LUNGCANCERDETECTIONUSINGSUPPORTVECTORMACHINE

N.Ashok Kumar, M.Vyshali, P.Sravya, J.Dedipya, S. Alekhy
Department of Computer Science and Engineering, Vignan'sNirula Institute of Technology and Science for Women,Pedapalakaluru,AP,India.

Abstract
The Lung Cancer detection is a very challenging task. It generally occurs in both male and female due to uncontrollable growth of cells in lungs. Thus, treatment qualityyisakeystagetoimprovethequalityoflifeofpatients. MachineLearninghas a great influence to healthcare sector because of its highcomputationalcapability for detection of disease with accurate data analysis. Hence, the trustedand automatic classification scheme is essential to prevent the death rate ofhumans. Support Vector Machine (SVM) algorithm in Machine Learning (ML)isusedtodetecttheLungCancerbyclassification.TheLungCancerdetectionisdividedintotwophasessuchastrainingandtestingphases.Inthetrainingphase,preprocessing, feature extraction and SVM classification is performed to make aprediction model more efficient. Finally, Support Vector Machine is used forLung Cancer detection. The cancer dataset is taken from Kaggle. The proposedmethoddetectsthelungcancerwithbestaccuracy.

1. Introduction:

MachineLearningalgorithmshavethe potential to be invested deeply inall the fields of medicine from drugdiscoverytoclinicaldecisionmaking,significantlyalteringthewaymedicineispracticed.Thesuccessofmachine learningalgorithms at computer vision tasksinrecentyearscomesatanopportunetime when the medical records areincreasinglydigitalized.Medical imagesareanintegralpartofapatient'sHERandarecurrentlyanalysed by human radiologists, whoarelimitedbyspeed,fatigue,andexperience. And it also takes years,greatfinancialcosttotrainaqualifiedradiologist. A delayed or erroneousdiagnosismaycauseharmtothepatients. Therefore, it is an ideal formedical image analysis to be carriedoutbyanautomated,accurateand efficientmachinelearningalgorithms. Lung cancer is the mostcommonandaggressivedisease,leadingtoaveryshortlifeexpectancyintheirhighestgrade.The major types of lung cancer arenon-smallcelllungcancerandsmallcell lung cancer. They differ in sizeof the cell. Reports sayaround 84%of lung cancer cases are non-smallcelland13%aresmallcelllungcancer. Diagnosis of lung cancer isdonemedically.Ingeneral,thepatientwillbeawarewhenthecancerisinadvancedstage.Accidentallysometimespatientsmaygottheexistenceofmalignantcellsearlyandifapersonhasacoughandgeneratingsputum,testingthesputum beneath the microscope canperiodicallyrevealthelungcancercells.Whenmovingimagesareconsidered, MRI scan does not workwelllikeincaseoflungswhichexpandsaswebreadthinandcontractswhenwebreadthout.Inthesevere case it can affect brain, bonesor distant sites. The major focus is topredict the risk level of lung cancer.In the past, the outcome for patientsdiagnosedwiththesetumourswasvery poor, with typical survival ratesofjustseveralweeks.Moresophisticateddiagnostictools,inaddition to innovative surgical andradiationapproaches,havehelpedsurvivalratesexpanduptoyearsand also allowed for an improved qualityof life of patients. The classificationtechniqueshelpedtodetectthesurvival possibilities of lung cancervictimsandhelpthephysicianstotakedecisionontheforecastofdisease.

2. Literaturesurvey:

Allresearchershaveaimtodevelopasystem which will predict and detectthe cancer in its early stages. Alsotried

to improve the accuracy of theEarlyPredictionandDetectionsystembypre-processing,featureextractionandclassificationtechniquesofextracteddatabase.

