A Study of Liver DiseasePrediction Using Machine Learning Approaches

BY

MD. ABDUR RAHMAN BHUIYAN

ID: 181-25-645

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science and Engineering

Supervised By

Dr. Md. Ismail Jabiullah
Professor
Department of Computer Science & Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

06 DECEMBER, 2019

DECLARATION

I hereby declare that, this project has been done by me under the supervision of, **Dr. Md. Ismail Jabiullah**, **Department of CSE**, Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised By:

Dr. Md. Ismail Jabiullah
Professor
Department of Computer Science & Engineering
Daffodil International University

Submitted By:

Md. Abdur Rahman Bhuiyan
ID: 181-25-645
Department of CSE
Daffodil International University

APPROVAL

This Project/Thesis titled "A Study of Liver Disease Prediction Using Machine Learning Approaches", submitted by Md. Abdur Rahman Bhuiyan, ID No: 181-25-645 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 06 December, 2019.

BOARD OF EXAMINERS

Dr. Syed Akhter Hossain — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Ismail Jabiullah — **Internal Examiner**
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Sheak Rashed Haider Noori — **Internal Examiner**
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Mohammad Shorif Uddin — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

ACKNOWLEDGEMENT

First,I express my heartiest thanks and gratefulness to almighty Allah for His divine blessing makes me possible to complete the final year thesis successfully.

I sincerely and heartily grateful to my advisor **Dr. Md. Ismail Jabiullah, Professor,** Department of Computer Science & Engineering, Daffodil International University, Dhaka.Deep Knowledge & keen interest of my supervisor in the field of "*Machine Learning*" to carry out this project.His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would also like to wish my deepest heartiest gratitude to **Prof. Dr. Syed Akhter Hossain, Head**, Department of CSE for his lot of deepest help to fulfill my final year thesis and also thanks to other faculty members and the employees of CSE department of Daffodil International University.

Finally, I must acknowledgement with due respect the constant support and patients of my parents.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Objective: The key aspect of this study is to appraise the variousMachine learning (ML) classifiers performance for forecast of liver disease.

Methods: In this work, I have used six supervised MLmethods for the prediction of Liver Disease. For example: “SVM, NB, KNN, RF, DT and LR” were used for early prediction of Liver Disease. Therefore, I evaluated the Liver Disease dataset through $sensitivity_i$, $specificity_i$, f 1 measure and classifiers accuracy.

Results: The prediction performance of Liver Disease analysis shows that LR obtained the uppermost performance with utmost classification accuracy of 75%. Whereas, RF and DT has achieved the next highest accuracy by prediction.

Conclusion: My findings can be used to help reduce the occurrence of the Liver Disease through developing a machine learning based predictive system for early prediction.

CHAPTER ONE

INTRODUCTION

1.1 Background

For perdition of different disease to make the best feasible clinical care decisions it is very essential to take right decision form medical doctors. If the doctor takes their improper decisions, it could be likely to cause interruptions in medical action or drawn to loss of life. We know the medical services is a big commercial viewpoint in every time. The business stream always running in this fields. Patients are always searching there is a good platform for better services. But there is no 100% affordable platform for every patient. Therefore, in this domain need of one excessive platform for problem solves in healthcare and medical fields. Here is my main idea for improving healthcare services is to place more highlighting on early detection of chronic disease and less on treatment and live for better life.

1.2 Motivation of the Research

Liver Disease are the prominent causes of women death and disability in the global perspective.A report shows that the 508000 peoples died in 2011 by chronic disease, specially Liver Disease[1]. In 2015, around 17.7 million people were global death caused by CHD [2]. The World Health Organization (WHO)estimated that above 23.6 million peopleswill be affected by 2030, because of such liver disease[3]. Very few peoples can get their treatment but most of scenario affected by chronic disease treatment is very expensive and complex [4]. Moreover, this reason to takes long time, mistaken or delayed decisions are possible to cause of death. However, the cost of Liver Disease diagnosis and replacement is very extreme and it can be calls as extreme level of financial expenses. A study reported that the Liver Disease cause the commercial benefits with cost over $79 billion, and treating

people with end stage renal disease cost around $35 billion [5]. Liver Disease is chronic in nature and take long time forcured. This causes most of the patients cannot afford the cost of the cure for Liver Disease. Furthermore, chronic disease prediction is most prominent matter for clinical practitioners and medical services center in order to take accurate decision of such disease. Therefore, machine learning based extensive platform can solve these Liver Disease problems through early detection and diagnosis. This works main aspect is to improve early treatment and diagnosis of Liver Disease for peoples of low-income and developing countries. Hence, our study can be a significant approach for the detecting Liver Disease outbreak with machine learning algorithms.

1.3 Problem Statement

In the last 10 years, the growth rate of medical data is going to large amount from enormousarenas [6]. From the art of Machine learning (ML) algorithms have portrayed that purpose to resolve various health and scientific problem [7]. An establishment of several studies show that MLmodels already have obtained dramatically excessive accuracies in disease based medical problems. However, supervised basedmodels are one of the utmost operative method for the academic and health products on clinical fields. [8]. This works main aspect is to improve early treatment and diagnosis of chronic disease for peoples of low-income and developing countries. Hence, our study can be a significant approach for the detecting chronic disease outbreak with machine learning algorithms.

1.4 Research Question

In this work we have a specific research question for Liver Disease detection by prediction using computational techniques which is addressed below:

(1) Does the different type of ML algorithms affect the performance between the various classifiers to prediction of Liver Disease?

