

# Analysis of Liver Disease Prediction Using Machine Learning

SYED SHAZIA<sup>1</sup>, B. MURALI<sup>2</sup>

<sup>1</sup>PG Student, CSE, Quba College of Engineering & Technology

<sup>2</sup>Associate professor, CSE, Quba College of Engineering & Technology

**Abstract**— *Liver disease is becoming one of the most fatal diseases in several countries. Patients with the Liver disease have been continuously increasing because of excessive consumption of alcohol, inhaling of harmful gases, intake of contaminated food, pickles, and drugs. These days, AI strategies have generally been utilized in clinical science for guaranteeing precision. In this work, we have precisely built computational model structure procedures for liver infection forecast. We utilized some effective characterization calculations (Decision Tree, ANN and Support Vector Machine) for chronic liver disease patients which lasts over six months. We proposed an investigation model to predict liver infection with a high exactness value. Then, we analysed the good and bad values using a machine learning classifier which improvises the classification resultant. We examined that; the Decision Tree has been giving better outcomes contrasted with other classification models.*

**Index Terms**— *Liver Disease, Chronic Liver Disease (CLD), Artificial Intelligence (AI) in Healthcare, Machine Learning (ML) in Medicine, Liver Disease Prediction, Computational Modeling, Classification Algorithms, Decision Tree, Artificial Neural Network (ANN), Support Vector Machine (SVM)*

## I. INTRODUCTION

In India Cirrhosis of the liver is a big health issue. According to the latest WHO data, Liver disease mortality in India reached 259,749 in 2017, or 2.95 percent of total deaths, accounting for 18.3 percent of all cirrhosis deaths worldwide [1]. The intricacy of the symptoms in the early stages of liver disease makes it difficult to identify. Because the liver continues to operate even when partially damaged, problems with liver illnesses are frequently not recognized until it is too late. An early diagnosis may be able to save a person's life.

It is possible to identify the early signs of certain diseases, even if they are not visible to a doctor's trained eye. Patients who are diagnosed early have a far greater chance of living a long life. Liver failure is associated with a high risk in Indians. India is expected

to become the world's capital of liver disease by 2025 [2]. The epidemic of liver infections in India is caused by a destined lifestyle, increased alcohol consumption, and smoking. There are about 100 types of liver infections.

The human liver is a fascinating internal organ that can perform about 500 different functions. Its main functions are boosting immunity, detoxifying and purifying, and producing proteins and hormones. It maintains blood sugar levels and prevents blood loss due to coagulopathy. In more complex metabolic activities, the liver is known as the internal organ and the liver can regenerate and repair its tissues. Failure of such organs can cause serious health problems[3]. According to a survey of the main causes of medical accidents in India, liver disease is on the top10 list of diseases, and worldwide, India was ranked 63rd in liver disease in 2017. Liver diagnosis is made by both imaging and liver function tests. The liver is the biggest organ of the frame and it smiles are important for digesting meals and freeing the poison .

## II. LITERATURE SURVEY

Machine Learning in Healthcare: A Comprehensive Review of Chronic Disease Prediction. AUTHOR:Smith,A., &Patel, S DESCRIPTION: This comprehensive review explores the application of machine learning in healthcare, specifically focusing on the prediction of chronic liver diseases. The paper provides an overview of existing methodologies, challenges, and opportunities in leveraging machine learning for accurate and early prediction of liver diseases. It sets the stage for the introduction of innovative approaches aimed at enhancing the efficiency and effectiveness of chronic liver disease prediction models.

Feature Selection Techniques for Optimized Chronic Liver Disease Prediction AUTHOR:Wang,Q.,&Kim,J

DESCRIPTION: Focusing on feature selection, this paper presents a detailed analysis of methodologies for optimizing chronic liver disease prediction models using machine learning. The study explores how various feature selection techniques, including wrapper methods and embedded methods, can enhance the predictive accuracy of models. Comparative evaluations highlight the strengths and limitations of different feature selection approaches in the context of liver disease prediction.

