InternationalJournalofAdvancesinEngineeringandManagement(IJAEM)Volume4,Issue6June 2022,pp:425-428www.ijaem.netISSN:2395-5252

Chronic Liver Disease Detection SystembasedonMLAlgorithms

G.M.Padmaja[1],M.Gayatri[2],K.Raviteja[3],M.Tarun[4],K.Raviteja[5]
[1]AssistantProfessor,RaghuInstituteofTechnology,Visakhapatnam,AndhraPradesh,India.
[2,3,4,5]Student,RaghuInstituteofTechnology,Visakhapatnam,AndhraPradesh,India.

Submitted:01-06-2022 Revised:05-06-2022 Accepted:08-06-2022

ABSTRACT: Chronic liver disease is one of theleadingcausesofdeathworldwideandaffectsmany people throughout theworld. Thischronicliver disease is caused by a variety of factors. Weknow many of them, such as undiagnosed hepatitis,obesity,andalsoalcoholmisuse.Causesofabnormalnervefunction,coughingorvomitingblood, hepatic encephalopathy, kidney failure, liverfailure, jaundice, and many more are just symptomsofwhat'smoretocome.Diagnosisofthisdiseaseis also quite expensive and at the same time, it isvery complicated. Therefore, the goal of this paperwastoevaluatetheperformanceofvariousmachine learning algorithms to reduce the high costofdiagnosingthischronicliverdiseasebypredicting the disease. In this paper, we used threeMachinelearningalgorithmsi.eRandomForest, XGBoost,andExtraTreeClassifier.Theperformance of these classification techniques wasevaluated against different measurement techniquessuch asprecision, accuracy,f1-score, andrecall.Furthermore, our current study's only main focuswas on using clinical data to predict liver diseaseand exploring different ways to represent these datathroughouranalysis.

KEYWORDS:Chronic,MachineLearning,Classification,ExtraTreeClassifier,XgBoost,RandomForest.

I. INTRODUCTION

The liveris the largest organin the body,it is needed to digest food and remove toxins fromthe body. Alcohol consumption and Virusesleadtoliver damage and sometimes they may also lead tolife-threateningconditions.Manytypesofliverdiseases include hepatitis, cirrhosis, liver tumors,liver cancer, and many more. Among them, liverdisease and cirrhosis are the main causes of death.Therefore,thisliverdiseaseisoneofthemajorhealthproblemsintheworldandthatisthereason why it is called a chronic disease. Every year, morethan 2 million people around the world die fromliver disease. According to the Global Burden ofDisease report, which was later published by BMCMedicine, in the year 2010, over one million peoplediedfromcirrhosisdiseasewhichisoneofthesymptoms of liver disease, and one million fromLiver-Cancer.Machinelearninghashadasignificantimpactinthemedicalfieldforpredicting and giving diagnosing methods for liverdisease.

Now,Machinelearningguaranteesimproved detection and prediction of diseases thathave aroused interest in the biomedical field andalsoincreases theobjectivity in decision-making.By using these ML techniques, medical problemscan be solved easily and the cost of diagnosiscanbe reduced.In this paper, themain goalwastopredictoutcomesmoreefficientlyandreducediagnostic costs in the medical field. In this paper,onlythreemachinelearningtechniqueswereapplied: RF, XGB, and XTC. The performance ofthese techniques has been estimated from variousaspectssuchasaccuracy,recall,f1-score,andprecision.