1.5 Research Scope

To date, machine learning classification techniques have created a significant impact and obligation in the chronic disease research society for primary discovery of chronic disease. Moreover, ML algorithms are given more accurate results in chronic disease prediction as compared to others data classification techniques [8][9]. Many of studies already shows that the supervised based classification techniques have obtained excellent accuracies in the field of disease prediction [10][4][11]Motivated by this, the authors have used six prominentML techniques for prediction and proper treatment of chronic patients. The main purposeof this work is to inspect the classifiers performance measurement of various prominent supervised methods and gained more efficient outcome by reducing extremely cost of diagnosis and dialysis of chronic diseases. For this study, six supervised learning techniques were used including "KNN, Support Vector Machine, Decision Tree, Random Forest, Naïve Bayes and Logistics Regression". Moreover, the classifiers performance of selected learning techniques is evaluated using the confusion matrix and different statistical methods. Henceforth, the outperform classification technique will donated for the decision support system and diagnosis of chronic disease.

1.6 Thesis Organization

The rest of the work is ordered as following, chapter one describes the objectives of this thesis, motivation behind this thesis, research scope and thesis organization. Chapter 2 depicted the literature review and related works in these clinical areas. And the materials and methodology are described with the evaluation benchmark of different classification techniques in Section 3. Therefore, the performance results and discussion are demonstrated in section four. Finally, the conclusions and future viewpointsof research and recommendations are deliberated in section five.

CHAPTER 2

BACKGROUND

Through related studies were done on applying and using severalML classifiers to determine early detection and prediction of Liver Disease using MLtechniques. However, the outcomes of the previous work on machine learning used in Liver Disease prediction as follows: Jain et al. [12]presented a survey to attribute assortment and machine learning techniques for identification and forecast of chronic disease. This works focused on a comprehensive review of different feature selection methods and their advantage and limitation. The contribution of this study is to use adaptive with parallel classification techniques for chronic complaint prediction.Bartz-Kurycki et al.[12], introduced a new model to forecast "neonatal surgical site infections (SSI)" using diverse classification processes.. The contribution of this study is to examine the ensemble model with more medication relatedattributes. Carvalho et al. [13] presented a new hybrid methods to sustenance the early verdict of Liver Disease. This study tries to find optimum accuracy to provide sustenance to verdict in circumstances whereas Bayesian Network does not provide a satisfactory outcome. The contribution of the study is to advance an automatic device to contribute a precise identification and prediction of Liver Disease. Kumari [14], presented a new prediction system that can predict the occurrence of Liver Disease at early stage by analyzing nominal set of attributes that has been selected from medical datasets. The KNN classifier obtain the best performance (99.28%) than others classifiers. The contribution of this study is to use the proposed system to predict the Liver Disease at early stage with greatly reduces the cost of treatment and improves the quality of life.Tapak et al.[15], introduced a comparative study between Naïve Bayes, Random Forest, AdaBoost, Support Vector Machine, LSSVM, Adabag, Logistics Regression and LDA to predict Liver Disease survival and metastasis. "LR and LDA" were achieved the maximumaccuracy of (86%). The support vector machine and "LDA" have excellentsensitivity in assessmentto other classifiers. The contribution of this study is to use SVM to predict existence of Liver Disease.Asri et al. [16]presented a comparative study between SVM, DT (C4.5), NB, K-NN to predict early stage of Liver Disease. The intelligent

techniques are applied on WEKA data mining tool. Experiment results show that the SVM have the best performance accuracy, it is 97.13%. The contribution of this study to use SVM to predict the early stage of Liver Disease.Chougrad et al. [16]developed a deep convolutional neural networks based computer aided treatment system. The CNN model achieved the best performance, it is 98.94%. And they tested the CNN model on independent database and they've got the accuracy 98.23% and 0.99 AUC. The contribution of this study to use the high performer classifiers within the proposed structure and that can be used to forecast the patients are "benign or malignant".Wang et al. [17], presented a new model to use Liver Disease diagnosis based on patient's historical data from clinical data. The proposed WAUCE model reduces the variance by around 97.98% and increase accuracy by 33.34%. The contribution of this study can be further applied to safer, more reliable illness diagnosis process.Madhuri & Bharat et. Al [18], present a comparative study to diagnosis the Liver Disease patients through supervise machine learning techniques. They applied multiple machine learning algorithms including LR, RF, DT and Multi-layer perception. Multi-layer perception gives high performance compare to others algorithms.Layla &Hana [19],implement a feed forward back propagation network (FFBPN) to classify the "benign Liver Disease or malign Liver Disease". They showed that for an Artificial Neural network best design for classification is that three hidden layer and twenty-one neurons in every hidden level. Propose design gives highest accuracy 98%.Amrane et al. [20] presented a comparison between two machine learning classifier NB and KNN to provide an effectively diagnosis Liver Disease patients. The comparison result was KNN gives high accuracy at 97.51% and NB has 96.19% accuracy.Al-hadidi et al.[21], presented a model to detect the liver disease with a higher accurateness. Their model was divided into two part, first one was processing the images for extract the features and second one was used two supervise machine learning techniques to get the accuracy. Xiao et al. [22], introduce a new strategy for gene expression analysis to five different classification algorithms with deep learning methods. The contribution of this study is shown to be accurate and effective prediction results for Liver Disease prediction.

CHAPTER THREE

MATERIALS AND METHODS

3.1 Experimental Setup

In this work, this section represents the working process (figure 3.1.1) of the experiment including machine learning techniques. Liver Disease data sets have been considered in this work. Firstly, we focused on preparing and combined data from the main datasets. Moreover, I extracted features from the liver datasets. Then, we have checked the missing values and co related values. Secondly, Data set splitting is an important task of this machine learning based fields. In this dataset, I have not found split and test datasets. Figure 3.1.1 shows the Parkinson data set has split into trainset and test sets. After that, 6 supervised based classifiers performed the operation. After successfully executed these algorithms LR obtained the highest performance.

