



# A Novel Machine Learning Approach for Earlier Diagnosis of Liver Disease

D.Shanthi<sup>1</sup>, S.T.Shenbagavalli<sup>2</sup>, S.Santhana Prabha<sup>3</sup>, M.Jayanthi<sup>4</sup>, J.Biju<sup>5</sup>, T.Yogameera<sup>6</sup>

## Abstract

Healthcare system is becoming a significant factor of every human being. So in this paper we are discussing about the vital organ of the human body i.e. Liver. Early analysis of liver disease prediction is highly challenging. As a result, liver disease has become more common which is gradually improved, enhancing people's troubles. To recognize the causes and the identification phases are more important. For this, we applied a machine learning technique that was highly promising for studies with regard to healthcare and health. To uncover hidden data for precise diagnosis and decision-making, ML can also be used. Any form of liver disease is hazardous to your health. Poisons, narcotics, and excessive drinking are the main causes of this condition. In this study, we will use several associated features and KNN technology to enhance liver disease prediction. The accuracy forecast is provided by this algorithm. The main aspect is to predict the results more efficiently and reduce the cost of diagnosis in the medical sector. Therefore, we used different classification techniques for the classification of patients has liver disease.

2474

**KeyWords:** Machine learning, Hypertext Mark-up Language, Cascade Style Sheet, Support Vector Machine

**DOI Number:** 10.48047/NQ.2022.20.16.NQ880249

**NeuroQuantology2022; 20(16):2474-2479**

## Introduction

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. Machine learning algorithms are often categorized as supervised or unsupervised. Supervised machine learning algorithms can apply what has been learned in the past to new data using labeled examples to predict future events. Starting from the analysis of a known training dataset, the learning algorithm produces an

inferred function to make predictions about the output values. The system is able to provide targets for any new input after sufficient training. The learning algorithm can also compare its output with the correct, intended output and find errors in order to modify the model accordingly.

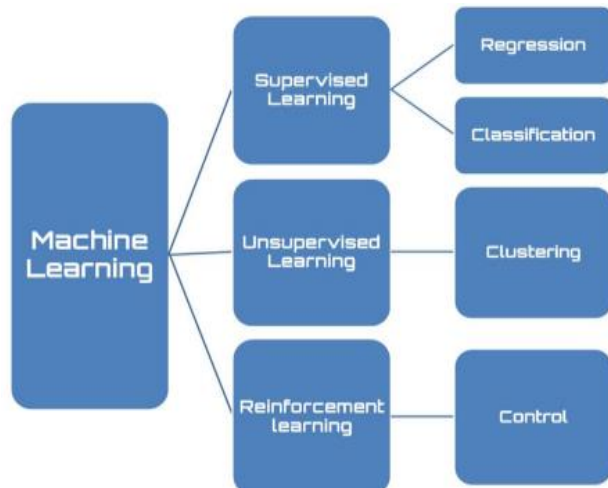
In contrast, unsupervised machine learning algorithms are used when the information used to train is neither classified nor labeled. Unsupervised learning studies how systems can infer a function to describe a hidden structure from unlabeled data. The system doesn't figure out the right output, but it explores the data and can draw inferences from datasets to describe hidden structures from unlabeled data. [1]

**Corresponding author:** D.Shanthi

**Address:** <sup>1</sup>Professor and Head, Department of Computer Science and Engineering, PSNA College of Engineering & Technology, Dindigul, India, <sup>2,3,4</sup>Assistant Professor, Department of Computer Science and Engineering, PSNA College of Engineering & Technology, Dindigul, India, <sup>5</sup>Assistant Professor, Department of Artificial intelligence & Data science, KGISL Institute of technology, Coimbatore, India, <sup>6</sup> Assistant Professor, Department of computer science and Engineering, Theni kammavarsangam, College of technology, India

E-mail: dshan71@gmail.com, stshenbagavalli@gmail.com, ssanthanaprabhasivakumar@gmail.com, jayantheemurugesan@gmail.com, jbijuinfo@gmail.com, yogameeraarasu@gmail.com





**Fig.1: Machine Learning**

Liver disease (also called hepatic disease) is a type of damage to or disease of the liver. The liver is the largest organ in the body. Among its 500+ roles, the liver is responsible for food processing, energy storage, blood filtration, and immune response. Specifically, the liver contributes by secreting bile for lipid breakdown, storing excess glucose as glycogen, and removing bacteria and toxins from the blood. Whenever the course of the problem lasts long, chronic liver disease ensues. Therefore, this makes the way to develop a system that helps in prediction of liver disease.

