A REVIEW OF LIVER PATIENT ANALYSIS METHODS USING MACHINE LEARNING

1. INTRODUCTION

Liver is the largest internal organ in the human body,it is essential for digesting food and releasing the toxic element of the body and plays a major role in metabolism and serving several vital functions. The liver is the largest glandular organ of the body. It weighs about 3 lb (1.36 kg) .The liver's main job is to strain the blood coming from the digestive tract, before passing it to the rest of the body. The liver also detoxifies chemicals and metabolizes drugs. As it does so, the liver hides bile that ends up back in the intestines. The liver also makes proteins important for blood clotting and other functions.The liver supports almost every organ in the body and is vital for our survival. Liver disease may not cause any symptoms at earlier stage or the symptoms may be vague, like weakness and loss of energy. Symptoms partly depend on the type and the extent of liver disease. Liver diseases are diagnosed based on the liver functional test.

Several diseases states can disturb the liver. Some of the diseases are Wilson's disease, hepatitis (an inflammation of the liver), liver cancer, and cirrhosis (a chronic inflammation that progresses ultimately to organ failure). Alcohol alters the metabolism of the liver, which can have on the whole detrimental effects if alcohol is taken over long periods oftime. Hemochromatosis can cause liver problems .

Common Liver Disorder

- **Fatty liver** is a revocable condition where large vacuoles of triglyceride fat acquire in liver cells via the process of limit. It can occur in people with a high level of alcohol consumption as well as in people who never had alcohol.

- **Hepatitis**(usually caused by a virus spread by excess contamination or direct contact with infected body fluids).

- **Cirrhosis** of the liver is one of the most serious liver diseases. It is an action used to indicate all forms of diseases of the liver characterized by the significant loss of cells. The liver gradually contracts in size and becomes leathery and hard. The regenerative action continues under liver cirrhosis but the progressive loss of liver cells exceeds cell replacement.

- **Liver cancer** The risk of liver cancer is higher in those who have cirrhosis or who had valid types of viral hepatitis; but more often, the liver is the site of secondary (metastatic) cancers spread from other organs.
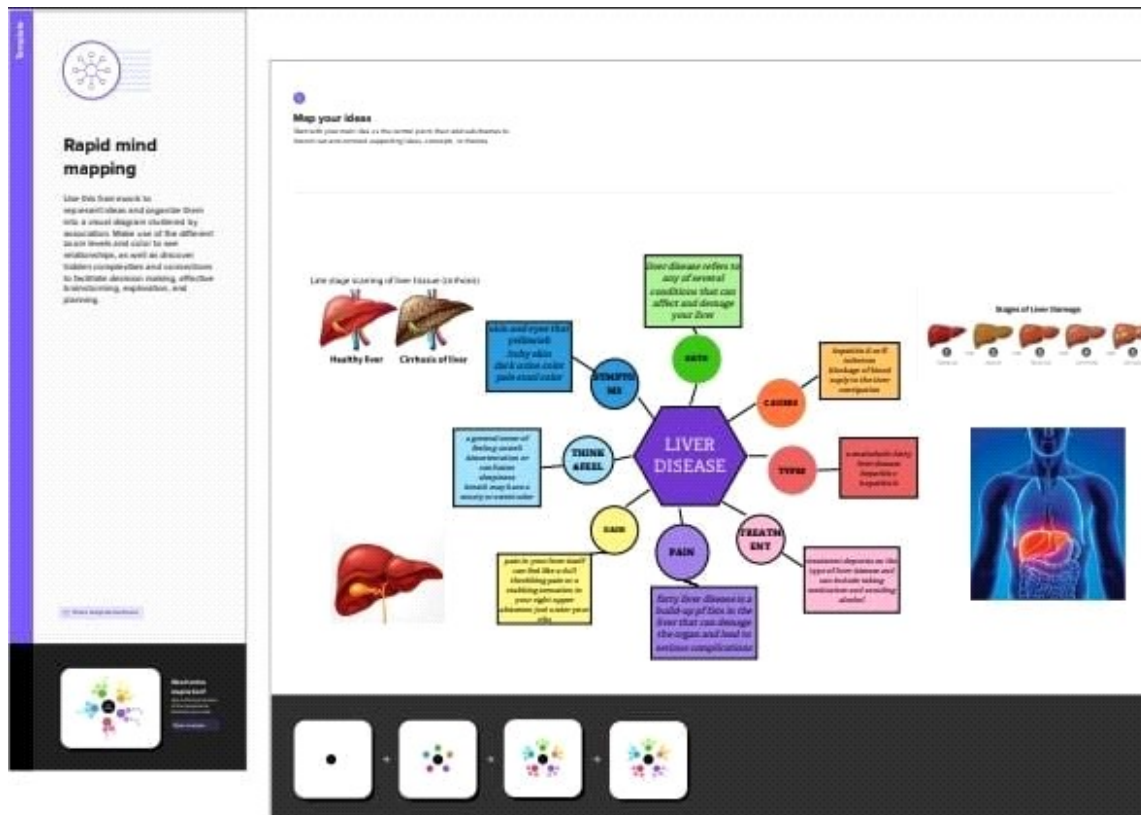
1.1 Overview

