

***GURU NANAK DEV ENGIN***

***EERING COLLEGE***

***BIDAR***

***-***

***585***

***403***

***(***

***Affiliated to***

***VTU Bel***

***a***

***ga***

***vi***

***, Approved by AICTE)***

**PROJECT**

**SYNOPSIS**

**O**

**F**

**“**

**LIVER DISEASE PREDICTION**

**”**

Submitted by:

1. **MD SALAM (3GN17CS043)**
2. **MOHAMMED ABDUL BASEER (3GN17CS045)**
3. **MD TAIZAN ALI (3GN17CS044)**
4. **MOHAMMED FUAADUDDIN (3GN17CS046)**

Under the guidance of

##### **Prof. Ashwini Mankal**

**DEPARTMENT OF COMPUTER SCIENCE**

**ENGINEERING**

**GURU NANAK DEV ENGINEERING COLLEGE**

**BIDAR**

**–**

**585 403**

2020 – 21



**VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELAGAVI**

GURU NANAK DEV ENGINEERING COLLEGE,

**BIDAR-585403, KARNATAKA**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**CERTIFICATE**

This is to certify that the Internship work entitled **“Liver Disease Prediction”** carried out by **Mohammed Abdul Baseer, Md Taizan Ali, Mohammed Fuaaduddin, Md Salam**, a bonafide student of Guru Nanak Dev Engineering College in partial fulfilment for the reward of **Bachelor of Engineering** in **Computer Science and Engineering** under **Visvesvaraya Technological University**, **Belagavi** during the academic year **2020-2021** is true representation of Internship work completed satisfactorily.

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Guide Co-Ordinator HOD**

**EXTERNAL VIVA**

**Examiners: 1) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**LITERATURE SURVEY OF LIVER DISEASE CLASSIFICATION**

# ABSTRACT

Currently there is one of the prevalent diseases of 21st century is liver disorders annually killing so many people’s round the worlds. A range of therapies have been provided by researcher to evaluate results. Early diagnosis is of considerable amount of significance in treating the disease. Diagnosis is of the physician skills conducting based on their knowledge’s and experience yet an error might occurrence is here.

Diagnosis of liver disease at a preliminary stage is important for better treatment. It is a very challenging task for medical researchers to predict the disease in the early stages owing to subtle symptoms. Often the symptoms become apparent when it is too late. To overcome this issue, this project aims to improve liver disease diagnosis using machine learning approaches. The main objective of this research is to use classification algorithms to identify the liver patients from healthy individuals. This project also aims to compare the classification algorithms based on their performance factors. To serve the medicinal community for the diagnosis of liver disease among patients.

Using various Artificial Intelligence methods for liver disorders diagnosis has recently become wide spreading data. These intelligent helps systems help physicians as diagnosis assistants. Now, various Artificial Neural Network system, Expert Systems, Classification, this paper provides a review of different Artificial System and expert system method in diagnosis and detections of liver disease disorders acuteness is the key for results.

# INTRODUCTION

The use of intelligence systems in medical diagnosis is increasing gradually. There is no doubt that evaluation of data taken from patient and decisions of experts are the most important factors in diagnosis. But, expert systems and different artificial intelligence techniques for classification also help experts in a great deal.

Classification systems, helping possible errors that can be done because of fatigued or inexperienced expert to be minimized, provide medical data to be examined in shorter time and more detailed. Problems with liver patients are not easily discovered in an early stage as it will be functioning normally even when it is partially damaged. An early diagnosis of liver problems will increase patient’s survival rate. Liver disease can be diagnosed by analyzing the levels of enzymes in the blood.

Furthermore, nowadays mobile devices are widely used for supervision humans’ body conditions. In addition, automatic classification algorithms are demanded. According to their implements for liver diseases (almost definitely mobile capable or web capable), that decline the patient queue at the liver experts such as endocrinologists. Liver disorders are also an important disease in medicine. Levels of enzymes combined to blood are analyzed in Liver Disorders diagnosis. It can be a lot of possible errors in this diagnosis due to the number of enzymes to be many as well as the effects of different taken alcohol rates to be very from one patient to the other.

This paper is structured as follows. Section two briefly Data set of liver disorders, Section three reviews the various methods of liver disorders finally, Section four is a conclusion. The used data source in this study is UCI machine learning repository.

