proposed model they have used 212 dataset sample for analysis of MLP & Naive Bayes on WEKA Tool. After computation they have obtained more 98.6% accuracy of MLP compare than Naive Bayes for detect malignant brain tumor. MATLAB tool was used feature retrieved.

Table 3. Compare various Data mining method for analysis of Heart Disease

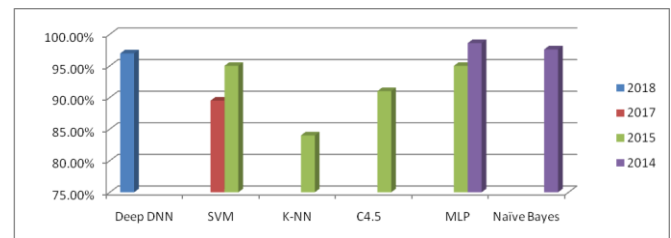


Figure 5. Machine learning methods for analysis of Brain tumor diseases.

Analysis: The MLP methods are very strong form of ANN that is used for classification. The multilayer Perceptron is also call feed forward neural network. Komal Sharma et al are giving 98.6% maximum accurate result in 2014 by MLP. Dena Nadir et al. are give optimum accurate 95% result in 2015. Use of MLP achieves high accuracy of classification of brain tumor into good, possible to apply, effective of rate.

Merits and Demerits of Multi Layer Perceptron:

Merits: Multilayer Perceptron contain more application for resolve complicated task. MLP consist loop and outcome it only dependent only input item data this reason that is call feed forward neural network. MLP are useful for mapping with non-linear wise. It may be N- no. of input & M- no. of output. This algorithm is very good for classification.

Demerits: The main disadvantage of these algorithms is that it is trained so it can not secure of minima. Other demerits of MLP algorithm is that user must be set internal layer neurons, if after setting value is minimum may outcome of MLP show under fitting and if value is high may outcome of MLP show over fitting.

IV. Observation of Machine Learning Algorithms: Various machine learning algorithms perform too nice for Diagnosis & Prediction on various medical diseases like breast cancer, brain tumor, heart disease. We have to observe that these existing algorithm SVM & Naive Bayes algorithms are much more used for Diagnosis &

Prediction of disease. Both algorithms provide well accuracy compare than other algorithm. Decision tree algorithm has more complexity but it provides accurate result. Artificial neural network show the much outcome but it takes maximum time compare to another algorithms.

V. Conclusion: These papers give different machine learning algorithms and observe on diagnosis of various diseases such as Breast Cancer, Brain tumor, Heart Disease. It is identify that above survey provide well diagnosis and prediction analysis of applied on various machine learning algorithms which can help to medical practitioners for take decision in early treatment of disease. From before study, analysis of Breast cancer, SVM give maximum accuracy of 97%. Brain tumor, give maximum accuracy of 98.6% by MLP. Heart disease, Naïve Bayes provide highest accuracy of 96.5%. Machine learning algorithms proposed for different possible solution of disease. This paper provides useful knowledge for analysis of improved medical diseases. Also determine of improvement graph in detail for analysis of different machine learning algorithm. So machine learning technique offer for improved accuracy of decision.