Liver diseases averts the normal function of the liver. This disease is caused by an assortment of elements that harm the liver. Diagnosis of liver infection at the preliminary stage is important for better treatment. In today's scenario devices like sensors are used for detection of infections. Accurate classification techniques are required for automatic identification of disease samples. This disease diagnosis is very costly and complicated. Therefore, the goal of this work is to evaluate the performance of different Machine Learning algorithms in order to reduce the high cost of liver disease diagnosis. Early prediction of liver disease using classification algorithms is an efficacious task that can help the doctors to diagnose the disease within a short duration of time. In this project we will analyse the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier for determining the liver disease. This project compares various classification algorithms such as Random Forest, Logistic Regression, KNN and ANN Algorithm with an aim to identify the best technique. Based on this study, Random Forest with the highest accuracy outperformed the other algorithms and can be further utilised in the prediction of liver disease and can be recommended to the user.

2.1 Purpose

The liver disease dataset which is select for this study is consisting of attributes like total bilirubin, direct bilirubin, age, gender, total proteins, albumin and globulin ratio. The main purpose of this work is to calculate the performance of various decision tree techniques and compare their performance.

2.Problem definition & Design Thinking

2.1 Empathy map

2.2 Ideation & Braintorming Map

2.RESULT

Liver Patient Analysis

Introduction

Liver diseases averts the normal function of the liver. Mainly due to the large amount of alcohol consumption liver disease arises. Early prediction of liver disease using classification algorithm is an efficacious task that can help the doctors to diagnose the disease within a short duration of time. Discovering the existence of liver disease at an early stage is a complex task for the doctors. The main objective of this paper is to analyse the parameters of various classification algorithms and compare their predictive accuracies so as to find out the best classifier for determining the liver disease. This paper focuses on the related works of various authors on liver disease such that algorithms were implemented using Weka tool that is a machine learning software written in Java. Various attributes that are essential in the prediction of liver disease were examined and the dataset of liver patients were also evaluated. This paper compares various classification algorithms such as Random Forest, Logistic Regression and Separation Algorithm with an aim to identify the best technique. Based on this study, Random Forest with the highest accuracy outperformed the other algorithms and can be further utilised in the prediction of liver diseaserecommended.

Liver Patient Prediction

Age:

Gender:

Enter 0 as male, 1 as female

Total_Bilirubin:

Direct_Bilirubin:

Alkaline_Phosphotase:

Alamine_Aminotransferase:

Aspartate_Aminotransferase:

Total_Protiens:

Albumin:

Albumin_and_Globulin_Ratio:

Predict

Liver Patient Prediction

You have a liver desease problem, You must and should consult a doctor. Take care

3.ADVANTAGES & DISADVANTAGES

ADVANTAGES

- Clear diagnostic criteria

- Diagnostic value confirmed

- May suggest the etiology

- Can perform differential diagnosis

- Assess the degree and stage of liver damage

- It can decide the therapy

DISADVANTAGES

- Major invaise test

- Comlications include death

- Significant sampling errors

- HIgh cost

- Inter-observer variability

5.APPLICATION

The liver is the largest solid organ in the body. It removes toxins from the body's blood supply, maintains healthy blood sugar levels, regulates blood clotting, and performs hundreds of other vital functions. It is located beneath the rib cage in the right upper abdomen.

Key Facts

- The liver filters all of the blood in the body and breaks down poisonous substances, such as alcohol and drugs.

- The liver also produces bile, a fluid that helps digest fats and carry away waste.

- The liver consists of four lobes, which are each made up of eight sections and thousands of lobules (or small lobes).

Functions of the Liver

The liver is an essential organ of the body that performs over 500 vital functions. These include removing waste products and foreign substances from the bloodstream, regulating blood sugar levels, and creating essential nutrients. Here are some of its most important functions:

- **Albumin Production**: Albumin is a protein that keeps fluids in the bloodstream from leaking into surrounding tissue. It also carries hormones, vitamins, and enzymes through the body.