### III. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM: Machine Learning has pulled in an immense measure of investigation and has been applied in different fields. In medication, ML has demonstrated its force where it has been utilized to take care of numerous crisis issues like malignant growth treatment, coronary illness, dengue fever analysis, etc. A few remarkable strategies such as Logistic regression, Random Forest, and other data mining algorithms to calculations have been utilized for liver examination

#### 3.1.1 DISADVANTAGES OF EXISTING SYSTEM :

1. Accuracy is Low.
2. Only using Data Mining algorithms and ML algorithms in existing methods

3.2 PROPOSED SYSTEM: In this work, we have precisely built computational model structure procedures for liver infection forecast. We utilized some effective characterization calculations (Decision Tree , Decision Tree has been giving better outcomes contrasted with other classification models

#### 3.3. SYSTEM REQUIREMENTS

Functional requirements for a secure cloud storage service are straightforward:

Software requirements:

- Operating System: Windows
- Coding Language: Python 3.7
- Script:
- Database :

Hardware Requirements:

- Processor - Pentium –III
- Speed – 2.4 GHz
- RAM - 512 MB (min)
- Hard Disk - 20 GB
- Floppy Drive - 1.44 MB

- Key Board - Standard Keyboard
- Monitor – 15 VGA Colour

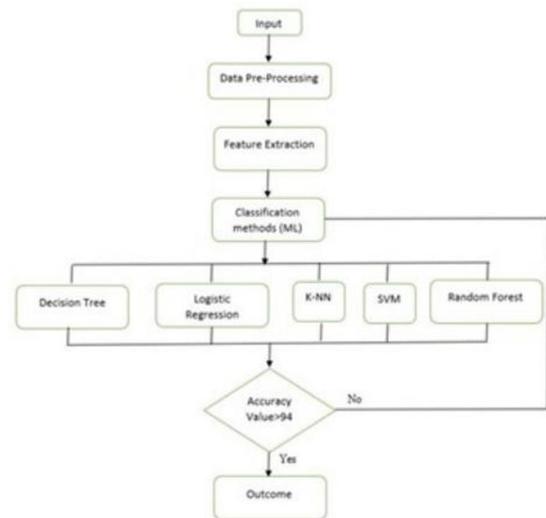
Module Description: Tensor flow: TensorFlow is a free and open-source differentiable programming software library for dataflow and across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks . It is used for both research and production at Google . TensorFlow was developed by the Google Brain was released under the Apache 2.0 team for internal Google use. It open-source license Numpy: on November 9, 2015.

Numpy is a general-purpose array-processing package. It provides a high performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contain various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting)
- functions Tools for integrating C/C++ and Fortran code

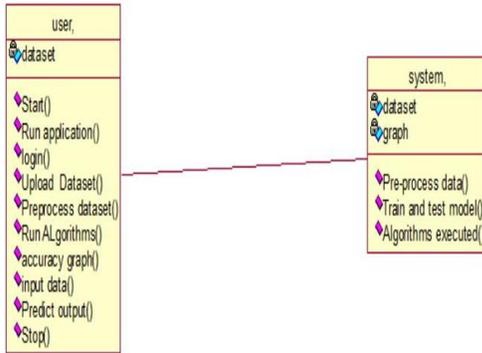
### IV. SYSTEM ARCHITECTURE



### V. SYSTEM DESIGN

5.1 CLASS DIAGRAM: In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that

describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information



5.2 UMLDIAGRAMS: UML is an acronym that stands for Unified Modeling Language. Simply put, UML is a modern approach to modeling and documenting software. In fact, it's one of the most popular business process modeling techniques . It is based on diagrammatic representations of software components. As the old proverb says: "a picture is worth a thousand words". By using visual representations, we are able to better understand possible flaws or errors in software or business processes. UML was created as a result of the chaos revolving around software development and documentation. In the 1990s, there were several different ways to represent and document software systems. The need arose for a more unified way to visually represent those systems and as a result, in 1994-1996, the UML was developed by three software engineers working at Rational Software . It was later adopted as the standard in 1997 and has remained the standard ever since, receiving only a few updates.

## VI. SOFTWARE ENVIRONMENT

What is Python?

Below are some facts about Python.

Python is currently the most widely used multi-purpose, high-level programming language.

Python allows programming in Object-Oriented and Procedural paradigms. Python programs generally are smaller than other programming languages like Java.

Programmers have to type relatively less and indentation requirement of the language, makes them readable all the time.

Python language is being used by almost all tech-giant companies like – Google, Amazon, Facebook, Instagram, Dropbox, Uber... etc.

