A DETECTION AND SEGMENTATION OF MEDICAL IMAGE USING MACHINE LEARNING ALGORITHMS

BY

MD. NAZMUL HASSAN
ID: 183-15-11805
AND

TABASSUM PROGGA
ID: 183-15-12001

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

Supervised By

Professor Dr. Md. Ismail Jabiullah
Professor
Department of CSE
Daffodil International University

Co-Supervised By

Md. Jueal Mia
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

SEPTEMBER 2022

APPROVAL

This Project titled **A Detection and Segmentation of Medical Image Using Machine Learning Algorithms,** submitted by Md. Nazmul Hassan and Tabassum Progga, ID No. 183-15-11805 and 183-15-12001 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements of the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation was held on *12/09/2022*.

BOARD OF EXAMINERS

23/09/22

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman

Dr. Md. Monzur Morshed [DMM]
Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Ms. Samia Nawshin [SN]
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Dewan Md Farid
Professor
Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

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

Supervised by:

Professor Dr. Md. Ismail Jabiullah
Professor
Department of CSE
Daffodil International University

Submitted by:

(Md. Nazmul Hassan)
ID: 183-15-11805
Department of CSE
Daffodil International University

Progga

(Tabassum Progga)
ID: 183-15-12001
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project successfully.

We really grateful and wish our profound our indebtedness to **Professor Dr. Md. Ismail Jabiullah, Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our 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.

We would like to express our heartiest gratitude to Md. Ismail Jabiullah and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Steatosis, often known as fatty liver disease, is a common disorder brought on by an overgrowth of fat in the liver. There is a modest amount of fat in a healthy liver. A problem arises when fat makes up 5% to 10% or more of the weight of our liver. This illness is linked to a high risk of morbidity and mortality. The goal of our research proposal is to create a machine learning model that can help physicians identify high-risk patients, provide a distinct diagnosis, and prevent and treat FLD. 55 of the 105 people who were recruited had their FLD assessed. The sickness was found using classifier models. Among the models available are decision trees, K-Neighbors classifiers, logistic regressions (LR), random forests (RF), and linear regressions and Decision Tree Classifier. The link between the model's actual and anticipated values was examined using the confusion matrix. According to the study's findings, machine learning algorithms can accurately predict FLD. According to our research, the Logistic Regression model is the most reliable regression model. The Random-Forest-Classifier has a higher accuracy of 68 percent than the other classification models. Our method could be used to find FLD patients who could have a big impact on how treatments are delivered. The use of this algorithm for early prediction could lower medical expenses and save money on treatment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Introduction

A typical condition is fatty liver disease. It has a high rate of mortality and morbidity. It has been determined that FLD causes a financial expense. Hepatocellular carcinoma and non-cholestatic cirrhosis are long-term effects of FLD [1]. Furthermore, obesity, metabolic syndrome, and diabetes have all increased along with FLD prevalence [2].

1.2 Motivation

FLD is thought to be a negative for the economy. The diagnosis, treatment, and prevention of the disease could be significantly impacted by earlier detection and more precise identification of those who are at high risk for it. In recent years, the biopsy has been primarily used to diagnose patients with this condition and is thought to be the gold standard. This method, however, is more expensive and intrusive. Its use may have specific adverse consequences or sampling flaws. The development of more efficient methods is necessary since the accuracy of ultrasonography in disease diagnosis greatly depends on the user. Machine learning (ML) is an optional approach that could be able to overcome these restrictions. As a result, we suggested a machine learning model that will predict FLD, and we have high hopes that both patients and doctors will benefit from our approach.

1.3 Objective

Due to the use of various ultrasound devices, poor picture quality, and patient physical differences, the diagnosis of fatty liver by ultrasonic imaging differs. On the other hand, a prediction model based on easily accessible clinical traits will help medical professionals in accurately detecting and making judgments on effective preventative measures, early diagnosis, and targeted interventions. There hasn't been a lot of in-depth, extensive study done on the use of models based on electronic medical data to forecast FLD. This is why we want to create a predictive model for fatty liver disease using cutting-edge machine learning techniques, particularly the classification method.

1.4 Research Questions

i) How early detection and more accurate identification of Fatty Liver Disease at high risk for the disease could have a significant impact on its diagnosis, treatment, and prevention?

ii) Is machine learning (ML) able to address the limitations of conventional approach to categorize Fatty Liver Disease patients? If yes then how?

1.5 Problem Statement

Higher FLD prevalence has been associated to higher economic costs. Biopsy has been used tostratify persons with this disease in recent years, andit has been recognized as the gold standard of diagnosis. This surgery, however, is more painful andcostly. It has the potential to cause various adverse effects or sampling problems when utilized. Ultrasonography's accuracy in illness diagnosis is highly dependent on the operator [3], demanding the development of more efficient methods. As a result, accurate risk identification and early detection of FLD could have major implications for diagnosis, prevention, and, eventually, proper treatment. Machine learning (ML) is a different approach that might be able to solve these problems. The field of computer science known as machine learning (ML) use computer algorithms to find patterns in massive amounts of data and forecast different outcomes [4]. In a number of fields, machine learning techniques have emerged as a viable tool for prediction and decision-making [5]. Clinical data are readily available, hence ML has also been crucial in helping doctors make decisions [6, 7]. Machine learning can enhance prediction and uncover latent features that aren't visible but can be extracted from other data. Several machine learning (ML) techniques, such as logistic regression (LR), random forest (RF), artificial neural networks (ANNs), K- nearest neighbors (KNNs), support vector machines, extreme gradient boosting (XGBoost), and linear discriminant analysis, are currently being used in disease prediction with significantly higher accuracy than conventional methods (LDA) [8].

