International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.136

Prediction and Classification of Binary and Multi-class Heart disease with Artificial Intelligence

Naga VenkatMarneni
Department of Data Science and Business Systems SRM Institute of Science and Technology,
Kattankuathur
mn8362@srmist.edu.in

VarunKhareedu
Department of Data Science and Business Systems SRM Institute of Science and Technology,
Kattankuathur
kv6122@srmist.edu.in

T. Veeramakali
Department of Data Science and Business Systems School of Computing SRM Institute of Science and Technology
Kattankuathur
veeramat@srmist.edu.in

Abstract—Heart disease posesa significant threat to human existence because of its high mortality and morbidity rates. For early treatment, localization, and countermeasure, precise anticipation and conclusion are becoming increasingly important. The Internet of Things and artificial intelligence make it easier for doctors to find, evaluate, and diagnose cardiac disease. However, the majority of prediction models only assess the severity of an individual's illness and only predict whether they are ill. In this study, wepresentahypothesismodelforsimultaneousassumptionforboth matchedanddistinct request coronary illness that is based on machine learning (ML). We first develop a fuzzy-GBDT strategy that combines cushy reasoning with gradient boosting decision tree (GBDT) to advance the matched portrayal assumption and reduce data complexity. To avoid over fitting, the pressing is completed with fleecy GBDT. Theseverityofcardiacdiseaseisalsoincludedinthemulticlassification assumption based on Bagging-Fuzzy-GBDT.TheassessmentshowsthatthePackingFluffyGBDTisincrediblydependableandexact for both twofold and different get-together presumptions.

Keywords :Fuzzy logic, gradient boosting decisiontree, theInternetofMedicalThingscardiacdiseasepredictionanddiagnosis,andmachine learning

I. INTRODUCTION

One of the most difficult and life-threatening diseasesaffecting humans, heart disease has a high mortalityand morbidity rate [1]. People's quality of life suffersasaresult,andtreatmentandmonitoringincursignificantcosts.Itispossibletoanticipate,recognize,anddiagnosehealthconditionswithartificialintelligence(AI)[2].Bymakingitpossibleforpatientstoreceivetheappropriatemedicalinformation,treatment,andintervention,itmaylessenthedevastationcausedbyheart disease.Continuouscardiacillnesspredictionandsuggestiveoutcomesmay be possible for e-medical care frameworks thatheavily rely on Internet of Medical Things (IoMT)data using AI learning algorithms [3]-[6]. In addition,italleviatesthefinancialandadministrativechallenges associated with intelligent systems for the treatment,monitoring,andpreventionofchronicdiseases.However,howtoguaranteethestrength,generalizability,andhighaccuracyofML-basedexpectationmodelsandcomputationsmustbeaddressed.

Intoday'ssociety,the ideaofanticipatingcardiacdisease is a significant one that is altering people'sperceptionsofhealth.ThebasicconceptistousetheRandom forest algorithm to figure out the age groupandheartrate.Basedonuser-suppliedinputslikeblood pressure and other variables, our study showshowasystemanalyzesheartrateandcondition.Compared to other algorithms, RFA produces resultsthataremoreaccurateandoffersabetteruserexperience. The assessment of a person's heart rate inrelation to their overall health is just one of the manyusesfor this.Additionally,itaidsintheearlydetectionofdisease.

The bootstrap totalling (bagging) method is added tothelearningmodeltoincreasetheareaunderthecurve(AUC)whiledecreasingchange.However,theaccuracy of any prognosis has not been establishedbecause previous studies on the predictionof heartdiseasehaveheavilyreliedoncomplexdata.Theinitial multi-category method for diagnosing variousriskgroupsforheartdiseasestillhasroomforimprovementintermsofaccuracy.