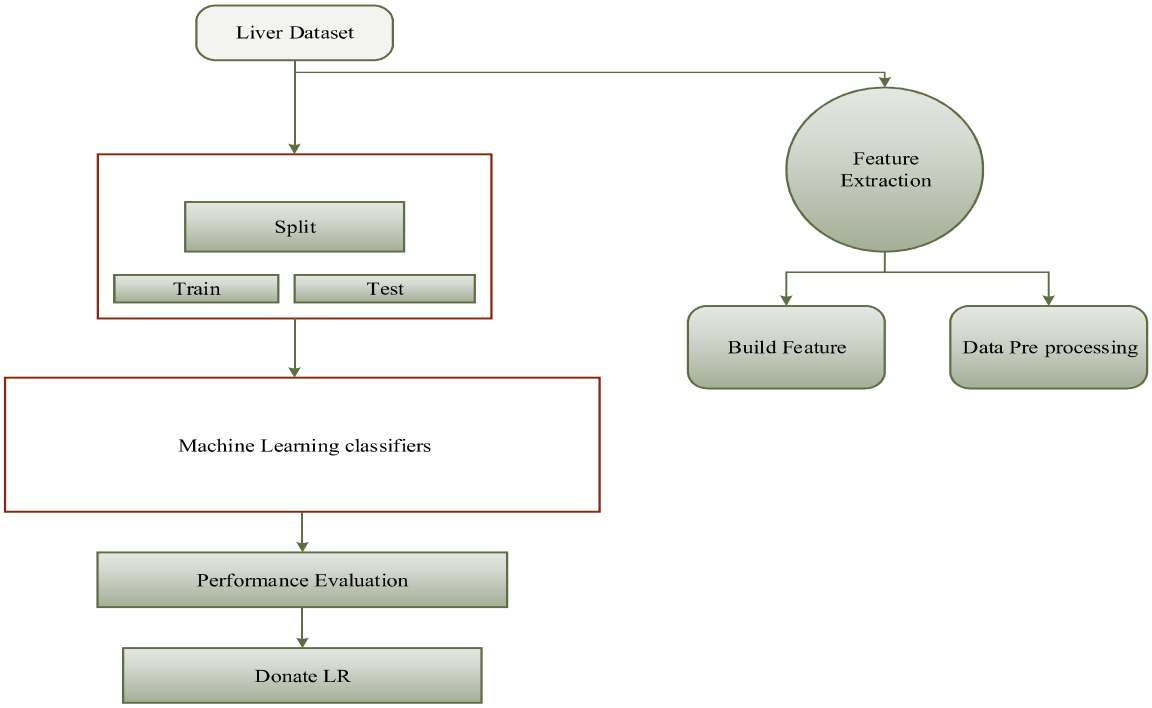


Figure 3.1.1. The experimental setup

3.2 Data Collection

In this study, we use the liver patient data from provided by the "University of California, Irvine"[23]. The objective of the dataset is to diagnostically predict whether or not a patient has liver disease based on certain diagnostic parameters. Moreover, in the dataset contains 583 liver patient's data whereas 416 samples are liver patients and 167 samples are non-liver patient. This data set contains 75.64% samples are male patients and 24.36% samples are female patient. However, I chose the particular parameters for data analysis which are depicted in table 3.2. 1

Table 3.2.1 Paraments for data analysis

No		Attributes		Indication	Description
1		Age		Numerical	Years
2		Gender		Male or Female	Sex of the patients
3		Total_Bilirubin		Numerical	mg/dL
4		Direct_Bilirubin		Numerical	mg/dL
5		Alkaline_Phosphotase		Numerical	ALP in IU/L
6		Alamine_Aminotransferase		Numerical	ALT in IU/L
7		Aspartate_Aminotransferase		Numerical	AST in IU/L
8		Total_Protiens		Numerical	g/dL
9		Albumin		Numerical	g/dL
10		Albumin_and_Globulin_Ratio		Numerical	A/G ratio
11		Dataset		Yes or No	patient has liver disease or not

3.3. Data Pre-Processing

Our dataset contains 416 samples diagnosed with liver disease and 167 samples not diagnosed with liver disease. Figure 3.3.1 shown that the ratio of the liver patients. In this dataset, there is most of the patient are male, Fig 3.3.2shown that the ratio of gander the liver patients (the number of male patients is 441 and 142 are female).