1.6 Expected Output

Our intention is to build a model that can assist doctors identify high risk Fatty Liver Disease patients so that they can start early treatment and prevent heavy damage to patients. We expect that our model will be able to classify patients with high risk of having FLD with more than 60 percent of accuracy.

1.7 Report Layout

Chapter 2: Here, we review the history of our chosen approach. A variety of existing models are compared to us, and we also give details regarding related projects, opportunities, and obstacles.

Chapter 3: Give the work's engrossments. We established the various diagrams and other specifications. Additionally, we talked about the features.

Chapter 4: In this chapter we discussed about the step-by-step methods we have applied to make our project successful.

Chapter 5: Here we discussed about a descriptive analysis of the inputs and outputs of our project.

Chapter 6: In this chapter we present experiment result of our work and discussed about comparative analysis of the result and accuracy of our work with some existing work.

Chapter 7: Discussions were held about the project's scope for expansion and its conclusion.

CHAPTER 2

Background Study

2.1 Introduction

Many research has already looked into the classification of fatty liver disease using medical imaging techniques as ultrasound, computed tomography, and magnetic resonance imaging (MRI). Imaging with ultrasound is non-invasive, affordable, simple to use, and portable. However, due to the use of various ultrasound machines, low picture quality, and patient physical differences, the diagnosis of fatty liver on ultrasound imaging varies. A prediction model based on readily accessible clinical characteristics, however, would assist doctors in accurately identifying and making a decision regarding prevention, early diagnosis, and focused intervention. The advantages of using classification models with information from the electronic medical record to predict FDL have not yet been thoroughly assessed. Therefore, our goal was to develop a predictive model for fatty liver disease employing cutting-edge machine learning techniques, particularly in the classification method. This is the largest study that has, to our knowledge, used machine learning models to predict FDL. In this chapter, we make an effort to provide an overview of earlier works that are relevant to our work as well as a sense of what we plan to accomplish.

2.2 Related Works

Ma, H., Xu, C. F., Shen, Z., Yu, C. H., and Li, Y.M. (2018) [9] conducted a cross sectional investigation utilizing machine learning approaches to predict Non-Alcoholic Fatty Liver Disease (NAFLD). They employed ultrasonographic methods to diagnose NAFLD, which has certain drawbacks but is still the most widely used method with good accuracy. Their findings offer valuable insight into the use of innovative AI approaches to predict NAFLD. Their model has some flaws as well, such as the lack of patient liver tissue biopsy data. Patients' liver tissue biopsy data could be added in future studies to construct a more accurate predictive ML model. Andrade, A., Silva, J. S., Santos, J., and

Belo-Soares, P. (2012) [10] evaluated the effectiveness of three classifier algorithms for hepatic steatosis diagnosis (Fatty Liver). This study provides a semi- automated categorization technique for evaluating steatotic liver tissues based on B-scan ultrasound data. Several features were extracted using Artificial Neural Networks (ANN), Support Vector Machines (SVM), and k-Nearest Neighbors (k-NN) and then used in three different classifiers (kNN). The classifiers were trained using the 10-cross validation approach. A stepwise regression-based feature selection method was also applied, which enhanced prediction accuracy. To build a more accurate and robust model more patients data is required. Feature extraction methods upon the original US images could be implemented to get more promising results. Furthermore, other classification methods like neural network and support vector machine classifiers should be implemented to enhance the efficiency of the process. R. Ribeiro and J. Sanches published an automatic classification system for the diagnosis of liver steatosis (FLD) from ultrasound pictures in June 2009 [11]. The features were chosen so that clinicians could diagnose the condition simply by looking at the ultrasound images. The outcome of their planned work is quite promising in terms of sensitivity and specificity. This project has several limits as well. More patient data is needed to construct a more accurate and robust model. To generate more promising findings, feature extraction methods on the original US photos could be used. Other classification approaches, such as neural networks and support vector machine classifiers, should also be used to improve the process' efficiency. Based on the features of liver ultrasonography imaging and clinical fatty liver diagnostic criteria A SVM classification model is built by Li, G., Luo, Y., Deng, W., Xu, X., Liu, A., and Song, E., (2008) [12]. They were able to diagnose fatty and normal livers with 97.1 percent and 84 percent accuracy, respectively. Although this model has decent accuracy, more work is needed to add more enhancing features. Future study could focus on how to get consistent diagnosis and quantitative analysis results from ultrasound images acquired on various B-scan equipment. Pei, X., Deng, Q., Liu, Z., Yan, X., and Sun, W., 2021 [13] found that when the variables are few, ML models, particularly XGBoost, have superior accuracy for FLD prediction. This method may be more beneficial in real-world clinical practice. In order to predict FLD in a variety of datasets, this model will need to be tested in the future. The authors of [14] wanted to create a machine learning model that could help doctors classify high-risk patients and make a novel diagnosis, as well as prevent and manage FLD. Chieh-Chen Wu, Wen-Chun Yen, Wen-Ding Hsu, Md. Mohaimenul Islam, Phung Anh(Alex) Nguyen, Tahmina Nasrin Poly, Yao-Chin Wang, Hsuan-Chia Yang Their approach has the ability to detect FLD early, which would improve therapy precision and effectiveness. They only obtained data from a single medical facility. External validation and multicenterdatasets, on the other hand, may improve performance and dependability. They were unable to distinguish between patients with fatty and non-fattyliver disease due to a lack of data. They simply utilized a classification strategy; however, a deeplearning approach may have been used to improve prediction. Birjandi et al. [22] created another classification model that uses a classification tree (CT) to predict the likelihood of developing NAFLD and identifies the most crucial variables that affect the condition. BMI, WHR, triglycerides, glucose, SBP, and alanine aminotransferase were the primary potential variables for predicting NAFLD based on the CT, and the model achieved a prediction accuracy of 80% with a receiver operating characteristic (ROC) curve area of 78%. Additionally, Jamali et al. [23] created a model using serum adipokines to distinguish between NAFLD and healthy people as well as between nonalcoholic steatohepatitis (NASH) and simple steatosis. 86.4% of the initially grouped cases were appropriately categorized based on NAFLD discriminant score. A machine learning model based on laboratory parameters was created and validated by Yip et al. [24] to identify NAFLD in the general population. 922 participants from a population screening study were randomly assigned to training and validation groups, and 23 standard clinical and laboratory parameters were measured following elastic net regulation.