Forbothbinaryandmultiple-orderheartdiseaseexpectations,wepresentastableandpreciseexpectation method in this work. By simplifying theinput,fuzzylogic,ontheotherhand,encouragesmodel generalization and reduces model deviation. Byreducing the change and deviation of the assumptionmodel, the unrivaled Bagging-Fuzzy-GBDT enhancesestimate exactness and adequacy. Early detection ofheartdiseaseinhigh-risk,upgradedsymptomaticindividualsusinganexpectationmodelhasbeenwidely proposed to cut down on deaths and furtheradvancetreatmentandpreventionoptions.InCDSS,aforecast modelismadeandusedtohelpdoctorsfigureout why people are so likely to get heart disease andgivethemtherightdrugstolowerthatrisk.Inaddition,a number of studies have demonstrated that utilizingCDSS may enhance the quality of decisions, clinicalnavigation,anddeterrentconsideration.Coronaryartery disease (CAD), also known as ischemic heartdisease (IHD), is the most common cause of deathamong people over the age of 35 in some nations. Italso rose to become the most common cause of deathin China over the same time period. IHD occurs whencoronary artery stenosisreducesbloodflow totheheart. Myocardial damage can result in a potentiallyfatal myocardial infarction, which can either result inventriculararrhythmiaorsuddencardiacdeath.

II. LITERATURE SURVEY

[1] Theauthorsinthispaperexploredtheuseoftrendanalysisintelemonitoringdatatopredictdecompensationeventsinheart failurepatients.Decompensation events are acute worsening of heartfailure symptoms that oftenlead to hospitalization,reducedqualityoflife,andincreasedhealthcarecosts.Earlypredictionandinterventionarecrucialtopreventthese eventsandimprovepatientoutcomes.

[2] The authorsexplorethe conceptof home carerobotic systems offering avisionforthe future ofhealth care and

DOI: rp-9788770040723.136
discussing the enabling technologiesthat can facilitate the development and deployment ofthese systems. Home care robotic systems have thepotentialtorevolutionizehealthcarebyprovidingpersonalized, convenient, and efficient care in patients'homes, which is particularly relevant given the aginggl obal population and increasing demand for healthcareservices.

[5] TheauthorscombinedeepLongShort-TermMemory(LSTM)recurrentneuralnetworkswithadaptive kernel spectral clustering to provide a novelmethod for monitoring health with AI.. The primarygoal of this research is to enhance the effectiveness ofmachinehealthmonitoringsystems,whichareessential for preventing failures, reducing maintenancecosts, and ensuring the safe operation of machinery.theauthorsemploydeepLSTMrecurrentneuralnetworks to model the temporal dependencies withinthe extracted features. This approach is particularlyeffectiveinhandlingtimeseriesdata,asLSTMnetworkscancapturelong-termdependenciesandlearncomplexpatterns. The deepLSTM built andtrained to predict the machine's health state, enablingearlyidentificationofpotentialfailuresordeteriorating conditions.

[6] This study gives a novel approach to the detectionandlocalizationofischemicheartdisease(IHD)using magnetocardiography (MCG) and machine learningtechniques.Reduced blood flow to the heart muscle istheprimarycauseofischaemicheartdisease,whichisalso a leading cause of death globally. The authors'work is significant as it offers a promising alternativetotraditionaldiagnosticmethodsanddemonstratesthepotentialofmachinelearninginimprovingtheaccuracyandefficiency ofIHDdetection.

[9] The authors investigate the potential of GraphicsProcessing Units (GPUs) for improving the efficiencyof Gradient Boosting Decision Tree (GBDT) training.Becauseto itsexcellent accuracy andcapacity forhandlingbigdatasets,GBDTisapotentmachinelearningmethodthatisfrequentlyusedforavarietyoftasks,includingclassificationandregression.However,thetrainingprocessforGBDTcan becomputationallyintensiveandtime-consuming,especially when dealing with large-scale data. Thisresearchissignificantbecauseitaddressesthechallenge of GBDT training efficiency by leveragingthe parallel processing capabilities of GPUs, whichcanleadtofasterandmoreefficientmodeltraining.

[14] This study investigated the prediction of prefixavailability in the Internet. Prefix availability refers tothe reachability of aspecific IP prefix,whichcanimpactthestabilityandperformanceofInternet communication.Theauthorsusedavarietyofmachinelearning approaches, including decision trees, logisticregression, and support vector machines. to developmodelsforpredictingprefixavailability.Manymetrics, such as precision, recall, F1-score, and areaunderthereceiveroperatingcharacteristic(ROC)curve,wereusedtoassess themodels'performance.