- **Bile Production**: Bile is a fluid that is critical to the digestion and absorption of fats in the small intestine.

- **Filters Blood**: All the blood leaving the stomach and intestines passes through the liver, which removes toxins, byproducts, and other harmful substances.

- **Regulates Amino Acids**: The production of proteins depend on amino acids. The liver makes sure amino acid levels in the bloodstream remain healthy.

- **Regulates Blood Clotting:** Blood clotting coagulants are created using vitamin K, which can only be absorbed with the help of bile, a fluid the liver produces.

- **Resists Infections**: As part of the filtering process, the liver also removes bacteria from the bloodstream.

- **Stores Vitamins and Minerals**: The liver stores significant amounts of vitamins A, D, E, K, and B12, as well as iron and copper.

- **Processes Glucose**: The liver removes excess glucose (sugar) from the bloodstream and stores it as glycogen. As needed, it can convert glycogen back into glucose.

6.CONCLUSION

Bioartificial liver therapy for bridging patients with ALF to liver transplantation or liver regeneration is promising. Its clinical value awaits further improve- ment of BAL devices, replacement of hepatocytes of animal origin by human hepatocytes, and assessment in con- trolled clinical trials.

7.FUTURE SCOPE

The proposed liver disease prediction (LDP) method has provided the right path for liver disease detection. From the results of this study, after balancing the dataset, SVM has 78.1%, and Naïve Bayes has 65.1%. This balancing of the dataset using ROSE significantly changes the accuracy compared to the accuracies produced by Auxilia (2018), which is 77% for SVM and 37% for Naïve Bayes.

Singh et al. (2020) also focused on the same dataset of liver patients with feature engineering done with WEKA. After feature engineering, only five attributes are selected for the analysis, and algorithms are applied. The common algorithm from this research and Singh et al. (2020) is Naïve Bayes and has an accuracy of 55.9% with only five attributes selected. By comparing that to the results of this study, Naïve Bayes has an accuracy of 65.1%. As shown in Figure 11, only three attributes are less likely correlated with class attributes, but the rest are correlated with the class attribute, affecting the accuracy. The attributes that are not correlated with class attributes can be removed, which gives the better performance of algorithms and maximised accuracy. So, from Singh et al. (2020) research, some relatable features are dissolved in feature engineering, impacting accuracy. Thus, if this research needs to be done differently, it can include some more instances for better prediction. As the given dataset has only 583 instances, they can be increased in number for a better prognosis. Along with increasing the instances, different attributes important to predict liver disease like triglycerides, urine copper, serum cholesterol, and serum glutamic-oxaloacetic transaminase (SGOT) could be added to improve the chances of liver disease prediction (Assegie et al., 2022).