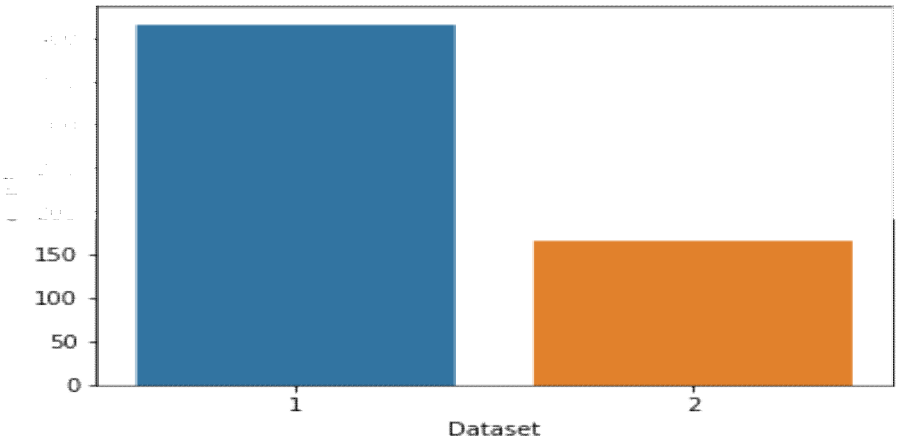


Figure 3.3.1 Count plot shows the ratio of the liver patients

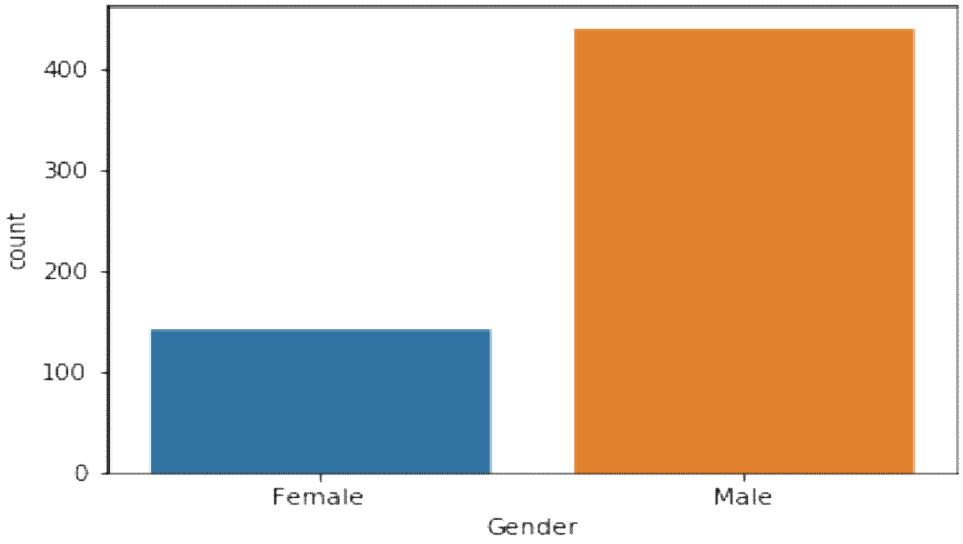


Figure 3.3.2 Count plot shows the ratio of gander the liver patients

This dataset has ten features column and output column. In the output column, value (1)specifiesthat the samplehas negative condition and '0' indicates the samplehas positive condition.Therefore, we examine between column to find the correlated column(figure 3.3.3), the we found some correlated features from heat map.

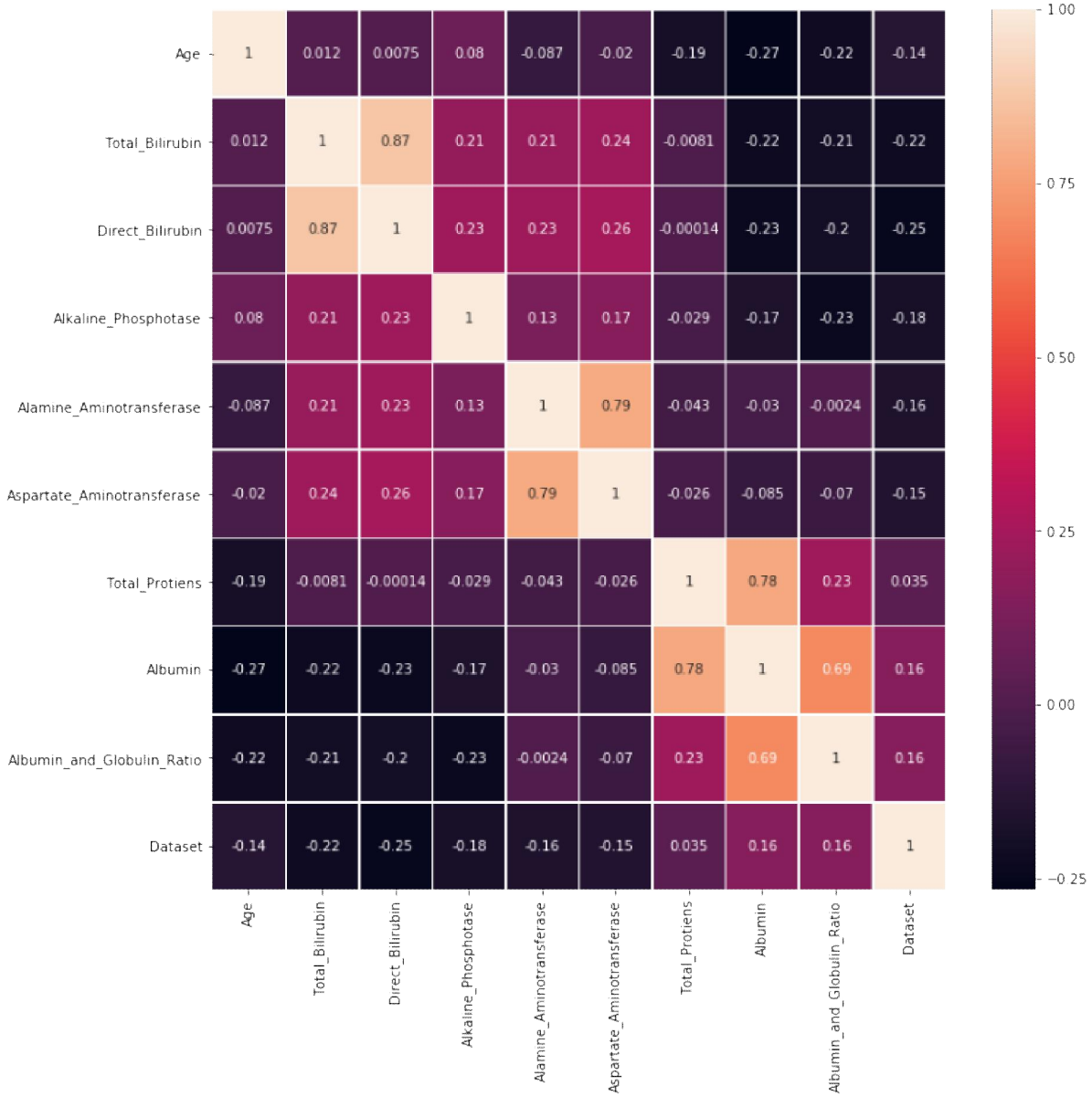


Figure 3.3.3. Heat map for checking correlated columns for Liver Diseases

3.4 Evaluation Criteria

In this thesis, we have depleted6MLmethodsfor the forecastof Liver Disease. Therefore, the performance measurements of the ML methodsare appraised by diversestatistical measurement techniques. For example, we have chosen the confusion_matrix(True_Positive, False_Positive, True_Negative, False_Negative), $Recall_i$, $Precision_i$, f1- measure etc.[24].