[15] Thisstudyproposesmethodthatcombines Traditionalclassificationtogeneticalgorithmstoenhance accuracy and efficiency of diagnosis. Heartdisease is a major global health concern, and earlydetectioniscrucialforsuccessfultreatmentandmanagement. The authors' work is significant becauseitpresentsanovelapproachtoheartdiseaseprediction,leveragingthestrengthsofbothassociativeclassification and genetic algorithms to improve theperformanceoftraditionalpredictionmethods.Agenetic algorithm to optimize the rule set generated inthe first phase. The genetic algorithm searches for thebest combination of rules that maximizes predictionaccuracy while minimizing the complexity of the ruleset.Thisoptimizationprocessresultsinamoreefficient and accurate prediction model compared tousingassociativeclassificationalone.

III. Methodology

As previously demonstrated, most current assumptionmodelsandcalculationsonlyaddressthematchedportrayal issue of heart disease speculation withouttakingintoaccounttheactualseverityofthecondition.Based on angiographic data, the level severity of thedisease is divided into five classes, ranging from modiseasetolevelfour.Inanycase,theprimarymulticlassificationmethod'saccuracyinrequestingvariousheartdiseasebetclassesneedstobeimproved.Reducing fluctuation and deviation is frequently usedinmachinelearningtoimprovecomputationaccuracy.Thelearningmodelaccomplishesabiggerregionunderthebend(AUC)andconfineschangebyincludingthebootstrapcollecting (packing)technique.However,theexactnessofeachtypeofcoronarydisease prediction has not been provided in previousstudiesoncardiacillnessforecastingduetokeyinformationcomplexities.Infactthere'sstillchancetoincrease the accuracy in multiclassification problem.Although it does not cover every kind of heart diseaseanticipation, previous research on the prediction ofheartdiseasecontainsasignificantamountofinformationcomplexity.

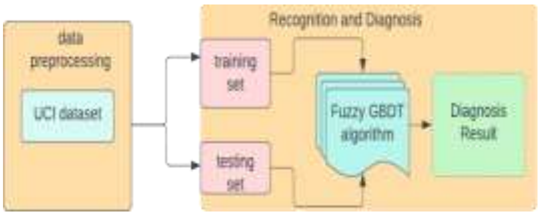
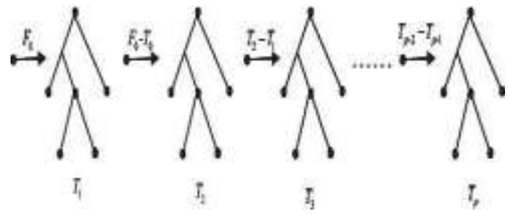


Fig1SystemArchitecture

Thomas et al. forecasted each player's wager. In anycase,theprimarymulticlassificationmethod'saccuracyinrequestingvariousheartdiseasebetclassesneedstobeimproved.Fromonepointofview,by making informationsimpler,fuzzy logic makesmodelspeculationbetterandreducesmodeldeviation.To improve accuracy, we incorporate fluffy logic intotheGBDTcomputation.Incontrast,thesackingmethod reduces model change by repeating erroneousinspections. Accordingly, we incorporate the puttingaway strategy to

DOI: rp-9788770040723.136



build the model's solidarity. Whencontrastedwithebbandflowestimations,ourproposedrecipecanforeseewhetherpeoplearedebilitatedaswellastheseriousnessofheartsickness.

Fig2SchematicdiagramofGBDTAlgorithm

ThebenefitsofaFuzzy-GBDT-basedtwofoldcharacterizationexpectationtechniqueforheartdisease conclusion are presented. These advantagesinclude an increase in the GBDT's speculation limitand a reduction in the complexity of the informationregardingheart disease.

ThesuperiorBagging-Fuzzy-GBDTmanagesstrengthand accuracy of figures. By minimizing the changeand deviation in the assumption model, we present inthisarticleaconsistentandhigh-precisionsuppositionmethod for both twofold and numerous heart diseasesolicitations.

Modules

The below modulesweredevelopedby usforthenotioninourproject.