References: -

- [1]. A.Suganya, K.K.Kavitha, “*Diagnosing Benign and Malignant Breast Cancer Using Data Mining classification*”, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 12, Pg. no. 21006-21010, 2016.
- [2]. Aarti Sharma, Rahul Sharma, Vivek Kr. Sharma, Vishal Shrivastava, “*Application of Data Mining – Computer Science and Information Technologies*”, Vol. 5, Issue 2, Pg. no. 2023-2025. 2014.
- [3]. Ashish Kumar Sen, Shamsher Bahadur Patel, Dr. D. P. Shukla, “*A Data Mining Technique for Prediction of Coronary Heart Disease Using Neuro-Fuzzy Integrated Approach Two Level*”. International Journal Of Engineering And Computer Science, Vol. 2 Issue 9, Page No. 2663-2671, 2013.
- [4]. Ayon Dey, “*Machine Learning Algorithms: A Review*”. International Journal of Computer Science and Information Technologies, Vol. 7, Issue 3, Pg. no. 1174-1179, 2016.
- [5]. B. Gayathri, Sumathi M., T. Santhanam, “*Breast Cancer Diagnosis using Machine Learning Algorithms – A Survey*”. International Journal of Distributed and Parallel Systems, Vol. 4, Issue 3, Pg. no.105-112, 2013.
- [6]. Deepashri, K., S. & Ashwini Kamath, “*Survey on Techniques of Data Mining and its Applications*”. International Journal of Emerging Research in Management & Technology, Vol. 6, Issue 2, Pg. no.198-201, 2017.
- [7]. Deepika Verma & Nidhi Mishra, “*Analysis and Prediction of Breast cancer and Diabetes disease datasets using Data mining classification Techniques*”. IEEE International Conference Intelligent Sustainable Systems, Pg. no.188-195. 2017.
- [8]. Dena Nadir George, Hashem B. Jehlol, Anwer Subhi Abdulhussein Oleiwi, “*Brain Tumor Detection Using Shape features and Machine Learning Algorithms*”. International Journal of Scientific & Engineering Research, Vol. 6, Issue 12, Pg. no. 454-459, 2015.
- [9]. B.Balakumar, P.Raviraj, E.Divya Devi, “*Brain Tumor Classification Using Machine Learning Algorithms*”. Engineering Research & Management, Vol. 4, Issue 2, Pg. no. 30-41, 2017.
- [10]. C Nalini & D.Meera, “*Breast cancer prediction system using Data mining methods*”. International Journal of Pure and Applied Mathematics, Vol. 119, Issue 12, Pg. no.10901-10911, 2018.
- [11]. Dursun Delen, Glenn Walker, Amit Kadam, “*Predicting breast cancer survivability: a comparison of three data mining methods*”, Artificial Intelligence in Medicine Vol. 34, Pg.no.113—127, 2005.
- [12]. G. Parthiban & S.K.Srivatsa, “*Applying Machine Learning Methods in Diagnosing Heart Disease for Diabetic Patients*”. International Journal of Applied Information Systems (IJ AIS), Vol. 3, Issue 7, Page no. 25-30, 2012.
- [13]. G. Sujatha and K. Usha Rani, “*A Survey on Effectiveness of Data Mining Techniques on Cancer Data Sets*”. International Journal of Engineering Science Research, Vol. 4, Issue 1, Pg. no. 1298-1304, 2013.
- [14]. G. Sujatha and K. Usha Rani, “*Evaluation of Decision Tree Classifiers on Tumor Datasets*”. International Journal of Emerging Trends & Technology in Computer Science, Vol. 2, Issue 4, Pg. no. 418-423. 2013.
- [15]. Hamid Karim Khani Zand, “*A Comparative study on Data Mining Technique for Breast Cancer Diagnosis*