Proposed model should able to get the data from user and compare that with existing model and

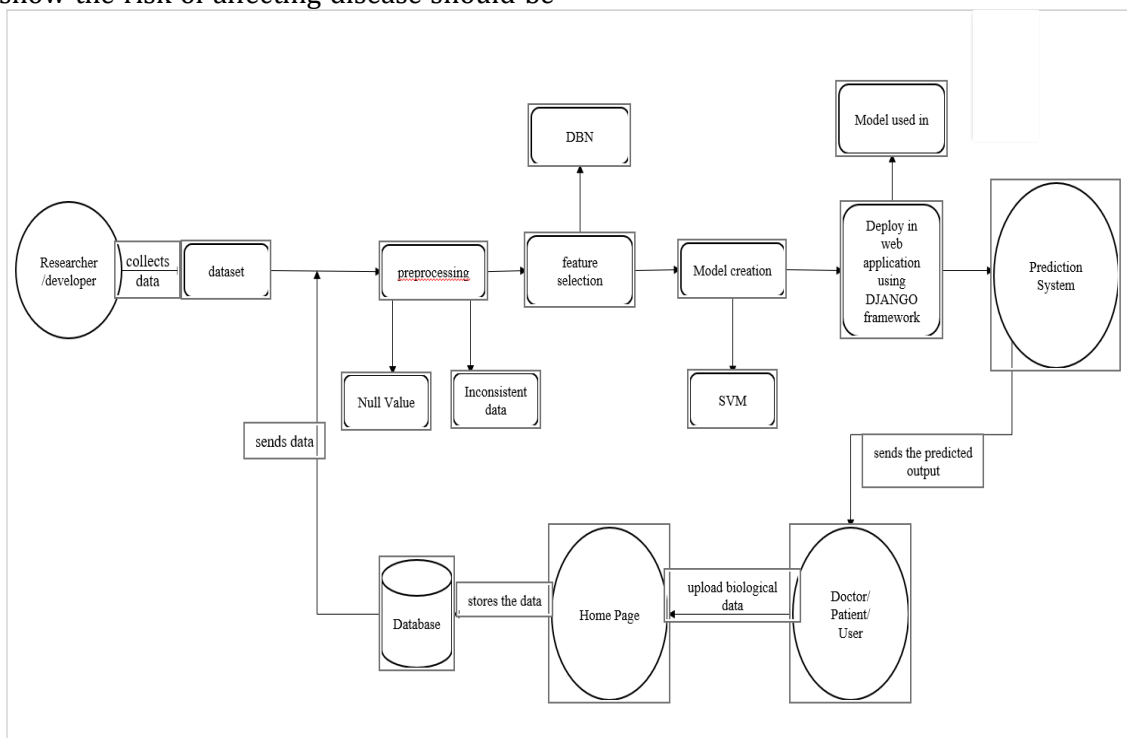
shown. With help of the prediction, the patient should be able to advance their medication in order to avoid getting disease at severe level.

2475

**Methodology**

The proposed system will predict the risk factor of the HCC affected patients using Machine Learning. We are using DBN (Deep Belief Network) for feature selection. SVM (Support Vector Machine) is used to classify the processed datasets. Pre-processing and correct replacement of missing value with different attribute type will improve the model. [2]