The Liver Disorders data is named as BUPA Liver Disorders. BUPA Liver Disorders database make ready by BUPA Medical Data Company includes 345 specimens consisting of six fields and two classes. Each sample is taken from an single man. Two hundred of these samples are of one class with remaining 145 are possessed by to the other. First Five attributes of the collected data samples are the outcomes of blood test while the last attribute includes daily alcohol consuming.

The attributes are as follows:

1. Total\_Bilirubin,
2. Alkaline phosphotase,
3. Alamine\_Aminotransferase,
4. Aspartate\_Aminotransferase,
5. Total\_Protiens
6. Albumin
7. Albumin\_and\_Globulin\_Ratio

# LITERATURE SURVEY

Classification algorithm is one of the greatest significant and applicable data mining technique used to apply in disease prediction. Classification algorithm is the most common in several automatic medical health diagnoses. Many of them show a good classification accuracy listed below,

Paul Mangiameli et al., [2] proposed model selection affects the decision support systems accurately. In their model selection, how to affects the accuracy of decision support system hydrides by single model and ensembles. They proposed single model is not more accurate than ensembles.

Ahmed M. Hashem et al., [18] proposed to predict Liver Cirrhosis or fibrosis single stage classification model and multistage classification model. In their model based on Decision Tree, Neural Network, Nearest Neighborhood clustering and Logistic Regression.

Ziol.M etal.,[3] proposed to evaluated liver fibrosis with chronic hepatitis C for patients using liver stiffness measurement (LSM).

Z. Jiang.Z.,[4] proposed for discovering the corresponding degree of fibrosis by support vector machine (SVM).

Kemal Polat et al.,[22] proposed resource allocation mechanism of AIRS was changed with a new one decided by Fuzzy-Logic. This approach called as Fuzzy- AIRS was used as a classifier in the diagnosis of Liver Disorders.

In this Classification accuracies were evaluated by comparing them with reported classifier’s accuracy, time and number of resources.

Piscaglia et al.,[6] proposed to predict Liver cirrhosis and other liver-related diseases used by Artificial neural network.

Dong-Hoi Kim et al.[19] proposed machine learning technique and decision tree(C4.5).In this method is used for to predict the susceptibility to two liver diseases such as chronic hepatitis and cirrhosis from single nucleotide polymorphism(SNP) data . They also used to identify a set of SNPs relevant to those diseases.

Anh Pham,[8] developed optimizing the classification accuracy when analyzing some medical datasets. This proposed work done by new meta-heuristic approach, called the HomogeneityBased Algorithm (or HBA).This approach used to predict error rates and associated penalty costs. These costs may be dramatically different in medical applications as the implications of having a false-positive and a false-negative case may be tremendously

Rong-Ho Lin [9] proposed to predict accuracy of liver disease using case-based reasoning

(CBR) and classification and regression tree (CART) approach. He also integrates CART and CBR for the diagnosis of liver diseases. In this model included two major steps. (1) CART To diagnose whether a patient suffers from liver disease using CART. (2)To predict which types of Liver disease affected for patients using CBR.He also [18], proposed to determine whether patients suffer from liver disease or not using case-based reasoning, artificial neural networks and analytic hierarchy methods . They also predict which types of liver disease suffered human body. Chun-Ling Chuang et al.,[15] proposed to diagnosis early Liver disease and predict classification accuracy by integrated case- based reasoning into classification and regression tree, back-propagation neural network (BPN), discriminatory analysis and logistic regression of classification methods in data mining techniques. In their methods used a ten-fold crossvalidation to select a best.

# Problem Definition:

# The present system shares the same objective but encompass different methodologies to arrive at a relatively less accurate conclusion. The qualitative superiority that these methods have over one another is dependent on the accuracy of the results produced. There are different aspects of the data that are used in order to parametrically come to a definite conclusion over the prediction of liver disease. Fuzzy logic has been developed for the classification of patients with liver cirrhosis. In gastroenterology, the Child-Pugh score is used to assess the prognosis of chronic liver disease, mainly cirrhosis. It was originally made to predict the mortality during surgery. It is now used to determine the prognosis, the required strength of treatment and the necessity of liver transplantation. The following score is instrumental in utilizing the five clinical measures of liver disease and each of these measures are scored between 1 and 3, with 3 indicating a serious condition of organ deterioration.

• Risk factors can be predicted early by machine learning models: The machine learning algorithms predict the risk factors through simple methodologies of analysis the inconsistencies in the collective training data set and their respective parameters.

The proposed system is built expert system which employs for the diagnosis of LDs is developed in an environment Neuroph and Crystal's reports were used for analysis. Approach for analyzing clusters to identify a meaningful pattern for determining whether a patient suffers from LD or not is presented.