II. LITERATURESURVEY

Norziahet.al[1]predictedahepatitisprognosisdiseaseusingtheSupportVectorMachine(SVM)and WrapperMethod.First,forthe classificationprocess,they'veusedwrappermethods to remove the noise features and let theSVMcarryoutfeatureselectiontogetbetteraccuracy.Featuresselectionwasimplementedtominimizenoisyorirrelevancedata.Theyhaveachievedthetargetresultbythecombinationofboth the Wrappers Method and SVM techniques.Accuracy:81%

Dayanandet.al[2]havepredictedthreemajor liverdiseasesLivercancer, Cirrhosis,andHepatitis with the help of distinct symptoms. They'their project using the Naïve Bayes algorithm andSVMalgorithm.Whenacomparisonisdonebetween these two algorithms has been done basedon their classification accuracy measure, from theexperimental results, they concluded that the NBalgorithm was the better algorithm that predicteddiseaseswithmaximumclassificationaccuracythan theother algorithm.Accuracy:55.6%

Rakshitet.al[3]have useda computingtechnique combined with the intelligent diagnosis todetectchronicliverdiseasewhichisbasedonimplementingtheclassificationandthetypedetection for this project in the most complex way.For the project, they have used the Artificial NeuralNetwork(ANN)classificationalgorithm.Accuracy:79%

III. PROBLEMSTATEMENT

This paper aims to analyze the dataset ofliver disease patients, based on certain diagnosticmeasurementsincludedinthedataset.Severalconstraints were placed on the selection of theseinstancesfromalargerdatabaseandalsothisdataset, male patients were more when comparedthe females. In many machine learning papers, itcontains only one type of algorithm. But here, wehaveusedonlythreetypesofalgorithmsandcompareall ofthemfor thebest result.

IV. METHODOLOGY

Our main goal of this paper is to find a suitableMachineLearningtechniquethatcandetectChronicLiverDiseasewithhighaccuracy.
AlthoughwehavetakenonlythreeMachinelearning techniques, (i.e Random Forest, XgBoost,and eXtra Tree Classifier)we have also used a fewmore popular ML algorithms in our sample tests, inwhich these 3 ML algorithms yield high results forthedatasetthatwehave obtained.
From thedataset,wehaveremovedinconsistentdata, Encoded the categorical data, and splitthedata for further analysis. For the execution part, wehave used python programming language and theeditor used was Google Colab as it is useful forcollaborativework andhelpedustoexecutethecodeeasilywithoutanydirectlyinstallinglibraries.

4.1 SystemArchitecture

A system architecture is a conceptual model thatdefinesthestructure,behavior,andviewsofasystem.Figure4.1describesthebehaviorandconceptthat weareusinginthispaper.

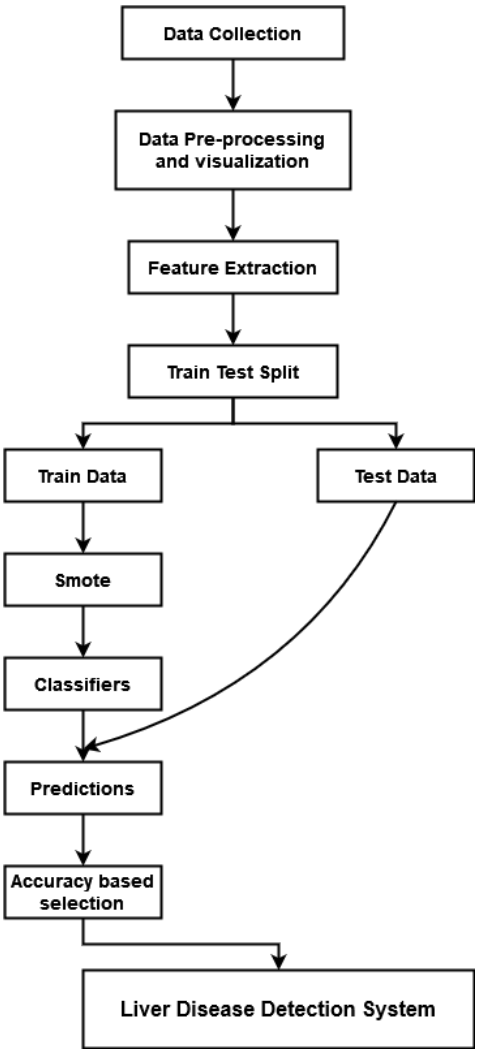

Figure4.1:SystemArchitecture

4.2 DataCollection

In this paper, we acquired a dataset that ismadeavailablebytheNationalInstituteofgastroenterology and LiverDiseases(NIGLD)which has the survey records ofLiver disease patients. These survey records wereprocessed by the National Health Portal and wereuploadedtovariouspublicwebsites.