The biggest strength of Python is huge collection of standard library which can be used for the following

- Machine Learning
- GUI Applications (like Kivy, Tkinter, PyQt etc. )
- Web frameworks like Django (used by YouTube, Instagram, Dropbox)
- Image processing (like OpenCV, Pillow)
- Web scraping (like Scrapy, BeautifulSoup, Selenium)
- Test frameworks
- Multimedia

Advantages of Python :-

Let's see how Python dominates over other languages.

Extensive Libraries

Python downloads with an extensive library and it contain code for various purposes like regular expressions, documentation-generation, unit-testing, web browsers, threading, databases, CGI, email, image manipulation, and more. So, we don't have to write the complete code for that manually.

Extensible

As we have seen earlier, Python can be extended to other languages. You can write some of your code in languages like C++ or C. This comes in handy, especially in projects.

Embeddable

Complimentary to extensibility, Python is embeddable as well. You can put your Python code in your source code of a different language, like C++. This lets us add scripting capabilities to our code in the other language.

Improved Productivity

The language's simplicity and extensive libraries render programmers more productive than languages like Java and C++ do. Also, the fact that you need to write less and get more things done.

## VII. SYSTEM IMPLEMENTATION

Sample code:

```

from django.shortcuts import render
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
import joblib

from keras.models import Sequential

from keras.layers.core import Dense, Activation, Dropout
from keras.optimizers import Adam

import matplotlib.pyplot as plt
import numpy as np
import os

from os import path

from keras.models import load_model

# Create your views here.

def adminlogin(request):
    return render(request, "AdminApp/Login.html")

def logaction(request):
    name = request.POST.get('username')
    apass = request.POST.get('password')
    if name == 'Admin' and apass == 'Admin':
        return render(request, "AdminApp/AdminHome.html")
    else:
        context = {'data': "Admin Login Failed..!!"}
        return render(request, "AdminApp/Login.html", context)

def AdminHome(request):
    return render(request, "AdminApp/AdminHome.html")

global data
def UploadDataset(request):
    global data
    filename = "dataset\\Liver_Patient_Dataset.csv"
    data = pd.read_csv(filename, encoding='unicode_escape')

```

```

context = {'data':data}
return render(request, "AdminApp/Upload.html", context)

global X, y, X_train, X_test, y_train, y_test
def Preprocess(request):
    global X, y, X_train, X_test, y_train, y_test
    filename = "dataset\\Liver_Patient_Dataset.csv"
    data = pd.read_csv(filename, encoding='unicode_escape')
    data.dropna(inplace=True)
    data['Gender of the patient']=data['Gender of the patient'].map({'Female':0,'Male':1})
    X=data.iloc[:, 0:10]
    y =data.iloc[:, 10:11]
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)
    context={'data':str(len(data)), 'train':str(len(X_train)), 'test':str(len(y_test))}
    return render(request, "AdminApp/Preprocess.html", context)

global decacc, decmodel
def runDT(request):
    global decacc, decmodel
    svmmodel = DecisionTreeClassifier(random_state=0)
    svmmodel.fit(X_train, y_train)
    joblib.dump(svmmodel, "model/DecModel.joblib")
    pred=svmmodel.predict(X_test)s

```

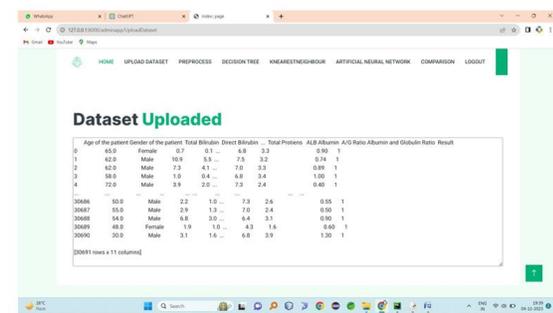
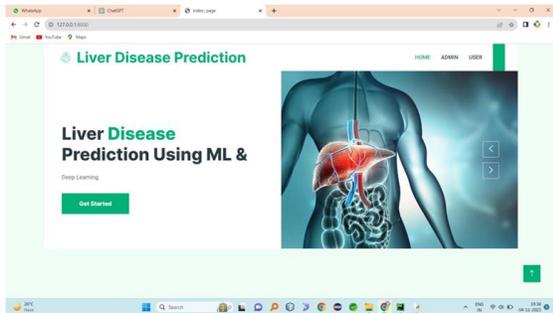
## VIII. SYSTEM TESTING

**SYSTEM TESTING:** The purpose of testing is to discover errors. Testing is the process of trying to discover every conceivable fault or weakness in a work product. It provides a way to check the functionality of components, sub-assemblies, assemblies and/or a finished product. It is the process of exercising software with the intent of ensuring that the Software system meets its requirements and user expectations and does not fail in an unacceptable manner. There are various types of tests. Each test type addresses a specific testing requirement.