should show the risk of affecting disease should be



**Fig.2: System Architecture**

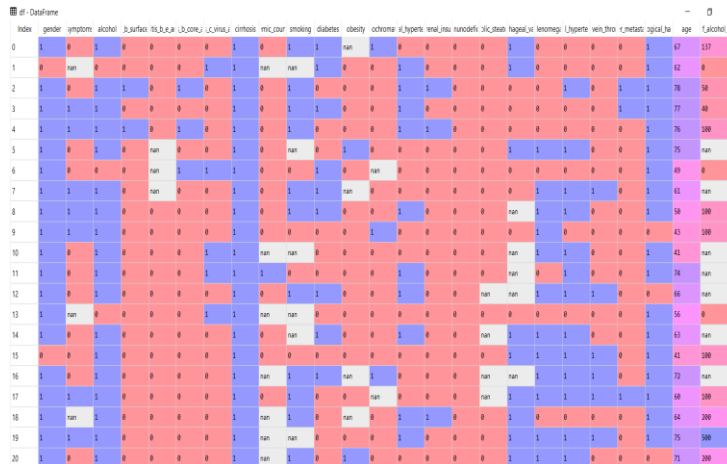
**Modeling And Analysis**

**Data Collection**

To achieve the goal, Data Engineering is the first step. Data Engineering consists of two processes, they are Data Collection and Data Pre-processing. Data Collection will be collected with meaningful parameters like age, blood test and so on. [3]

**Data Pre-Processing**

Collected data will be Pre-processing which means encoding the categorical information in the data. Dropping unwanted parameters, scaling the parameter values to achieve normal distribution (Zero mean and Standard Deviation as one), handling missing values and so on. Here Data set from Kaggle where biological data of 165 people are collected [4]

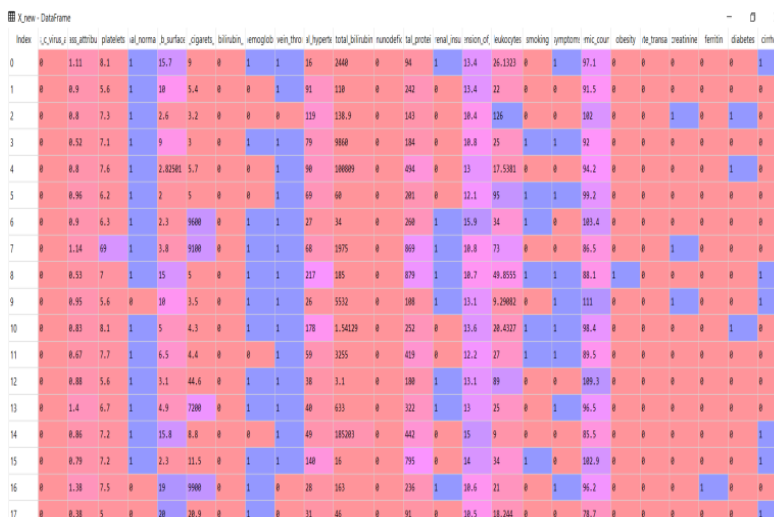


**Fig.3: Data Collection Pre-Processing**

**Feature Engineering**

After the Data Engineering process, Feature Engineering will be done. Feature Engineering is an important step to predict our output. The advantage of Feature Engineering is minimizing the parameter. For example, if our whole datasets contains 10

parameters, after feature engineering only three-parameter enough to predict the output with high efficiency. Feature Engineering based on correlation, co-variance, co-linearity and etc. Feature Engineering has many algorithms to predict correct correlated parameters. [5-6]