S Vishukumar, K. Patela and PavanShrivastavab - In this paper authorsmostlyfocusonsignificantimprovement in contrast of massesalongwiththesuppressionofbackgroundtissuesisobtainedbytuningtheparametersoftheproposedtransformationfunctioninthespecified range. The manual analysisofthesputumsamplesistimeconsuming, inaccurate and requiresintensivetrainedpersontoavoiddiagnostic errors. The segmentationresults will be used as a base for aComputerAidedDiagnosis(CAD)system for early detection of cancer,whichimprovesthechancesofsurvivalforthepatient.In thispaper, authorsproposedGaborfilterforenhancementofmedicalimages.Itisaverygoodenhancementtoolformedical images.

3. **ExistingSystem:**

RandomForestClassifierismostaccuratealgorithmmainlyusedinhealthcaredomain.Thisalgorithmisappliedondataasettoachieveaccuracy which helps in predictinglung cancer at early stage. Randomforest builds multiple decision tressandmergesthemtogethertogetaccurate and stable prediction. It willhandlethemissingvaluesandmaintaintheaccuracyofalargeproportionofdata.

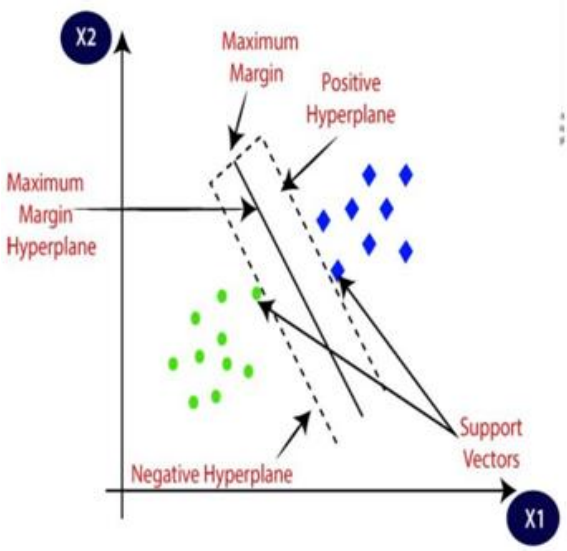
DisadvantagesofExistingSystem:

- RandomForestrequiresmoretimetotrain as it generatesalotof trees.
- Requires more computationalpowerandresources.
- Makesdecisiononmajorityofvoting.
- Low accuracy with morecomplexity.

4. **ProposedSystem:**

Support Vector Machine is powerfulandflexiblesupervisedmachinelearningalgorithmusedfor classification and is popular becauseof its ability to handle multiple andcategorical attributes.

Support Vector Machinealgorithmisimplementedwithkernelthattransforms an input data space intorequired form. Kernel makes SVMmorepowerful,accurateandflexible.InourprojectSVMalgorithmisimplementedbyusingLinearKernel.It can be used when there are a largenumber of features in particular dataset. Training a linear kernel is fasterthanotherkernels.TheobjectiveofaLinear SVC is to fit to the data weprovide and returning a best fit hyperplane that divides or categorizes thedata.



TherearesomeimportantconceptsinSVMlikeHyperplane,supportvectors, Margin. Hyperplane is usedtodivideasetobjectshavingdifferent classes.
Support vectors are the data pointsclosest to hyperplane. Margin is thegapbetweentwolinesonclosestdatapoints of different classes. At first,SVMwillgeneratehyperplanesiterativelywhichsegregatestheclassesinbestway.Thenitwillchoosethebesthyperplanethatseparatestheclassescorrectly.Itusesa technique called the kernel trick totransformthedataandbasedonthesetransformations it finds on optimalboundarybetweenpossibleoutputs.