- Data investigation: We will enter data into thesystemwiththismodule.
- Treatment: We will read data for processingusingthismodule.
- Separating train and test data: Train and testdatawillbeseparatedbythismodule.
- Models that can be generated include SVM,RF,DT,LR,KNN,XGBoost,GaussianNaiveBayes, VotingClassifier,GBDT,Bagging+GBDT,Fuzzy+GBDT ,andBagging+Fuzzy+GBDT.
- Login and registration for users: Sign Up andauthentication are required in order to accessthismodule.
- User-providedpredictioninformation:Prediction input will result from using thismodule.
- Prognosis:Thefinalpredictedvalueisshown.

IV. Implementation

1. DataPre-processing

Wemakeuseoftheheartdiseaseopen-sourcedatasetfrom the University of California, Irvine (UCI) [25].Fourdistinct medical databases, including those inCleveland,Hungary,Switzerland,andVALongBeach,providedtheinformation.Thereare836dataintotaland14significantattributesinthis database.Most of the datasets Switzerland and VA Long Beachrequirethecompletionofmissingvalues.Theaveragevalues from other complete datasets are used to fill intheblankfields. Thedataset inthisstudyissplitintoatrainingsetandatestsetintheratioof7to3.

2. BaggingFuzzyGBDT Algorithm

TherearesixparametersfortheBagging-Fuzzy-GBDT algorithm that must be determined. They aredepicted as follows. The values of these parametershave an impact on the predictability and precision ofthe suggested model. Hence, one of the main issues ishowtodeterminetheidealvalues.

(i) M decision trees are present. Each iteration ofthe Bagging-Fuzzy-GBDT growth process resultsinthecreationofadecisiontree.

(ii) Each decision tree can have a maximum depthof MD. MD cannot have an excessively high orlow value. The algorithm takes too long to run ifMD is too large since each tree's training period isprolonged.

(iii) To separate an internal node, MS samples areneeded as a bare minimum. Two circumstancesdefinethevalue ofMS.

(iv) AleafnoderequiresMLsamplestobepresentas a minimum. A splitting point of any depth willonlybetakenintoconsiderationifbothofitsleftandright branchesstillcontainatleastMLtrainingsamples.

(v) Therewillbemsamplescollectedforbagging.Tocreatemsubdatasets,mis the numberofsamplesfromtheoriginalsetthatmustbereplaced.

(vi) Myrateoflearning.Byregulatingthecontribution of a single decision tree in the modelusingtheregularisationtechniqueknownaslearningrate,anoverfittingoftheBagging-Fuzzy-GBDTcanbeprevented.

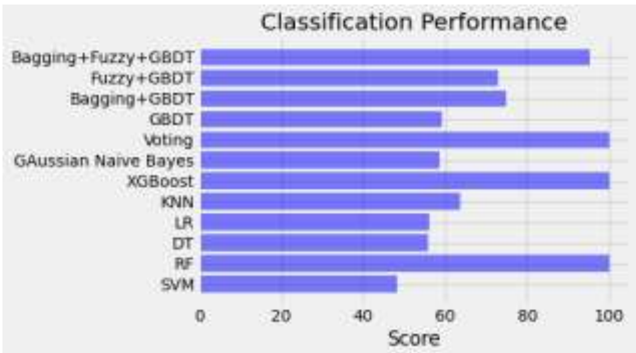
V. Performance Evaluation

GBDT

*Classification Report for GBDT:

	precision	recall	f1-score	support
0	0.73	0.82	0.77	207
1	0.51	0.55	0.53	161
2	0.47	0.42	0.44	71
3	0.39	0.24	0.30	71
4	0.53	0.48	0.50	21
accuracy			0.59	531
macro avg	0.52	0.50	0.51	531
weighted avg	0.57	0.59	0.58	531

Fig5PrecisionandRecall



DOI: rp-9788770040723.136

TheBagging-Fuzzy-GBDTapproachprovidesasignificant improvement in precisionand accuracy.First,GBDTperformsbetteratpredictingheartdiseasethantraditionaldecisiontreepredictionmodels.Second,using thecomparisonresultsbetweenBagging-GBDTandFuzzy-GBDTwithGBDT