This dataset that we have used consists of583 liver patients' data whereas75.64% were malepatients and 24.36% were female patients also anypatient whose age exceeded 89 is listed as being ofage "90" This dataset that we have used consists ofover11particularparameterswhereaswechoseonly ten parameters for our further analysis and oneparameter as a target class. The "Dataset" column isa class label used to divide groups into the liverpatient(liverdisease)ornot(nodisease).Theparametersinthedatasetwerelisted inTable4.2.

Table4.2:AllattributesinthegivenDataset

```
RangeIndex: 583 entries, 0 to 582
Data columns (total 11 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Age                         583 non-null    int64
 1   Gender                      583 non-null    object
 2   Total_Bilirubin             583 non-null    float64
 3   Direct_Bilirubin            583 non-null    float64
 4   Alkaline_Phosphotase        583 non-null    int64
 5   Alamine_Aminotransferase    583 non-null    int64
 6   Aspartate_Aminotransferase  583 non-null    int64
 7   Total_Protiens              583 non-null    float64
 8   Albumin                     583 non-null    float64
 9   Albumin_and_Globulin_Ratio  579 non-null    float64
 10  Dataset                     583 non-null    int64
dtypes: float64(5), int64(5), object(1)
memory usage: 50.2+ KB
```

4.3 Datasplitting

Afterhandlingthenoisyorinconsistentdata in the dataset, we split our data into a trainingset and a testing set. Most of the time we split ourdata intoeither70:30or80:20ratios.

For this project, we decided to divide our datasetinto an 80:20 ratio and which means eighty percentof the dataset is training data and the remainingtwentypercent ofourdataistestingdata.

V. ALGORITHMSUSED

5.1 RANDOMFOREST(RF)

Randomforestortherandomdecisionforest is one of the ensemble learning techniques ofMachine learning for classification, regression, andvarious assignments that works by spanning a largenumber of decision trees at the time of training andgeneratingclassisamethodofclassesorapredictive averageofeachofindividualtrees.Randomdecisionforestsadaptdecisiontrees'propensity to adapt to their given training set. Fromthesecombinedtrees,thereisanimmediateconnectionbetweenthecombinedtreesandtheresultstheycan achieve.

5.2 eXtremeGradientBoosting(XGBoost)

XGBoostoreXtremeGradientBoostingalgorithm is one of the popular and efficient open-sourcealgorithmswhichhelpsintheimplementationof theenhanced-gradienttreealgorithmfromClassificationtechniques ofMachine-learning.TheEnhanced-Gradienttreeisa supervisedmachinelearningalgorithmthatcanattempt to accurately predict a target variable bycombining some set of the estimates from a set of asimplerandaweakermodel.

5.3 ExtraTreeClassifier(XTC)

Extratreesclassifierisoneoftheensemble learning techniques which was helping intheaggregationofvariousresultsofhugeuncorrelated decision trees which were collected ina"forest"togenerateitsclassificationresults.Conceptually, it is very similar to a random forestclassifier and differs only in the way decision treesare constructed in the forest. Each decision tree inthe additional tree forest is built from the originaltrainingsamplesgiventoit.Then,ateachtestnode,every treereceivesarandom sampleofkfeaturesfromthefeaturesetfromwhicheach decision tree should choose the bestfeatureof allofittosplitthedataaccordingtosomemathematicalcriterion.