The system provides a guide for the diagnosis of LDs within the decision-making framework. The process for the medical diagnosis of LD starts when an individual consults a physician (doctor) and presents a set of complaints (symptoms). The physician then requests further information that will further aid in the proper diagnosis of the disease.

The doctor is likely to start with a health history and a thorough physical examination.

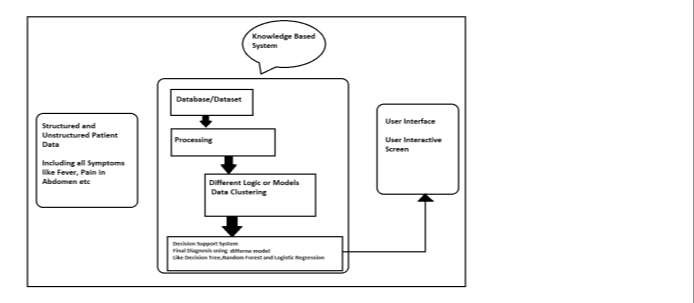
• Blood tests. A group of blood tests called liver function tests can be used to diagnose liver disease. Other blood tests can be done to look for specific liver problems or genetic conditions. • Imaging tests. An ultrasound, CT scan and MRI can show liver damage.

• Tissue analysis. Removing a tissue sample (biopsy) from your liver may help diagnose liver disease and look for signs of liver damage. A liver biopsy is most often done using a long needle inserted through the skin to extract a tissue sample. It is then analyzed in a laboratory.

**Strategy**

This seems to be a classic example of supervised learning. We have been provided with a fixed number of features for each data point, and our aim will be to train a variety of Supervised Learning algorithms on this data, so that , when a new data point arises, our best performing classifier can be used to categorize the data point as a positive example or negative. Exact details of the number and types of algorithms used for training is included in the Algorithms and Techniques' sub-section of the 'Analysis' part.

# PROPOSED SYSTEM



The proposed system is build expert system which employ for the diagnosis of LDs is developed in an environment characterized by Microsoft Window XP professional to Windsows 10 Operating System, Microsoft Access Database Management system, Python Language and Microsoft Excel. Neuroph and Crystal reports were used for analysis and graphical representation. An approach for analyzing clusters to identify meaningful pattern for determining whether a patient suffers from LD or not is presented. The system provides a guide for diagnosis of LDs within the decision making framework. The process for the medical diagnosis of LD starts when an individual consults a physician (doctor) and presents a set of complaints (symptoms). The physician then requests further information that will further aid in the proper diagnosis of the disease

Data collected include patient’s previous state of health, living condition and other medical conditions. When diagnosing LD, the physician looks at the patient’s symptoms and conducts a physical examination. In addition, the physician may request a liver biopsy, liver function tests, an ultrasound, CT scan, and/or a magnetic resonance imaging (MRI) scan From the symptoms presented by the patient, the physician narrows down the possibilities of the illness that corresponds to the apparent symptoms and make a list of the conditions that could account for what is wrong with the patient. These are usually ranked in possibility order . When the list has been narrowed down to a single condition, it is called differential diagnosis and provides the basis for a hypothesis of what is ailing the patient. The examining physician accounts for possibilities of having LD through an interview, physical examination, LFT tests or liver biopsy. LFT test is used for identifying the presence of certain liver enzymes in the blood. A thorough diagnostic evaluation may include a complete history of; when the symptoms started, how long the symptoms have lasted, how severe it has been and having occurred before, was it treated and what treatment was received

**Advantages of the Proposed System:-**

Considering the certain differences that have been adopted in the current system the following are the distinct advantages that are observed: • The performance classification of liver-based diseases is further improved: with the far deepened understanding of the different kinds of ailments in the field of medicine, the different set of parameters to distinctly determine the kind of liver disease and its occurrence has become a far less complicated task. With advancements in data mining paradigms and software architectures like Hive, R, easing up the data collection process, the preprocessing and evaluation stages are given more attention to. • Time complexity and accuracy can be measured by various machine learning models, so that we can measures different parameters, owing to the needs of the user: Every prediction system is based on the kind of parameters that it is expected to accept, compare and then finally come to a predictive conclusion. Accordingly, there are different algorithms that are used to model the predictive system to suit the context. The different machine learning algorithms judge the kind of disease and the testing parameters. • Different machine learning having high accuracy of the result: In comparison to other methodologies considered, the right machine learning algorithm can aptly increase the efficiency of the results that are expected out of the predictive system.