2.3 Summary

A frequent clinical consequence that is linked to high morbidity and death is fatty liver disease (FLD). The potential to develop a suitable plan for prevention, early diagnosis, and therapy is given by an early prediction of FLD patients. Our goal was to create a machine learning model that could predict FLD and help doctors identify high-risk patients, establish a new diagnosis, prevent, and manage FLD.

CHAPTER 3

Proposed System

3.1 Introduction

Fatty Liver Disease is a common condition. It has a significant morbidity rate. It has a great impact on economy too. Traditional system or technique of detecting fatty liver disease is expensive as well as painful for patients. An early detection of this ailment is possible with the help of Machine Learning technique and it will help both doctor and patient to a great extent. Therefore, we proposed a machine learning model to classify high risk patients. Our predictive model will give an early prediction result for high-risk fatty liver disease patients.

3.2 Description

At first, we will gather data and information from various clinics. Patients who had taken an initial fatty liver treatment will be included in study. After that a dataset will be prepared. In that dataset there should be predictive and predictor variable. The dataset needs to be preprocessed for building our model. Finally, after the preprocessing done, we will have our final dataset ready for model building.

The final dataset will be used to build our predictive model by implementing various classification and regression machine learning algorithms. The algorithms we will be using are Linear Regression, Logistic Regression. Random Forest, KNeighbour Classifier, KNeighbour Regressor etc. After building the predictive model we will evaluate it. After a successful evaluation our model will be prepared.

3.3 Process Diagram

Figure 3.3 illustrates the ML(machine learning) process.

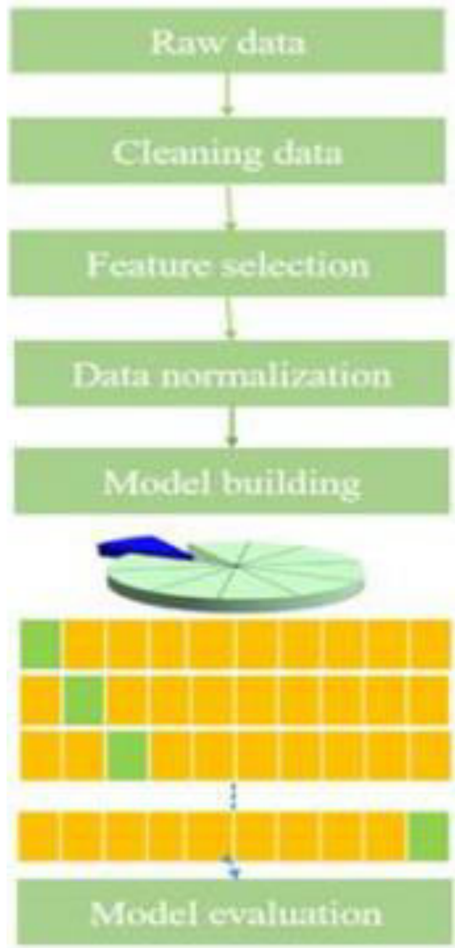

Figure 3.1: Process Diagram

To build our model we had go through a step-by-step process that is given above as a diagram. At first, we collected our raw data by which our initial dataset is comprised of. After that we cleaned our data. We deleted the data of any variables having more than 50% of missing values because missing values create outliers. After that we used feature selection technique to select the best set of variables for our model. After doing that we split our data into training and testing set to build our model. Training data was used to train our model and testing data was used to test our model. After building our model we evaluated it with confusion matrix.