VI. RESULTS&ANALYSIS

Inthispaper,weconsidereddifferentanalyses toexaminethethree-machinelearningclassifiers for the classification of chronic Liver-Disease datasets. In terms of metrics, i.e accuracy,recall,f1-score,andprecision,theeXtraTreeClassifier algorithm(XTC) that we have used hasachievedthehighestaccuracyofover83.8%,precision:0.86,recall:0.83,andf1-score:0.84whereas, XGBoost has achieved the leastperformanceamongallthreeMLmodelsusedwhich has an accuracy of77.8% for the dataset thatwehavegiven.

Accordingtothecomparisonofthesemeasurementcriteria,theXTC(ExtraTreeClassifier)classificationtechniqueinMachine-Learningismuchmoreeffectivethantheotherclassifiers that we used for predicting this chronicLiverdiseaseforourgivendataset.Theperformancecomparisonofthethreesupervisedmachine-learning techniques that we have used willbeshownintheresulttable (Table6.1)

Table6.1:ResultsoftheAlgorithms

	Accuracy	Precision	Recall	F1- Score
Random Forest	81.4%	0.84	0.81	0.82
eXtreme Gradient Boosting	77.8%	0.81	0.76	0.78
Extra Tree Classifier	83.8%	0.86	0.83	0.84

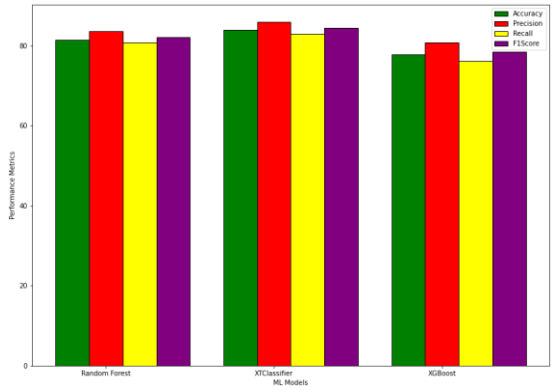

Figure6.1GraphicalRepresentation

VII. CONCLUSION

The main objective of this paperis for ustomakeaneffectiveandanefficientdetectionsystemf orliverdiseasepatientsbyusingthreedistinctive and supervised ML techniques where weresearched all classifier's execution of all patients'information along with parameters and the ExtraTreeclassifiergivesthemostelevatedorderexact ness of 83.8 %measure to predict liver diseaseand XGBoost only gives the least precision 77.8 %.This project has the option to predict liver infectionbefore advisingthewell-beingcondition.

This paperis very surprisingly gainful inlow-salarynationswhereifanyabsenceofmedicinal foundations and just particular specialistsoccurs. We only explored some popular supervisedmachine learning algorithms, more algorithms canbepickedtoassembleanincreasinglyprecisemodel of liver disease prediction and performancecan be progressively improved. Additionally, thisworks likewise ready to assume a significant role inhealth care research and just as restorative focusesto anticipateliverinfection.

REFERENCES

[1].APredictionofHepatitisPrognosisUsingSupport Vector Machine and WrapperMethod written by Roslina and A.Norziah,IEEE2209-22.

[2].ApaperonLiverDiseasePredictionusingSVMand NBAlgorithmswrittenbyS.DhayanandandS.V ijayaraniintheInternationalJournalofScience, Engineering, andTechnology ResearchVol4,Issue4,April2015.

[3].Apaperonliverdiseasedetectionsystemsusing MachineLearning Techniques writtenbyD.B.Rakshith,S.PGururaj,Ashwani Kumar,andMrigankSrivastavaInternationalJo urnalofEngineeringResearch & Technology, Vol. 10 Issue 06,June-2021.

[4].Thedataset that we acquired fromKaggle: kaggle.com//datasets//uciml//indian-liver-patient-records.

[5].Types,Symptoms,andCausesofChronicdisease were mentioned detailed inClevelandClinicmy.clevelandcli nic.org//health//diseases//17179-liver-disease.