**Scope of the Project:**

In this thesis the proposed system concludes that PSO feature selection methods for Indian Liver Patient Dataset. This thesis analyzed the liver disease using algorithms such as J48, MLP, SVM, Random Forest, and Bayesnet Classification. These algorithm gives various result based on PSO feature selection model .It has been seen that bayes net and J48 Classification gives better results compare to other classification algorithms.

There are many criterions for evaluating the selected feature subset, here this thesis used features such as Total bilirubin, Direct\_ bilirubin, Total\_protiens, Albumin, A/G ratio, SGPT, SGOT, Alkphos to evaluate the performance of different classification algorithm. In future, we have attempted to classify different feature selection algorithms into four groups: complete search, heuristic search, meta-heuristic methods and methods that use artificial neural network.

**System Requirement:**

**Hardware Requirement:**

RAM: 4GB.

System type: 64-bit OS.

Processor: Intel I3-6000U or AMD CPU.

Hard Disk: 125 GB Minimum.

**System Requirement:**

Operating System: Windows XP-10, Kali Linux, Mac OS.

Software : Visual Studio, Jupyter (IDE),etc.

Languages: Python(MarLab, Pandas, Numpy, Flask) .

User Interface: HTML, CSS, & JS.

**Methodology:**

**NORMALIZATION**

Normalization is the process of classify data into an associated table it also eliminates redundancy and increases the reliability which improves output of the query. To normalize a database, we divide the dataset into tables and establish relationships between the tables. Dataset normalization can essentially be defined as the practice of optimizing table structures. Optimization is accomplished as a result of a thorough investigation of the various pieces of

data that will be stored within the database, in particular concentrating upon how this data is interrelated.

**FEATURE SELECTION**

PSO feature extraction model for liver dataset and applied an improve probability in many medical application such as training artificial neural networks, linear constrained function optimization, wireless network optimization, data classification, and many other areas where GA can be applied. Computation in PSO is based on a swarm of processing elements called particles in which each particle represent a candidate solution.

**RANDOM FOREST (RF)**

Random forests is a machine learning regression method for classification that drive by constructing liver data into a multitude of decision trees at training time and outputting the class that is the mode of the classes output by individual trees [14]. It is unexcelled in accuracy among current algorithms. It output classification efficiently on large liver dataset. It can handle thousands of input attributers without variable deletion. It gives estimates of what variables are important in the classification. Random Forests grows many classification trees. To classify a new liver object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, and says the tree "votes" for that class. The forest chooses the classification having the most votes).

**SUPPORT VECTOR MACHINE (SVM)**

SVM have attracted a great deal of attention in the last decade and actively tested to various domains applications. SVMs are mostly used for learning classification, regression or ranking function. SVM are based on statistical learning theory and structural risk minimization principal and have the intent of determining the location of decision boundaries also known as hyperplane that produce the optimal separation of classes. SVM is the most robust and exact classification technique, there are many problems. The data analysis in SVM is based on convex quadratic programming, and it is computationally costly, as solving quadratic programming methods require large matrix operations as well as time consuming numerical computations.

**Expected Outcome:**

Here we are creating a web app which is based on the Machine Learning where a system will predict based on the data given by the user, the system will predict that the patient is having the liver diseases or not based on then data provided.

Where we are creating a Web pages if the patient is having the Liver diseases then a red colour page contain some information about the result will inform. Or a green colour page containing some information is pop up.

# CONCLUSIONS

In this paper the problem of summarizing the different algorithm of data mining is used in the field of medical prediction are discussed. The main focus is on using different algorithm and combination of several targets attributes for different types of disease prediction using data mining. A foremost class of problems in medical science absorbs the diagnosis of disease, based upon an assortment of tests carried out upon the patient. When several tests are involved, the ultimate diagnosis may be difficult to obtain, yet for a medical expert.

This has given rise, over the past few decades, to automated problem-solving tools, intended to assist the physician in making sense out of the welter of data. In healthcare, data mining is becoming increasingly more essential. The selection of data mining approaches depends on the nature of the dataset if the dataset consists of the labeled features then the classification techniques can be suggested for best prediction. If the dataset is with unlabeled features, then the clustering techniques are best suited for pattern recognition. If the optimization of the results needs to be improvised means, then bio inspirational based techniques are best suited. Keeping in consideration with these existing problems this paper aims to survey the existing approaches in the field of medical sciences and the importance of data mining techniques used by various authors. The study reveals the importance of life threading disease should be diagnosed.