3.4 Flow Chart Diagram

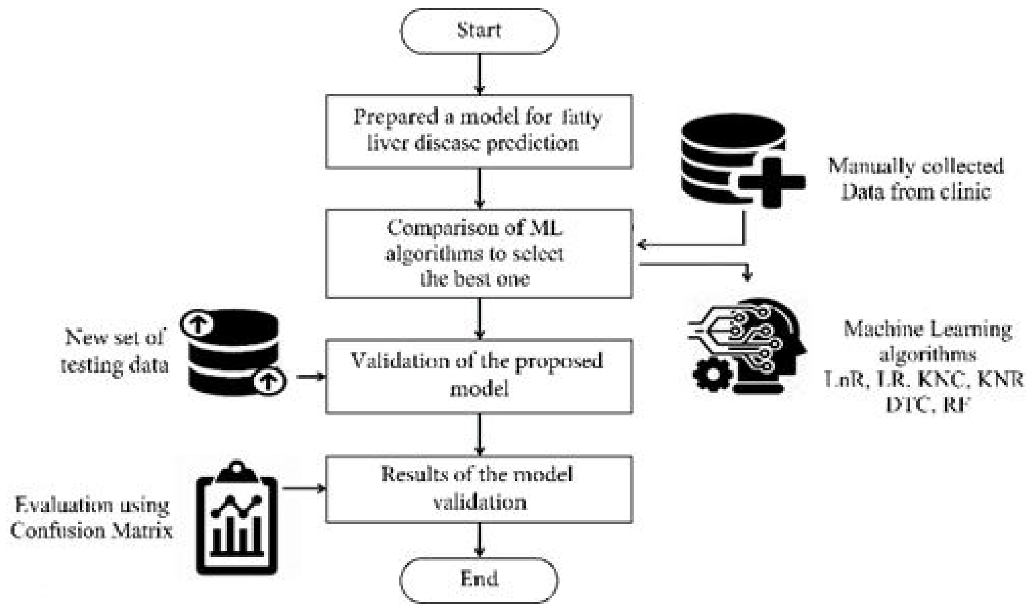

Figure 3.2: Flow Chart Diagram

3.5 Requirement Analysis

To make our project successful we will use google colab for our data analysis, data preprocessing and model building. The minimum software and hardware that are required to run google colab are as follows:

i) Software requirement (minimum):

OS: Windows 7 SP1

Browser: Google Chrome, Mozilla Firefox etc.

ii) Hardware Requirement:

Memory: 4 GB

Graphics Card: NVIDIA GeForce GTX 970

CPU: Intel Core i5-4590

File Size: 2 GB

3.6 Summary

FLD is a high-risk disease with greater morbidity rate. An early prediction could save patients life. Machine learning has the potential to predict FLD in the early stage. Therefore, we try to build a ML model that will help doctors to figure out high risk patients. We will collect data, prepare data and finally use those data to build our predictive model. After building we will evaluate the performance of our model.

CHAPTER 4

Methodology

4.1 Population Under Study

We gathered data from the Family Care Foundation Located in Sabamoon Tower, 152, 2, Green Rd, Dhaka-1205 as part of a liver protection initiative. All patients who had undergone initial fatty liver treatment were included in the study in December 2021. Patients were excluded if they were,

a) under 30 years old,

b) had a history of mental illness, or

c) had a history of substance abuse.

d) incomplete examination process

e) an ultrasound test revealed a suspected case of fatty liver.

Liver	Biliary channel	CBD	Gall bladder	Pancreas	Spleen	Kidney	Urinary bladder	Prostate	PVR	Prediction
Liver is normal in size with homogeneous echotexture. Parenchymal echogenicity is increased. No focal lesion is seen.	Intra hepatic biliary channel are not dilated.	Not dilated.	Normal	Appears normal	Normal in size & uniform echopattern.	Both kidneys are normal in size & shape. Cortical echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well-filled & regular in outline. No intravesical calculus is seen.	Normal in size.		Fatty change in Liver
Liver is enlarged in size with increased echogenicity. No focal lesion is seen.	Intra hepatic biliary channel are not dilated.	Not dilated.	Contracted (Post Prandial).	Appears normal	Normal in size & uniform echopattern.	Both kidneys are normal in size & shape. Cortical echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well-filled & regular in outline. No intravesical calculus is seen.	Enlarged in size (Volume – 29gm), with intravesical indentation.	Nil	Fatty Change in Liver
Liver is enlarged in size with increased echogenicity. No focal lesion is seen.	Intra hepatic biliary channel are not dilated	Not dilated.	No stone or lesion could be detected.	Appears normal.	Normal in size & uniform echopattern	Both kidneys are normal in size & shape. Cortical echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well-filled & regular in outline. No intravesical calculus is seen.	Normal in size, with no intravesical indentation.		Fatty Change in Liver.
Liver is enlarged in size with increased echogenicity. No focal lesion is seen.	Intra hepatic biliary channel are not dilated	Not dilated	No stone or lesion could be detected.	Appears normal.	Normal in size & uniform echopattern.	echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well-filled & regular in outline. No intravesical calculus is seen.	Normal in size, with no intravesical indentation.		Fatty Change in Liver(Moderate).
Liver is normal in size with increased echogenicity. No focal lesion is seen.	Intra hepatic biliary channel are not dilated	Not dilated.	No stone or lesion could be detected.	Appears normal.	Normal in size & uniform echopattern.	Both kidneys are normal in size & shape. Cortical echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well-filled & regular in outline. No intravesical calculus is seen.	Normal in size (15 cc), with no intravesical indentation		Fatty Change in Liver.
Liver is normal in size			No stone or							