*Classification Report for Bagging + Fuzzy + GBDT:

	precision	recall	f1-score	support
0	0.94	0.98	0.96	207
1	0.96	0.93	0.94	161
2	0.94	0.94	0.94	71
3	0.99	0.94	0.96	71
4	1.00	1.00	1.00	21
accuracy			0.95	531
macro avg	0.97	0.96	0.96	531
weighted avg	0.95	0.95	0.95	531

Fig 3 Precision and recallNaiveBayes

*Classification Report for NB:

	precision	recall	f1-score	support
0	0.62	0.91	0.74	207
1	0.63	0.65	0.64	161
2	0.74	0.39	0.51	71
3	0.65	0.21	0.32	71
4	1.00	0.10	0.17	21
accuracy			0.64	531
macro avg	0.73	0.45	0.48	531
weighted avg	0.66	0.64	0.60	531

Fig4PrecisionandRecall

In the experiment, the predicted accuracy foreachcategory ranges from 80 to 95%, showing that theBagging-Fuzzy-GBDTalgorithmperformsexceptionally well when multiclassification is beingused.Withanaverageaccuracyof93%,type2hadthebest prediction effect of the bunch. The 85% accuracyratefortype1andtype3predictionsisthesame.Eachtype's predictive performance is essentially the same,whichspeedsupandimprovesdiagnosiswhileproviding patientswithvarious therapiesbasedontheirindividualtypes.Therefore,comparingtheprecisionofeachindividualcategorizationformulticlassificationisunrealistic.

VI. CONCLUSION

For the cardiac disease prediction and detection, weproposedaconsistentandpreciseBaggingFuzzy-GBDTmethodinthisstudy.Inbothparallelanddistinct configurations, the proposed Bagging-Fuzzy-GBDT method predicted cardiac disease. To reduceinformation complexity and prevent overfitting, weintegrated fluffy logic and packing calculations intotheGBDTmethod.Themodel'ssecuritywassignificantlyimprovedwhentheborderswereexpanded using lattice search. In termsofperformance and accuracy, AUC, and other metrics,theevaluationrevealedthattheproposedmodelperformsbetterthanconventionalcomputationscurrently in use. In addition to accurately predictingillness, the Bagging-Fuzzy-GBDT computation alsodistinguishes the type of infection. In the field of e-medical services, it could be used to better understandthe conclusion and the board. We intend to refine theproposed model in the future and produce and test itspresentationwithauthenticandopendataincollaborationwithothernearbyinstitutions.

REFERENCES

[1] J. Henriques et al., "Prediction of heart failuredecompensationeventsbytrendanalysisoftelemonitoringdata",IEEEJ.Biomed.HealthInformat., vol.19,no. 5,pp.1757-1769,Sep.2014.

[2] G. Yang et al., "Homecare robotic systems forHealthcare 4.0: Visions and enabling technologies",IEEE J. Biomed. Health Informat., vol. 24, no. 9, pp.2535-2549,Sep. 2020.

[3] C.Lietal.,"Healthchain:SecureEMRsmanagementandtradingindistributedhealthcareservicesystem",IEEE InternetThingsJ., vol.8,no. 9,pp. 7192-7202,May2021.

[4] H.Li,K.OtaandM.Dong,"LearningIoTinedge:Deep learning for the Internet of Things with edgecomputing", IEEE Netw., vol. 32, no. 1, pp. 96-101,Jan. 2018.

[5] Y. Cheng, H. Zhu, J. Wu and X. Shao, "Machinehealthmonitoringusingadaptivekernelspectralclustering and deep long short-term memory recurrentneural networks", IEEE Trans. Ind. Informat., vol. 15,no. 2,pp.987-997, Aug.2018.

[6] P. Sun et al., "Modeling and clustering attackeractivitiesinIoTthroughmachinelearningtechniques",Inf.Sci.,vol.479,,pp.456–471,2019.

[7] R.Taoetal.,"Magnetocardiographybasedischemicheartdiseasedetectionandlocalizationusingmachinelearningmethods",IEEETrans.Biomed.Eng.,vol.66,no.6, pp.1658-1667,Jun. 2019.