The calculationapproachof the measurement techniques are following,

$$Accuracy_i = (TP_i + TN_i) / (TP_i + FP_i + TN_i + FN_i) \quad (1)$$

$$TPR_i \text{ or } Sensitivity_i \text{ or } Recall_i = TP_i / (TP_i + FN_i) \quad (2)$$

$$Specificity_i = TN_i / (TN_i + FP_i) \quad (3)$$

$$Precision_i = TP_i / (TP_i + FP_i) \quad (4)$$

$$f1_i = 2 * (Recall_i * Precision_i) / (Recall_i + Precision_i) \quad (5)$$

$$\text{False Positive Rate} = 1 - Specificity_i \quad (6)$$

The f1_measure is denoted by the weighted norm of the $recall_i$and$precision_i$. To classify as a better classifier this the value will be 1 and for the lowest performance, it will be 0.

3.5 Software and Tools

This work was employedin prominent programming language, i.e. python version 3.7.1 andusing Anaconda Distribution including Jupyter Notebook. The version of the notebook server is: 5.6.0-3badce9.

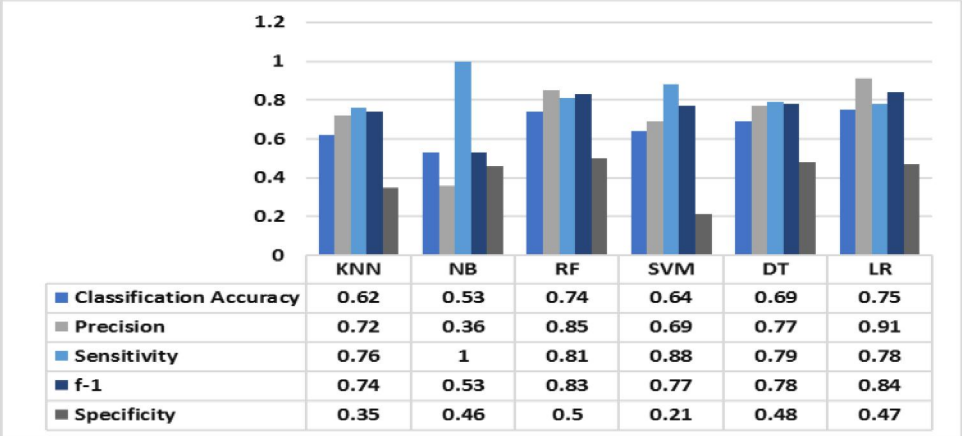
CHAPTER 4

RESULT AND DISCUSSION

4.1 Analysis of the Results

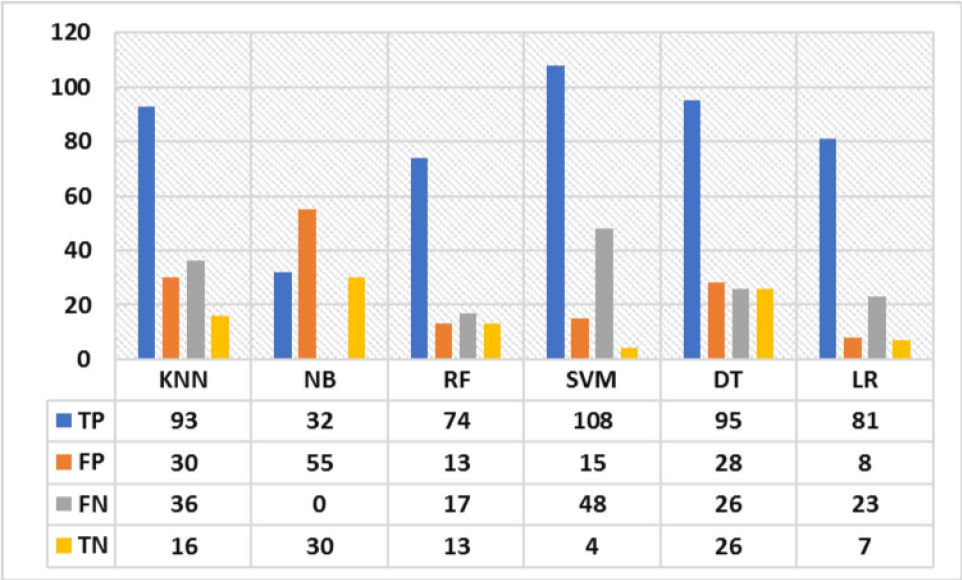
In this thesis, I have accompanied severalanalyses to examine the six ML based supervised techniques for diagnosis and forecast of Liver Disease. The performance comparison and performance measure of six machine learning classifiers for chronic disease prediction as detailed in the following.