Figure 4.1: Initial Dataset

Liver is normal in size with normal echogenicity. No focal lesion is seen.	Intra hepatic biliary channel are not dilated.	Not dilated .	No stone or mass lesion could be detected	Appears normal	Normal in size & uniform echopattern	Both kidneys are normal in size & shape. Cortical echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well filled. No stone or mass lesion could be detected.			Normal study.
Liver is normal in size with normal echogenicity. No focal lesion is seen.	Intra hepatic biliary channel are not dilated.	Not dilated .	No stone or mass could be detected.	Appears normal	Normal in size & uniform echopattern	Both kidneys are normal in size & shape. Cortical echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well filled. No stone or mass lesion could be detected.			Normal study.
Liver is normal in size with normal echogenicity. No focal lesion is seen.	Intra hepatic biliary channel are not dilated.	Not dilated .	No stone or mass lesion could be detected	Appears normal	Normal in size & uniform echopattern	Both kidneys are normal in size & shape. Cortical echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well filled. No stone or mass lesion could be detected.			Normal study.
Liver is normal in size with normal echogenicity. No focal lesion is seen.	Intra hepatic biliary channel are not dilated.	Not dilated .	No stone or mass lesion could be detected	Appears normal	Normal in size & uniform echopattern	Both kidneys are normal in size & shape. Cortical echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well filled. No stone or mass lesion could be detected.			Normal study.
Liver is normal in size with normal echogenicity. No focal lesion is seen.	Intra hepatic biliary channel are not dilated.	Not dilated .	Normal. No stone or mass lesion could be	Appears normal	Normal in size & uniform echopattern	Both kidneys are normal in size & shape. Cortical echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well filled. No stone or mass lesion could be detected.			Normal study.
Liver is normal in size with normal echogenicity. No focal lesion is seen.	Intra hepatic biliary channel are not dilated.	Not dilated .	No stone or mass lesion could be detected	Appears normal	Normal in size & uniform echopattern	Both kidneys are normal in size & shape. Cortical echogenicity is normal. Cortico medullary differentiation is well. Pelvicalyceal systems of both kidneys are not dilated.	Well filled. No stone or mass lesion could be detected.			Normal study.
Liver is normal in size	Intra hepatic	Not	No stone or mass			Both kidneys are normal in size & shape. Cortical	Well filled. No stone or			

Figure 4.2: Final Dataset

4.2 Machine Learning

The major goal of this work was to use classification machine learning models to choose prognostic markers for the prediction of fatty liver disease. We divided our machine learning approach into three pieces throughout this phase:

1. Data cleansing, missing data resolution, data transformation, and data imbalance reduction are a few examples of data preprocessing.

2. Variables selection: a method for choosing the most pertinent set of variables to include in the model's formulation (helps decrease overfitting, increase accuracy, and shorten training time).

3. Model construction: choose a suitable classification for better prediction.

4.3 Data Preprocessing

This study comprised a total of 120 patients, with 62 of them being diagnosed with fatty liver disease. Because data preparation is a key stage in machine learning, we eliminated any variables with more than 50% missing values.

We used the information gain ranking technique to determine the weight of each variable in this approach. It aided in determining the efficacy of factors included in the training dataset. Onlyvariables with a score greater than 0 were included inthe final model.

4.4 Model Building

Patients with FLD were successfully identified by predictive classifier systems. Prediction models were created using classification approaches such random forest (RF), decision tree classifier (DTC), logistic regression (LR), and linear regression. We settle on these four models due to the following characteristic. The random forest (RF) ensemble classification technique was created in 1999 by Leo Breiman and Adele Culter [15]. This system is composed of a sizable number of decision trees. A randomly selected sample of the training set is used to build the tree, which is constructed separately using the generic technique of bootstrap aggregation (also known as bagging). A simple majority vote of all trees determines the ultimate outcome. In a wide range of applications, including medical diagnostics, RF has demonstrated to be a highly accurate method. Artificial neural networks are computer simulations of organic brain networks (ANN). It is a potent nonlinear model that has been shown to yield precise predictions in a number of CDS [16]. Perceptrons, which are synthetic brain cells, make up this model [17]. Similar to how they do in a typical neural cell, dendrites of an ANN convey the signal into the neuron. Replicated signal transmission occurs from an input layer via several concealed levels, and then to an output layer. On the other hand, each layer has a sizable number of perceptrons, and the perceptrons between layers are connected by various weights that can be changed during training. Up until each input matches the corrected input, it learns from a huge number of samples in the training dataset. A discrete choice model that is a part of the multivariate analysis family is logistics regression (LR). It is the most popular method of empirical analysis and is frequently used to contrast with machine learning research in the fields of sociology, biostatistics, clinical medicine, quantitative psychology, econometrics, and marketing [18]. It offers a variety of advantages, including a great deal of strength and accuracy.