**Fig.4: Proposed Process Parameters**





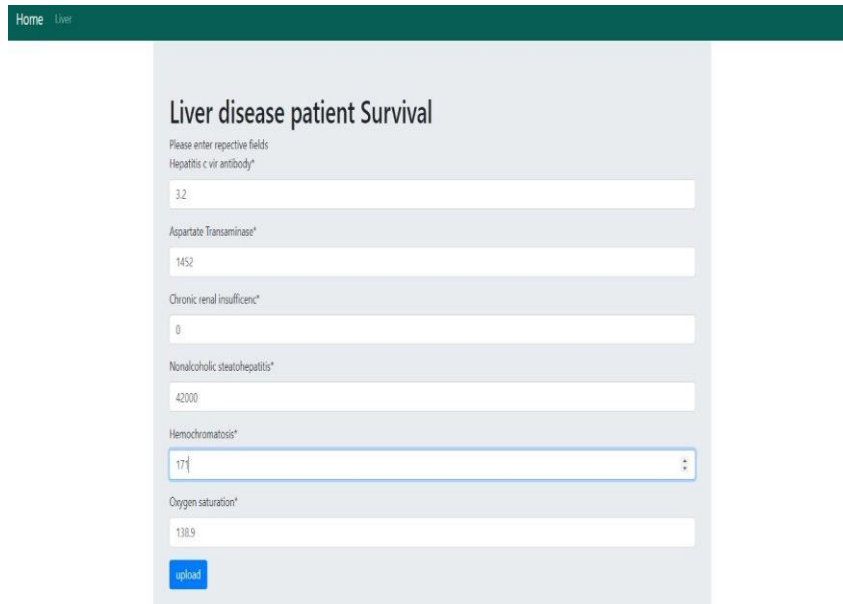


Fig.7: Web Development Images

2478

### Result And Analysis

The first step is the extraction of data. Data set from Kaggle where biological data of 165 people are collected. Since the data is collected on real time, data Pre-processing is needed. In Pre-processing us eliminating null values and inconsistent data and we will get datasets which will used be for processing.

Using DBN algorithm, we are getting best features for processing .By using this feature we can able to build best model. After getting datasets with best feature, we want the build suitable model which will give better accuracy. For our use case, we are using SVM algorithm. Once the model is created, deploy it in web server. So that user can predict the presence of disease with low risk or high risk.



Fig.8: System Output

### Conclusion

The HCC affected person's risk factor was classified with Support Vector Machine. This was achieved with feature selection method select-K parameter with chi square. The effective five features were selected from 50 features using feature selection

method. The result achieved was 95% accuracy. The trained model with SVM for 5 feature input are able to predict the low risk or high risk. Advantage of using feature selection has eliminated the unwanted feature which may increase the blood test cost of the person. In the proposed work, different classifiers were



implemented on liver patient diseases datasets to predict liver diseases based on developed software. Datasets was processed and implemented using feature selection techniques. The results of the proposed work were compared using feature selection and without using feature selection techniques after the implementation of different classifiers in terms of execution time and accuracy. The best result was achieved using Logistic Regression classifier with feature selection techniques and execution time of different classifiers was decreased after the implementation of feature selection technique. Finally, liver disease prediction Software (LDPS) is developed using concept of software engineering life cycle.

## References

- Performance Assessment of Classification Algorithms on Early Detection of Liver Syndrome Rashid Naseem, Bilal Khan, M. A. Shah, Karzan Wakil, Atif Khan, Wael Alosaimi, M. I. Uddin, BadarAlouffi Published on 2020 in "National Library of medicine, USA".
- Diagnosis of Liver Disease using Machine Learning Models A. Sivasangari, Baddigam Jaya Krishna Reddy, AnnamareddyKiran,P.Ajitha Published 2020 on "IEEE xplore".
- A.N.Arbainand, B.Y.P.Balakrishnan, "A comparison of data mining algorithms for liver disease prediction on imbalanced data, published in" International Journal of Data Science and Analytics, vol. 1, on 2019.
- Jagdeep Singha, SachinBaggab, Ranjodh Kaur Software-based Prediction of Liver Disease with Feature Selection and Classification Techniques International Conference on Computational Intelligence and Data Science (ICCIDS 2019).
- Binish Khan Piyush Kumar Shukla Manish Kumar Ahirwar Strategic Analysis in Prediction of Liver Disease Using Different Classification Algorithms INTERNATIONAL JOURNAL OF COMPUTER SCIENCES AND ENGINEERING · July 2019.
- A Comparison of Data Mining Algorithms for Liver Disease Prediction on Imbalanced Data "Ain Najwa Arba, B. P. Balakrishnan Published 2019 published in International Journal of Data Science and Advanced Analytics (ISSN 2563-4429).
- T.Choudhury, and A.Thakral. (2019), "Liver Disease Detection Due to Excessive Alcoholism Using Data Mining Techniques" published in IEEE International Conference on Advances in Computing and Communication Engineering (ICACCE), pp. 163-168.
- I.Arshad, C. Dutta, T. Choudhury, and A. Thakral. (2019), "Liver Disease Detection Due to Excessive Alcoholism Using Data Mining Techniques." Published in IEEE xplore.