In thisresearch, we have takensomeinvestigationto inspectthe performance of six ML methods for the detection of liver disease. Figure 4.1.1indicates thatthe performance analysis of six supervised based approachesfor liver disease data. With respect to precision LR achieved the highest score, 91% and NB perform worst (i.e. 0.36%). However, when considering the sensitivity, SVM attainedthe uppermostvalue (i.e. 0.88%) and KNN obtained the poorestvalue (i.e. 0.76%). Logistics Regression was also the best performer in terms of fi measure ("i.e. 0.83%"), and NB gainedthe poor performance (i.e. 0.53%). In the terms of accuracy, LR attainedthe maximumaccuracy (i.e. 0.75%), and NB accomplishedthe bad performance (i.e. 0.53%).The LR suggeststhat this ML method is more operationalthan the other algorithms for predicting Liver Disease. Moreover, the confusion matrix of this analysis is presented in figure 4.1.2.



	KNN	NB	RF	SVM	DT	LR
Classification Accuracy	0.62	0.53	0.74	0.64	0.69	0.75
Precision	0.72	0.36	0.85	0.69	0.77	0.91
Sensitivity	0.76	1	0.81	0.88	0.79	0.78
f-1	0.74	0.53	0.83	0.77	0.78	0.84
Specificity	0.35	0.46	0.5	0.21	0.48	0.47

Figure 4.1.1 The performance comparison of six supervised learning techniques



	KNN	NB	RF	SVM	DT	LR
TP	93	32	74	108	95	81
FP	30	55	13	15	28	8
FN	36	0	17	48	26	23
TN	16	30	13	4	26	7

Figure 4.1.2 The confusion matrix of prediction results

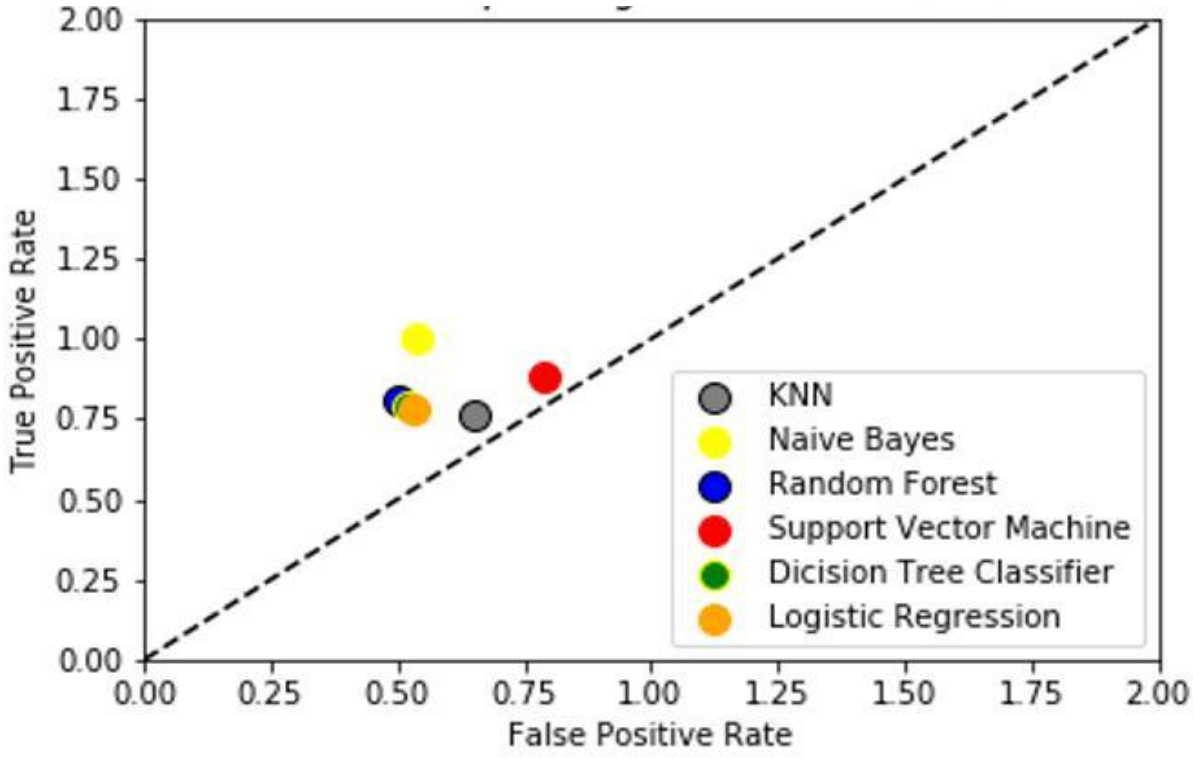


Figure 4.1.3 Receiver Operating Characteristics curve for Liver datasets

The prediction result shows the classifiers outcome above 70% for Liver Disease detection. Moreover, the six machine learning algorithms are doing very good for Liver Disease prediction. In our experiment, it is very essential to recognizethe "receiver operating

characteristics (ROC) curve", which is justifiedon "true positive rate (TPR) and false positive rate (FPR)" of these predictionoutcomes. According to 'ROC curve' (figure 4.1.3.)achieved highest AUC (area under curve) for ROC.

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Findings and Contribution

In this study, I have depicted several ML based classification techniques. Therefore, I deliver anexperimental process on ML based system for the early prediction of Liver Disease. Therefore, Icompeted the presentation of the six algorithms which are depleted in the forecast of Liver Disease diseases and assessed by their results using a statistical technique namely is confusion matrix. The experimental performance shows that the NB has achieved the outperform than the other classifiers within Liver Diseasedatasets. Thisinspection hasusage six MLmethodsfor the prediction of liver diseasewithsome clinical attributes. In addition, this study is part of a project that has the purpose to create an advanced tool to give more precise treatment to normal events and make a superior decision to complex situations. The application will be able to early detect in Liver Disease in a few minutes and notify the real

condition with extreme likelihood of having disease. This application can be remarkably beneficial in low-income countries where is lack of medical institutions and as well as specialized practitioners.