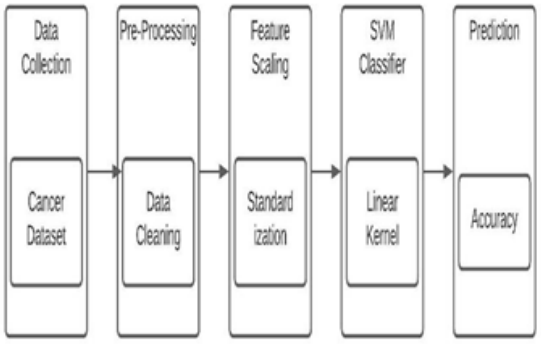


Fig:ProcessFlowDiagram

Theabovefigureshowstheproposedmodel.First,thecancerdatasetiscollectedfromKaggle.AtPre-Processingphase,datacleaningisperformedwhichdeletesunwantedcolumn Patient Id and then FeatureScalingisdone.Itisatechniquewhichwillstandardizetheindependent features present in thedata in a fixed range. It re-scales afeaturevaluesothatithasdistributionwith0meanvalueand variance equals to 1. SVM classifierusestrainingdatasetandgeneratesclassification rules. By using theseclassification rules with test datasetSVMclassifierpredictstheoutput.

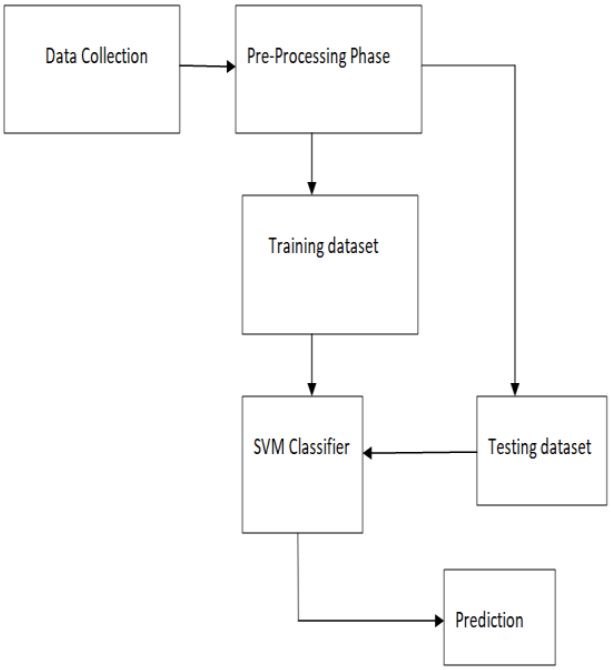


Fig: System Architecture

5. Result:

TheLungCancerdetectionisdividedintotwophasessuchastrainingphaseandtestingphases.Thedatasetiscollectedfromkagglewhichconsistsof1300records.Thetrainingdataset consists of 1000 records andtestingdatasetconsistsof300records.ByusingSupportVectorMachine,theclassificationreportis generatedwithhighaccuracyof98.33%.TheSVMalgorithmclassifiestherecordsintodifferentclasses based on the cancer level ofpatient. From the below output wecansaythat89peoplefallsunderclass-0withlowcancer,96fallsunderclass-1withmediumcancerand 115 falls under class-2 with highcancer.

Accuracy Using SVM: 0.9833333333333333

	precision	recall	f1-score	support
0	0.95	1.00	0.97	89
1	1.00	0.95	0.97	96
2	1.00	1.00	1.00	115
accuracy			0.98	300
macro avg	0.98	0.98	0.98	300
weighted avg	0.98	0.98	0.98	300

6. Conclusion:

Themaingoalistodesignanefficient model for classification oflungcancerwithhighaccuracy,performanceandlowcomplexity.Therandomforest classification isperformedbyusingpre-processingtechniques and the complexity is lowbut the computation power is

highmeanwhileaccuracyisalsolow.Further to improve the accuracy andtoreducethecomputationtime,SupportVectorMachine(SVM) classification algorithm is introducedintheproposedsystem.Also,theclassification report is given with thehighaccuracyof98.33%.Onthebasisofthisinformationthebesttherapy,surgery,radiationandchemo therapy is advised. So, earlydetection is needed to save the livesfromthedisease.