- and Prediction". Indian Journal of Fundamental and Applied Life Sciences, Vol. 5, Pg. no. 4330-4339, 2015.
- [16]. Heba Mohsen, El-Sayed A. El-Dahshan, El-Sayed M. El-Horbaty d, Abdel-Badeeh M. Salem, "Classification using deep learning neural networks for brain tumors". Future Computing and Informatics Journal, Vol. 3, Pg. no. 68-71, 2018.
- [17]. Jyoti Soni, Ujma Ansari and Dipesh Sharma, "Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction". International Journal of Computer Applications, Vol. 17, Issue 8, Pg. no. 43-48, 2011.
- [18]. K. Amarendra, "A survey on Data Mining Technique its Application". International Journal of Emerging Trends & Technology in Computer Science, Vol. 3, Issue 3, Pg. no. 163-167, 2014.
- [19]. Kajaree Das and Rabi Narayan Behera, "A Survey on Machine Learning: Concept, Algorithms and Applications". International Journal of Innovative Research in Computer and Communication Engineering, Vol. 5, Issue 2, Pg. no. 1303-1309, 2017.
- [20]. Keerti Yeulkar and Dr. Rahila Sheikh, "R Analysis of SEER Breast Cancer Dataset Using Naïve Bayes and C4.5 Algorithm", International Journal of Computer Science And Technology, Vol. 8, Issue 4, Pg.no.43-45, 2017.
- [21]. Komal Sharma, Akwinder Kaur, Shruti Gujra, "Brain Tumor Detection based on Machine Learning Algorithms". International Journal of Computer Applications, Vol. 103, Issue 1, Pg. no. 7 – 11, 2014.
- [22]. M. Durairaj and V. Ranjani, "Data Mining Applications in Healthcare Sector: A Study". International Journal of Computer Application, Vol. 2, Issue 10, Pg. no. 30-35, 2013.
- [23]. Meherwar Fatima & Maruf Pasha, "Survey of Machine Learning Algorithms for Disease Diagnostic". Journal of Intelligent Learning Systems and Applications, Vol. 9, Pg. no. 1-16, 2017.
- [24]. Milan Joshi & Anurag Joshi, "Technology On Comparative Study of Breast Cancer Classification". International Journal of Computer Science And Using Ensembles in Statistical Modeling, Vol. 8, Issue 1, Pg. no. 18 -21, 2017.
- [25]. N.Subash and J. Rajeeesh, "Brain Tumor Classification Using Machine Learning". IJCTA, Vol. 8, Issue 5, Pg. no. 2335-234, 2015.
- [26]. Otoom, A.F., Abdallah, E.E., Kilani, Y., Kefaye, A. and Ashour, M. Parthiban, G. Srivatsa, S.K., "Effective Diagnosis and Monitoring of Heart Disease". International Journal of Applied Information Systems, vol. 9, Issue no. 1, Pg. no.143-156, 2015.
- [27]. QANG Yang & Xindong WU, "10 challengingProblems in Data Mining Research". International journal of Information Technology & Decision Making, Vol. 5, Issue 4, Pg. no.597-603, 2006.
- [28]. R. Nithya & B. Santhi, "Classification of Normal and Abnormal Patterns in Digital Mammograms For Diagnosis of Breast Cancer". International Journal of Computer Applications, Vol. 28, Issue 6, Pg. no. 21-25, 2010.
- [29]. Razaqat Alam Khan et al, "Data Mining Algorithms for Classification of Diagnostic Cancer Using Genetic Optimization Algorithms". International Journal of Computer Science and Network Security, vol.17, Issue 12, Pg. no. 207 – 212, 2017.
- [30]. S. Padma Priya, P. Sowmiya, "Breast cancer prediction using data mining technique", International Journal of Advance Research and Engineering, Vol. 7, Issue 1, Pg.no. 41 - 44, 2018.
- [31]. Samiddha Mukherjee, Ravi Shaw, Nilanjan Haldar, Satyasan Changdar, "A Survey of Data Mining Applications and Techniques". International Journal of Computer Science and Information Technologies, Vol. 6, Issue 5, Pg. no. 4663-4666, 2015.
- [32]. Shelly Gupta, Dharminder Kumar, Anand Sharma, "Classification Technique Applied for Breast Cancer Diagnosis and Prognosis". Indian Journal of Computer Science and Engineering, Vol. 2, Issue 2, Pg. no. 188-195, 2011.
- [33]. Sudhir D. Sawarkar, Ashok A. Ghatol, Amol P. Pande, "Predicting Breast Cancer Recurrence using Data Mining Techniques". International Journal of Computer Applications, Vol. 122, Issue 23, Pg. no. 26 –31, 2015.
- [34]. Sunita singh, "Classification of Human Brain Tumors from MRI Using K-NN Algorithm". International Journal of Advances in Science Engineering and Technology, Vol. 3, Issue 1, Pg. no. 34-37, 2015.
- [35]. Syed Umar Amin, Kavita Agarwal, Dr. Rizwan Beg, "Genetic Neural Network Based Data Mining in Prediction of Heart Disease Using Risk Factors". Proceedings of 2013 IEEE Conference on Information and Communication Technologies ICT, Pg. no. 1227-1231, 2013.
- [36]. Tan K.C., Teoh E.J., Yu Q. and Goh K.C., "A Hybrid Evolutionary Algorithm for Attribute Selection in Data Mining". Journal of Expert System with Applications, Vol. 36, Pg. no. 8616-8630, 2009.
- [37]. Uma Ojha & Savita Goel, "A Study On Prediction of Breast Cancer Recurrence Using Data Mining Techniques". Cloud Computing, Data Science & Engineering, IEEE International Conference, Pg. no. 527 -530, 2017.
- [38]. Vembandasamy, K, Sasipriya, R. and Deepa, E., "Heart Diseases Detection Using Naive Bayes Algorithm". IJISSET-International Journal of Innovative Science, Engineering & Technology, Vol. 2, Pg. no. 441-444, 2015.
- [39]. Vikas Chaurasia & Saurabh Pal, "A Novel Approach for Breast Cancer Detection using Data Mining Techniques". International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, Issue 1, Pg. no. 2456–2465, 2016.