4.5 Summary

A total of 105 people were found, and FLDs were assessed on 55 of them. The disease was identified using classification methods. There are several models available, including decision tree classification, logistic regression, K-Neighbors classifier, and RF.

Chapter 5
Result and Discussion

5.1 Introduction

In previous chapter we discussed about the methodology of our project. In this chapter we will discuss about how our model is implemented in our data analysis platform. We also try to discuss about the input and output analysis in this chapter.

5.2 Input Analysis

At first, we imported our dataset in our data analysis platform. After that various data preprocessing techniques were applied.

5.3 Output Analysis

The confusion matrix was used to figure out howthe actual and predicted values in the model related. [19] Table 1 shows the structure of confusion matrixof the research.

Table 5.1: CONFUSION MATRIX REPRESENTATION

	Positive	Negative
Predicted true(+)	TP	TN
Predicted false (-)	FP	FN

Accuracy: To measure model accuracy, divide learning models in medical practice to provide tailored patient care. By dividing the total number of positive instances by the total number of cases, doctors may be able to extract the minimal amount of data required to make a therapeutic decision. The percentage of accurately detected cases is provided by the accuracy parameter. The correctness of the model is defined as

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

	precision	recall	f1-score	support
accuracy			0.68	28
macro avg	0.07	0.07	0.07	28
weighted avg	0.68	0.68	0.68	28

Figure 5.1: Classification report of Decision Tree Classifier Algorithm

	precision	recall	f1-score	support
accuracy			0.68	28
macro avg	0.07	0.08	0.08	28
weighted avg	0.59	0.68	0.63	28

Figure 5.2: Classification report of Logistic Regression Algorithm

	precision	recall	f1-score	support
accuracy			0.68	28
macro avg	0.08	0.09	0.08	28
weighted avg	0.59	0.68	0.63	28

Figure 5.3: Classification report of Kneighbor Classifier Algorithm

	precision	recall	f1-score	support
accuracy			0.68	28
macro avg	0.07	0.07	0.07	28
weighted avg	0.64	0.68	0.66	28

Figure 5.4: Classification report of Kneighbor Regressor Algorithm

	precision	recall	f1-score	support
accuracy			0.68	28
macro avg	0.07	0.07	0.07	28
weighted avg	0.64	0.68	0.66	28

Figure 5.5: Classification report of Random Forest Algorithm

Sensitivity: Sensitivity is defined as the ability todetermine the degree of a trait in order to appropriately classify a person with disorders. The equation is as follows,

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

Specificity: Specificity is defined as the degree to which an attribute is used to appropriately classify a person who is free of ailments.

$$\text{Specificity} = \frac{TN}{TN+FP}$$

To characterize the quality of the predicted class, quality measures such as sensitivity and specificity are utilized. Three metrics are used to evaluate the sensitivity, specificity, and accuracy of a medical diagnostic model. Precision, sensitivity, and specificity are the three criteria.

5.4 Discussion

The amount of data used in healthcare is always growing, and machine learning makes it possible to examine huge volumes of data quickly [20]. Therefore, there is a chance to employ machine choice [21]. The findings of this study demonstrate that machine learning techniques are suitable for reliably forecasting FLD. The most accurate regression model is the logistic regression. A typical consequence of critical illness that carries an increased risk of mortality and morbidity is fatty liver disease. Traditional diagnostic and therapeutic approaches have improved our understanding of FLD in recent years. However, it could also squander resources and have unfavorable effects. When compared to conventional statistical models, machine learning techniques always offer more information. To predict FLD, we created and assessed a classification machine learning model. The Random-Forest- Classifier, among the classification models, has a higher accuracy of 68 percent. This study has a number of issues that must be addressed. To begin with, we simply obtained data from one medical center. External validation and multicenter datasets, on the other hand, may increase performance and dependability. Second, there is a dearth of information. In our study, the patients' BMI was not mentioned.