[8] J. Li et al., "Heart disease identification methodusing machine learning classification in e-healthcare",IEEEAccess,vol.8,pp. 107562-107582,2020.

[9] Z. Wen et al., "Exploiting GPUs for efficientgradient boosting decision tree training", IEEE Trans.Parallel Distrib. Syst., vol. 31, no. 12, pp. 2706-2717,Dec.2019.

[10] L. Zhao et al., "InPrivate digging: Enabling tree-baseddistributeddataminingwithdifferentialprivacy", Proc. IEEE Int. Conf. Comput. Commun.,pp.2087-2095, 2018.

[11] B.Zhangetal.,"Healthdatadrivenoncontinuousbloodpressurepredictionbasedongradientboostingdecisiontreealgorithm",IEEEAccess,vol.7, pp. 32423-32433,2019.

[12] P.Chotwani,A.Tiwari,V.DeepandP.Sharma,"HeartdiseasepredictionsystemusingCART-C",Proc.Int.Conf.Comput.Commun.Informat.,pp.1-5,2018.

[13] J.ThomasandR.T.Princy,"Humanheartdiseasepredictionsystemusingdataminingtechniques", Proc. Int. Conf. Circuit Power Comput.Technol., pp. 1-5,2016.

[14] R. Khosla, S. Fahmy, Y. C. Hu and J. Neville,"Predicting prefix availability in the internet", Proc.IEEEInt.Conf.Comput.Commun.,pp.1-5,2010.

[15] M.A.Jabbar,B.L.DeekshatuluandP.Chandra,"Heartdiseasepredictionsystemusingassociativeclassification and genetic algorithm", Comput. Sci.,vol.1, pp.183-192,Dec. 2012.

[16] V. A. S. Hernndez et al., "A practical tutorial fordecisiontreeinduction:Evaluationmeasuresforcandidatesplitsandopportunities",ACMComput.Surv.,vol.54,no.1,pp.1-38,Jan.2021.

[17] D.Wu,C.T.Lin,J.HuangandZ.Zeng,"OnthefunctionalequivalenceofTSKfuzzysystemstoneuralnetworksmixtureofexpertsCARTandstackingensemble regression", IEEE Trans. Fuzzy Syst., vol.28,no.10,pp.2570-2580,Oct.2020. Moshika, A., Thirumaran, M., Natarajan, B., Andal, K., Sambasivam, G., &Manoharan, R. (2021).Vulnerability assessment in heterogeneous web environment using probabilistic arithmetic automata. IEEE Access, 9, 74659-74673.

[18] S.Pouriyehetal.,"Acomprehensiveinvestigationandcomparisonofmachinelearningtechniquesinthedomainofheartdisease",Proc.IEEESymp.Comput.Commun.,pp. 204-207,2017.

[19] Sitharthan, R., Vimal, S., Verma, A., Karthikeyan, M., Dhanabalan, S. S., Prabaharan, N., ...&Eswaran, T. (2023). Smart microgrid with the internet of things for adequate energy management and analysis.Computers and Electrical Engineering, 106, 108556.

DOI: rp-9788770040723.136

[20] PandZiemba,"NEATF-PROMETHEE—Anew fuzzy multiple criteria decision making methodbased on the adjustment of mapping trapezoidal fuzzynumbers", Expert Syst. Appl., vol. 110, pp. 363-380,Nov.2018.

[21] A.V.S.Kumar,"Diagnosisofheartdiseaseusingfuzzy resolution mechanism", J. Artif. Intell., vol. 5,no.1,pp. 47-55, Jan.2012.

[22] O. W. Samuel et al., "An integrated decisionsupport system based onANN and fuzzy-AHP forheart failure risk prediction", Expert Syst. Appl., vol.68,pp.163-172, Feb.2017.

[23] T. S. Polonsky et al., "Coronary artery calciumscoreandriskclassificationfor coronaryheartdiseaseprediction", JAMA, vol. 303, no. 16, pp. 1610-1616,Apr. 2010.

[24] UCIMachineLearningRepository-HeartDiseaseDataSet,[online]Available:http://archive.ics.uci.edu/ml/datasets/HeartDisease.