5.2 Recommendation for Future Work

In my experiments, related to most work in the study, each classification algorithms were trained and evaluated on a training dataset that comprises both of positive values and negative values. Moreover, the work can be helpful for chronic disease diagnosis and detection by collecting data from different devices and health related sensors, clinical and medical center and can deliver more accurate results for disease prediction and diagnosis. In my research perspective, there are several directions for the future work in this area of research. I only investigated to some popular supervised machine learning algorithms, it can be choosing more algorithm for build the accurate model of these chronic disease prediction and performance can be more improved. Additionally, I have underlinedthe studytrend and possibility in comparativeto liver disease research and medical data analysisthroughML based techniques.

REFERENCS

[1] "WHO | Breast cancer: prevention and control," *WHO*, 2016.

[2] "WHO | World Health Organization." [Online]. Available: http://www.who.int/. [Accessed: 05-Nov-2018].

[3] Purushottam, K. Saxena, and R. Sharma, "Efficient Heart Disease Prediction System," *Procedia Comput. Sci.*, vol. 85, pp. 962–969, Jan. 2016.

[4] P. Singh, S. Singh, and G. S. Pandi-Jain, "Effective heart disease prediction system using data mining techniques.," *Int. J. Nanomedicine*, vol. 13, pp. 121–124, 2018.

[5] "Chronic Kidney Disease Basics | Chronic Kidney Disease Initiative | CDC." [Online]. Available: https://www.cdc.gov/kidneydisease/basics.html. [Accessed: 12-Dec-2018].

[6] M. R. Ahmed, M. Arifa Khatun, A. Ali, and K. Sundaraj, "A literature review on NoSQL

database for big data processing," *Int. J. Eng. Technol.*, vol. 7, no. 2, pp. 902–906, 2018.

[7] R. Hossain, S. M. H. Mahmud, M. A. Hossin, S. R. Haider Noori, and H. Jahan, "PRMT: Predicting Risk Factor of Obesity among Middle-Aged People Using Data Mining Techniques," *Procedia Comput. Sci.*, vol. 132, pp. 1068–1076, Jan. 2018.

[8] A. K. Dwivedi, "Analysis of computational intelligence techniques for diabetes mellitus prediction," *Neural Comput. Appl.*, pp. 1–9, Apr. 2017.

[9] S. M. H. Mahmud and R. Ahmed, "Machine Learning Based Unified Framework for Diabetes Prediction," 2018.

[10] M. Heydari, M. Teimouri, Z. Heshmati, and S. M. Alavinia, "Comparison of various classification algorithms in the diagnosis of type 2 diabetes in Iran," *Int. J. Diabetes Dev. Ctries.*, vol. 36, no. 2, pp. 167–173, Jun. 2016.

[11] M. Kukar, I. Kononenko, C. Grošelj, … K. K.-A. intelligence in, and undefined 1999, "Analysing and improving the diagnosis of ischaemic heart disease with machine learning," *Elsevier*.

[12] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egypt. Informatics J.*, Apr. 2018.

[13] D. Carvalho, P. R. Pinheiro, and M. C. D. Pinheiro, "A Hybrid Model to Support the Early Diagnosis of Breast Cancer," *Procedia Comput. Sci.*, vol. 91, pp. 927–934, Jan. 2016.

[14] M. Kumari, "Breast Cancer Prediction system," *Procedia Comput. Sci.*, vol. 132, pp. 371–376, Jan. 2018.

[15] L. Tapak, N. Shirmohammadi-Khorram, P. Amini, B. Alafchi, O. Hamidi, and J. Poorolajal, "Prediction of survival and metastasis in breast cancer patients using machine learning classifiers," *Clin. Epidemiol. Glob. Heal.*, Oct. 2018.

[16] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using Machine Learning Algorithms

for Breast Cancer Risk Prediction and Diagnosis," *Procedia Comput. Sci.*, vol. 83, pp. 1064–1069, Jan. 2016.

[17] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *Eur. J. Oper. Res.*, vol. 267, no. 2, pp. 687–699, Jun. 2018.

[18] M. Gupta and B. Gupta, "A Comparative Study of Breast Cancer Diagnosis Using Supervised Machine Learning Techniques," in *2018 Second International Conference on Computing Methodologies and Communication (ICCMC)*, 2018, pp. 997–1002.

[19] L. Abdel-Ilah and H. Šahinbegović, "Using machine learning tool in classification of breast cancer," Springer, Singapore, 2017, pp. 3–8.

[20] M. Amrane, S. Oukid, I. Gagaoua, and T. Ensari, "Breast cancer classification using machine learning," in *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, 2018, pp. 1–4.

[21] M. R. Al-Hadidi, A. Alarabeyyat, and M. Alhanahnah, "Breast Cancer Detection Using K-Nearest Neighbor Machine Learning Algorithm," in *2016 9th International Conference on Developments in eSystems Engineering (DeSE)*, 2016, pp. 35–39.

[22] X. Liu *et al.*, "A Hybrid Classification System for Heart Disease Diagnosis Based on the RFRS Method.," *Comput. Math. Methods Med.*, vol. 2017, p. 8272091, 2017.

[23] B. Ramana, M. Babu, N. V.-I. J. of, and undefined 2012, "A critical comparative study of liver patients from USA and INDIA: an exploratory analysis," *Citeseer*.

[24] "Confusion Matrix." [Online]. Available: http://www2.cs.uregina.ca/~hamilton/courses/831/notes/confusion_matrix/confusion_matrix.html. [Accessed: 20-Dec-2018].