8.APPENDIX

A.Source Code

```
import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from matplotlib import rcParams

from scipy import stats


data=pd.read_csv('/content/indian_liver_patient.csv')

data.head()

data.info()

data.isnull().any()


data['Albumin_and_Globulin_Ratio']=data.fillna(data['Albumin_and_Globulin_Ratio'].mode()[0])

data.isnull().sum()
```

```python
from sklearn.preprocessing import LabelEncoder

lc=LabelEncoder()

data['Gender']=lc.fit_transform(data['Gender'])


data.describe()

sns.distplot(data['age'])

plt.title('Age Distribution Graph')

plt.show()


sns.countplot(data['outcome'],hue=data['Gender'])


plt.figure(figsize=(10,7))

sns.heatmap(df.corr(),annot=True)


from sklearn.preprocessing import scale

X_scaled=pd.DataFrame(scale(X),columns=X.columns)


X_scaled.head()


X=data.iloc[:,:-1]

Y=data.outcome


from sklearn.model_selection import train_test_split

X_train,X_test,y_train,y_test=train_test_split(X_scaled,y,test_size=0.2,random_state=42)
```

```python
from imblearn.over_sampling import SMOTE

smote=SMOTE()

y_train.value_counts()

X_train_smote,y_train_smote=smote.fit_resample(X_train,y_train)

y_train_smote.value_counts()



from sklearn.ensemble import RandomForestClassifier
model1=RandomForestClassifier()
model1.fit(X_train_smote,y_train_smote)
y_predict=model1.predict(X_test)
rfc1=accuracy_score(y_test,y_predict)
rfc1
pd.crosstab(y_test,y_predict)
print(classification_report(y_test,y_predict))



from sklearn.ensemble import DecisionTreeClassifier
model4=DecisionTreeClassifier()
model4.fit(X_train_smote,y_train_smote)
y_predict=model4.predict(X_test)
```

```python
dtc1=accuracy_score(y_test,y_predict)

dtc1

pd.crosstab(y_test,y_predict)

print(classification_report(y_test,y_predict))




from sklearn.neighbors import KNeighnorsClassifier

model2=KNeighborsClassifier()

model2.fit(X_train_smote,y_train_smote)

y_predict=model2.predict(X_test)

knn1=(accuracy_score(y_test,y_predict))

knn1

pd.crosstab(y_test,y_predict)

print(classification_report(y_test,y_predict))




from sklearn.linear_model import LogisticRegression

model5=LogisticRegression()

model5.fit(X_train_smote,y_train_smote)

y_predict=model5.predict(X_test)

logi1=accuracy_score(y_test,y_predict)

logi1

pd.crosstab(y_test,y_predict)

print(classification_report(y_test,y_predict))
```

```python
import tensorflow.keras

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense


classifier=Sequential()


classifier.add(Dense(units=100,activation='relu',input_dim=10))


classifier.add(Dense(units=50,activation='relu'))


classifier.add(Dense(units=1,activation='sigmoid'))


classifier.compile(optimizer='adam',loss='binary_crossentropy',metrics=['accuracy'])


model_history=classifier.fit(X_train,batch_size=100,validation_split=0.2,epochs=100)


model4.predict([[50,1,1.2,0.8,150,70,80,7.2,3.4,0.8]])


model1.predict([[50,1,1.2,0.8,150,70,80,72,3.4,0.8]])


classifier.save("liver.hs")


y_pred=classifier.predict(X_test)


y_pred
```

```python
y_pred=(y_pred>0.5)

y_pred


def predict_exit(sample_value):

    sample_value=np.array(sample_value)

    sample_value=sample_value.reshape(1,-1)

    sample_value=scale(sample_value)

    return classifier.predict(sample_value)


sample_value=[[50,1,1.2,0.8,150,70,80,7.2,3.4,0.8]]
if predict_exit(sample_value)>0.5:
    print('Prediction:Liver Patient')
else:
    print('Prediction:Healthy')


acc_smote=[['KNN
Classifier',knn1],['RandomForestClassifier',rfc1],['DecisionTreeClassifier',dtc1],['Logistic
Regression',logi1]]

Liverpatient_pred=pd.DataFrame(acc_smote,columns=['classification
models','accuracy_score'])

Liverpatient_pred
```

```python
plt.figure(figsize=(7,5))

plt.xticks(rotation=90)

plt.title('Classification models & accuracy scores after SMOTE',fontsize=18)

sns.barplot(x="classification models",
y="accuracy_score",data=Liverpatient_pred,palette="Set2")


from sklearn.ensemble import ExtraTreesClassifier

model=ExtraTreesClassifier()

model.fit(X,y)


model.feature_importances_


dd=pd.DataFrame(model.feature_importances_,index=X.columns).sort_values(0,ascending=False)

dd


dd.plot(kind='barh',figsize=(7,6))

plt.title("FEATURE IMPORTANCE",fontsize=14)


import joblib

joblib.dump(model1,'ETC.pk1')


from flask import Flask,render_template,request

import numppy as np

import pickle
```

```python
app=Flask(_name_)

@app.route('/')

def home():

    return render_temlate('home.html')

@app.route('/predict')

def index():

    return render_template("index.html")


@app.route('/data_predict',methods=['POST'])

def predict():

    age=request.form['age']

    gender=request.form['gender']

    tb=request.form['tb']

    db=request.form['db']

    ap=request.form['ap']

    aa1=request.form['aa1']

    aa2=request.form['aa2']

    tp=request.form['tp']

    a==request.form['a']

    agr=request.form['agr']



data=[[float(age),float(gender),float(tb),float(db),float(ap),float(aa1),floa(aa2),float(tp),float(a),float(agr)]]
```

```python
model=pickle.load(open('liver_analysis,pl1','rb'))


prediction=model1.predict(data)[0]

if(prediction==1):

    return render_template('nochance.html',prediction='You have a liver deseas
problem,You must and')

else:

    return render_template('chance.html',predivtion='You dont have a liver desease poblem')


if __name__=='__main__':

    app.run()


if __name__=='__main__':

    app.run()
```