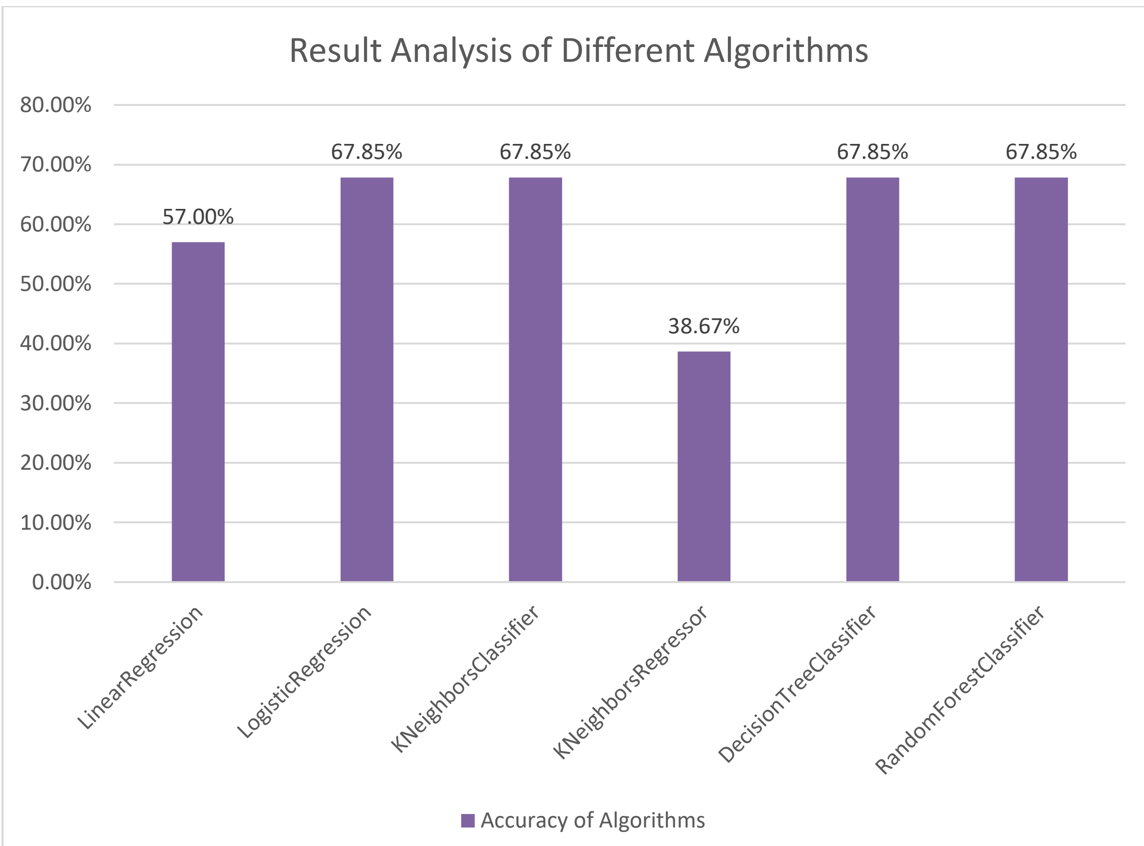

Figure 5.6: Result Analysis of Different Algorithms

CHAPTER 6
Comparative Analysis

6.1 Introduction

The results of this study suggest that machine learning algorithms can predict FLD accurately. Comparatively speaking, the random forest model performed better than other prediction categorization models. A frequent consequence of critical illness that is linked to greater mortality and morbidity is fatty liver disease. Traditional diagnostic and therapeutic approaches have recently helped us understand FLD better. However, it can occasionally have negative impacts and squander resources. However, when compared to conventional statistical models, machine learning models always offer a considerable insight. So, in order to forecast FLD, we created and assessed a classification machine learning model. Due to data insufficiency our model accuracy is not as expected, but our model has the potential to perform better if we collect and provide more data.

6.2 Comparative Chart

Table 6.1: COMPARATIVE CHART BETWEEN EXISTING WORKS AND PROPOSED WORK

Work	**Algorithm**	**Accuracy**
Application of machine learning techniques for clinical predictive modeling: a cross-sectional study on nonalcoholic fatty liver disease in China. BioMed research international	**LR**	**83.41%**
Classifier approaches for liver steatosis using ultrasound images.	**SVM**	**66.41%**
Prediction of Fatty Liver Disease using Machine Learning Algorithms	**RF**	**86.48%**
Proposed System	**RF**	**68%**

CHAPTER 7

Conclusion and Future Scope

7.1 Summary

In this work we have used various Machine Learning algorithms to build a predictive model that can predict high-risk fatty liver disease patients. This study is formed by a total of 120 patients data, with 62 of them being diagnosed with fatty liver disease. A few classification and regression algorithms were used in this work and these are Linear Regression, Logistic Regression, Kneighbor Classifier, Kneighbor Regressor, Decision Tree Classifier and Random Forest.

7.2 Conclusion

The results demonstrate that individuals with fatty liver disease may be reliably identified by a machine learning classification model, particularly the random forest model, using just a few clinical characteristics. This approach could be more useful in actual clinical settings. For both prevention and treatment, machine learning algorithms may be used to more accurately diagnose fatty liver.

7.3 Future Scope

In order to improve the diagnostic efficiency and operability, future work will address issues like how to quantify ultrasound image characteristics and categorize fatty liver into more precise categories, how to select the most useful features from a large number of ultrasound image features, and how to establish a useful man-machine interface.

References:

[1]Lazo, M., & Clark, J. M. (2008, November). The epidemiology of nonalcoholic fatty liver disease: a global perspective. In Seminars in liver disease (Vol. 28, No. 04, pp. 339-350). © Thieme Medical Publishers.