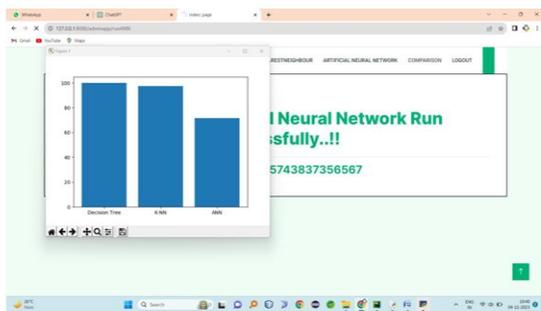
**TYPES OF TESTS:** Unit testing: Unit testing involves the design of test cases that validate that the internal program logic is functioning properly, and that program inputs produce valid outputs. All decision branches and internal code flow should be validated. It is the testing of individual software units of the application. It is done after the completion of an

individual unit before integration. This is a structural testing, that relies on knowledge of its construction and is invasive. Unit tests perform basic tests at component level and test a specific business process, application, and/or system configuration. Unit tests ensure that each unique path of a business process performs accurately to the documented specifications and contains clearly defined inputs and expected results

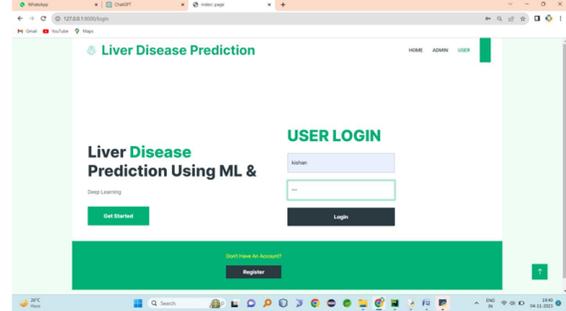
IX SCREENSHOTS



Comparison graph:



USER LOGIN SCREEN:



CONCLUSION

Diseases of the liver and heart become more common over time. With continuous technological advances, these will increase in the future. Today, people are becoming more health-conscious and are taking yoga and dance classes. Still, the sedentary lifestyle and luxury are constantly being introduced and improved. The problem will last for a long time. So, in such a scenario, our project is very useful to society. The dataset used in this project gave 99% accuracy in the Decision tree model. While it may be difficult to achieve such accuracy with such large datasets, the conclusion of this project is clear that liver risk can be predicted. In the future, philosophy is utilized to examine the liver area into distinct compartments for better classification accuracy. However, the technique requires further improvement generally to include the excretion of the liver into various parts: renal cortex, renal segment, renal medulla, and renal pelvis

REFERENCES

- [1] Mishra, Debakanta, Kaibalya R. Dash, Chittaranjan Khatua, Subhendu Panigrahi, Prasanta K. Parida, Sambit K. Behera, Rakesh K. Barik et al. "A study on the temporal trends in the etiology of cirrhosis of the liver in coastal eastern Odisha." Euroasian Journal of Hepato Gastroenterology 10, no. 1 (2020):
- [2] Sontakke, Sumedh, Jay Lohokare, and Reshul Dani. "Diagnosis of liver diseases using machine learning." In 2017 International Conference on Emerging Trends & Innovation in ICT (ICEI), pp. 129-133. IEEE, 2017.
- [3] Priya, M. Banu, P. Laura Juliet, and P. R. Tamilselvi. "Performance analysis of liver disease prediction using machine learning algorithms." International Research Journal of

Engineering and Technology (IRJET) 5, no. 1  
(2018): 206-211.

- [4] <https://bmcmedicine.biomedcentral.com/articles>
- [5] Garg, Arunim, and Vijay Mago. "Role of machine learning in medical research: A survey." *Computer Science Review* 40 (2021): 100370.
- [6] Rezai, Bahram, and Ebrahim Allahkarami. "Application of Neural Networks in Wastewater Degradation Process for the Prediction of Removal Efficiency of Pollutants." In *Soft Computing Techniques in Solid Waste and Wastewater Management*, pp. 75-93. Elsevier, 2021