Appendix

Appendix A: Liver Disease Dataset

Sample of Indian Liver Patient Data

Age	Gender	Total_Bili	Direct_Bil	Alkaline_I	Alamine_A	Aspartate_	Total_Pro	Albumin	Albumin_	Dataset
65	Female	0.7	0.1	187	16	18	6.8	3.3	0.9	1
62	Male	10.9	5.5	699	64	100	7.5	3.2	0.74	1
62	Male	7.3	4.1	490	60	68	7	3.3	0.89	1
58	Male	1	0.4	182	14	20	6.8	3.4	1	1
72	Male	3.9	2	195	27	59	7.3	2.4	0.4	1
46	Male	1.8	0.7	208	19	14	7.6	4.4	1.3	1
26	Female	0.9	0.2	154	16	12	7	3.5	1	1
29	Female	0.9	0.3	202	14	11	6.7	3.6	1.1	1
17	Male	0.9	0.3	202	22	19	7.4	4.1	1.2	2
55	Male	0.7	0.2	290	53	58	6.8	3.4	1	1
57	Male	0.6	0.1	210	51	59	5.9	2.7	0.8	1
72	Male	2.7	1.3	260	31	56	7.4	3	0.6	1
64	Male	0.9	0.3	310	61	58	7	3.4	0.9	2
74	Female	1.1	0.4	214	22	30	8.1	4.1	1	1
61	Male	0.7	0.2	145	53	41	5.8	2.7	0.87	1
25	Male	0.6	0.1	183	91	53	5.5	2.3	0.7	2
38	Male	1.8	0.8	342	168	441	7.6	4.4	1.3	1
33	Male	1.6	0.5	165	15	23	7.3	3.5	0.92	2
40	Female	0.9	0.3	293	232	245	6.8	3.1	0.8	1
40	Female	0.9	0.3	293	232	245	6.8	3.1	0.8	1
51	Male	2.2	1	610	17	28	7.3	2.6	0.55	1
51	Male	2.9	1.3	482	22	34	7	2.4	0.5	1
62	Male	6.8	3	542	116	66	6.4	3.1	0.9	1
40	Male	1.9	1	231	16	55	4.3	1.6	0.6	1

A Study of Liver Disease Prediction Using Machine Learning Approaches

by Abdur Rahman Bhuiyan

Submission date: 05-Dec-2019 01:45PM (UTC+0600)
Submission ID: 1227578130
File name: Thesis-MSc-Abdur_Rahman-Liver_disease_prediction.pdf (783.62K)
Word count: 4872
Character count: 25827

A Study of Liver Disease Prediction Using Machine Learning Approaches

ORIGINALITY REPORT

27%	20%	10%	24%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Kennesaw State University Student Paper	7%
2	Submitted to Daffodil International University Student Paper	6%
3	Submitted to NCC Education Student Paper	2%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	2%
5	Md. Razu Ahmed, S M Hasan Mahmud, Md Altab Hossin, Hosney Jahan, Sheak Rashed Haider Noori. "A Cloud Based Four-Tier Architecture for Early Detection of Heart Disease with Machine Learning Algorithms", 2018 IEEE 4th International Conference on Computer and Communications (ICCC), 2018 Publication	1%
6	Submitted to Korea University of Technology and Education Student Paper	1%

7	Emanuel Ontiveros-Robles, Patricia Melin, Oscar Castillo. "Relevance of Polynomial Order in Takagi-Sugeno Fuzzy Inference Systems Applied in Diagnosis Problems", 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2019 Publication	1%
8	"Machine Learning Methods for Diabetes Prediction", International Journal of Innovative Technology and Exploring Engineering, 2019 Publication	1%
9	Submitted to Infile Student Paper	<1%
10	Submitted to Universiti Kebangsaan Malaysia Student Paper	<1%
11	"Parkinson's Disease Prediction using Modified Gauss-Newton Method in Feed-Forward Neural Network", International Journal of Innovative Technology and Exploring Engineering, 2019 Publication	<1%
12	Submitted to University of Strathclyde Student Paper	<1%
13	S. M. Hasan Mahmud, Md Altab Hossin, Md. Razu Ahmed, Sheak Rashed Haider Noori, Md Nazirul Islam Sarkar. "Machine Learning Based Unified Framework for Diabetes Prediction",	<1%

	Proceedings of the 2018 International Conference on Big Data Engineering and Technology - BDET 2018, 2018 Publication	
14	Submitted to University of Dammam Student Paper	<1%
15	Mesut Toğaçar, Burhan Ergen, Zafer Cömert. "Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders", Medical Hypotheses, 2020 Publication	<1%
16	Submitted to CONACYT Student Paper	<1%
17	Submitted to The Center for Allied Health & Nursing Education Student Paper	<1%
18	Submitted to University of Texas at Tyler Student Paper	<1%
19	Submitted to University of Lancaster Student Paper	<1%
20	Submitted to National College of Ireland Student Paper	<1%

21	mafiadoc.com Internet Source	<1%
22	Submitted to Guru Nanak Dev Engineering College Student Paper	<1%
23	www.canberra.edu.au Internet Source	<1%
24	Submitted to Universiti Malaysia Pahang Student Paper	<1%
25	Submitted to University of South Australia Student Paper	<1%
26	Submitted to Blinn College Student Paper	<1%
27	es.scribd.com Internet Source	<1%
28	Submitted to Birkbeck College Student Paper	<1%
29	"Intelligent Systems and Applications", Springer Science and Business Media LLC, 2019 Publication	<1%
30	www.cddft.nhs.uk Internet Source	<1%
31	bioinfolab.unl.edu Internet Source	<1%

32	ethos.bl.uk Internet Source	<1%
33	ethesys.library.ttu.edu.tw Internet Source	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		