[2] Le, M. H., Devaki, P., Ha, N. B., Jun, D. W., Te, H. S., Cheung, R. C., & Nguyen, M. H. (2017). Prevalence of non-alcoholic fatty liver disease and risk factors for advanced fibrosis and mortality in the United States. PloS one, 12(3), e0173499.

[3] Anstee, Q. M., Targher, G., & Day, C. P. (2013). Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis. Nature reviews Gastroenterology & hepatology, 10(6), 330-344.

[4] Motwani, M., Dey, D., Berman, D. S., Germano, G., Achenbach, S., Al-Mallah, M. H., ... & Slomka, P. J. (2017). Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: a 5-year multicentre prospective registry analysis. European heart journal, 38(7), 500-507.

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. Health information science and systems, 2(1), 1-10.

[7] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[8] Hsieh, C. H., Lu, R. H., Lee, N. H., Chiu, W. T., Hsu, M. H., & Li, Y. C. (2011). Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks. Surgery, 149(1),87–93.

[9] Ma, H., Xu, C. F., Shen, Z., Yu, C. H., & Li, Y. M. (2018). Application of machine learning techniques for clinical predictive modeling: a cross-sectional study on nonalcoholic fatty liver disease in China. BioMed research international, 2018.

[10] Andrade, A., Silva, J. S., Santos, J., & Belo-Soares, P. (2012). Classifier approaches for liver steatosis using ultrasound images. Procedia Technology, 5, 763-770.

[11] Ribeiro, R., & Sanches, J. (2009, June). Fatty liver characterization and classification by ultrasound. In Iberian conference on pattern recognition and image analysis (pp. 354-361). Springer, Berlin, Heidelberg.

[12] Li, G., Luo, Y., Deng, W., Xu, X., Liu, A. and Song, E., 2008, August. Computer aided diagnosis of fatty liver ultrasonic images based on support vector machine. In 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 4768-4771). IEEE.

[13] Pei, X., Deng, Q., Liu, Z., Yan, X. and Sun, W., 2021. Machine learning algorithms for predicting fatty liver disease. Annals of Nutrition and Metabolism, 77(1), pp.38-45.

[14] Chieh-Chen Wu , Wen-Chun Yen , Wen-Ding Hsu , Md. Mohaimenul Islam , Phung Anh(Alex) Nguyen , Tahmina Nasrin Poly , Yao-Chin Wang , Hsuan-Chia Yang , Yu-Chuan (Jack) Li , Prediction of Fatty Liver Disease using Machine Learning Algorithms, Computer Methods and Programs in Biomedicine (2018)

[15] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

[16] Papadopoulos, M. C., Abel, P. M., Agranoff, D., Stich, A., Tarelli, E., Bell, B. A., ... & Krishna, S. (2004). A novel and accurate diagnostic test for human African trypanosomiasis. The Lancet, 363(9418), 1358-1363.

[17] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65(6), 386.

[18] Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. Journal of biomedical informatics, 35(5-6), 352-359.

[19] Provost, F., & Kohavi, R. (1998). Glossary of terms. Journal of Machine Learning, 30(2-3), 271-274.

[20] Condie, T., Mineiro, P., Polyzotis, N., & Weimer, M. (2013, April). Machine learning on big data. In 2013 IEEE 29th International Conference on Data Engineering (ICDE) (pp. 1242-1244). IEEE.

[21] Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. Jama, 309(13), 1351-1352.
[22] Birjandi, M., Ayatollahi, S. M. T., Pourahmad, S., & Safarpour, A. R. (2016). Prediction and diagnosis of non-alcoholic fatty liver disease (NAFLD) and identification of its associated factors using the classification tree method. Iranian Red Crescent Medical Journal, 18(11).
[23] Jamali, R., Arj, A., Razavizade, M., & Aarabi, M. H. (2016). Prediction of nonalcoholic fatty liver disease via a novel panel of serum adipokines. Medicine, 95(5).
[24] Yip, T. F., Ma, A. J., Wong, V. S., Tse, Y. K., Chan, H. Y., Yuen, P. C., & Wong, G. H. (2017). Laboratory parameter-based machine learning model for excluding non-alcoholic fatty liver disease (NAFLD) in the general population. Alimentary pharmacology & therapeutics, 46(4), 447-456.

A DETECTION AND SEGMENTATION OF MEDICAL IMAGE USING MACHINE LEARNING ALGORITHMS

ORIGINALITY REPORT

21%	13%	19%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	coek.info Internet Source	5%
2	Chieh-Chen Wu, Wen-Chun Yen, Wen-Ding Hsu, Md. Mohaimenul Islam et al. "Prediction of Fatty Liver Disease using Machine Learning Algorithms", Computer Methods and Programs in Biomedicine, 2018 Publication	3%
3	Guokuan Li, Yu Luo, Wei Deng, Xiangyang Xu, Aihua Liu, Enmin Song. "Computer aided diagnosis of fatty liver ultrasonic images based on support vector machine", 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2008 Publication	2%
4	Xieyi Pei, Qingqing Deng, Zhuo Liu, Xiang Yan, Weiping Sun. "Machine Learning Algorithms for Predicting Fatty Liver Disease", Annals of Nutrition and Metabolism, 2021 Publication	2%