7. FutureScope:

It is a challenging task in machinelearning to construct a specific andcomputationallyefficientclassifierfor the medical applications. Futurescope of this project is how the bigdatasetswillbehavefortheclassificationalgorithmsandtheidentification of particular stage oflungcancerisdoneinnearfuture.Toidentify the particular stage, we needtoperformthedetectiononimagedatasetusingDeepConvolutionalNeural Networks. A few innovativeapplicationsthatspanacrosstraditionalmedicalimageanalysiscategoriesaredescribedhere.Thefirst one is the lung cancer detectionwhichwillclassifythelungCTimages into benign or malicious andthus form the dataset abnormal lungCT can be extracted. Atlast, aftersegmentationlungcancerstageidentification is done which will findout the stages of all nodes that havebeensegmented.Basedonthesizeof nodules, it is classified into differentstages which will help to diagnosepatient's degree of spread of diseasemoreaccuratelythanthepreviousexisting methods.

References

[1] KwetisheJoroDanjuma,"PerformanceEvaluationofMachineLearningAlgorithmsinPost-operativeLifeExpectancyintheLungCancerPatients"DepartmentofComputerScience,ModibboAdamaUniversityofTechnology,Yola,AdamawaState,Nigeria.

[2] Survey of Intelligent Methods forBrainTumorDetection-IJCSIInternationalJournalofComputerScience Issues, Vol. 11, Issue 5, No1,September2014.

[3] ZehraKarhan1,TanerTunç2,"LungCancerDetectionandClassificationwithClassificationAlgorithms"IOSRJournalofComputerEngineering(IOSR-JCE)e-ISSN:2278-0661,p-ISSN:22788727, Volume 18, Issue 6, Ver.III(Nov.-Dec.2016),PP71-77.

[4] Ada, RajneetKaur, " A Study ofDetectionofLungCancerUsingDataMiningClassificationTechniques"InternationalJournalofAdvancedResearchinComputerScienceandSoftwareEngineering,Volume3,Issue3,March2013.

[5] Survey of Intelligent Methods forBrainTumorDetection-IJCSIInternationalJournalofComputerScience Issues, Vol. 11, Issue 5, No1,September2014.

[6] LungCancerdetectionandClassificationbyusingMachineLearning & Multinomial Bayesian-IOSRJournalofElectronicsandCommunication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN:2278-8735.Volume 9, Issue 1, Ver.III (Jan.2014),PP69-75.

[7] K.V.Bawane,A.V.Shinde"Diagnosis Support System for LungCancerDetectionUsingArtificialIntelligence"-International Journal ofInnovativeResearchinComputerandCommunicationEngineering,Vol.6,Issue1,January 2018.

[8] H.R.HAl-Absi,B.B.Samir,K. B.ShabanandS.Sulaiman,"Computer aided diagonosis systembasedonmachinelearningtechniquesforlungcancer",2012InternationalConferenceonComputerandInformationScience(ICCIS), Kuala Lumpeu, 2012, pp.295-300.

[9] SukhjinderKaur"ComparativeStudyReviewonLungCancerDetectionUsingNeuralNetworkandClustering Algorithm", InternationalJournalofAdvancedResearchinElectronicsandCommunication Engineering (IJARECE) Volume 4,Issue2,February2015.

[10] D.Vinitha,Dr.DeepaGupta,andKhare, S., "Exploration of MachineLearning Techniques forCardiovascularDisease",AppliedMedicalInformatics,vol.36,pp.23–32,2015.

[11] SathyanPanicker,J.V,,"LungNoduleClassificationUsingDeepConvNetsonCTImages",9thInternationalConferen

ceonComputing,CommunicationandNetworking Technologies, ICCCNT2018.

[12] Isaac,J.,Harikumar,"LogisticregressionwithinDBMS" ,Proceedingsofthe20162ndInternationalConferenceonContemporary Computing.

[13] PengGuan,DeshengHuang,Miao He, and Baosen Zhou, "LungCancerGeneExpressionDatabaseAnalysisIncorporatingPriorKnowledgewithSupportVectorMachineBasedClassificationMethod",JExpClinCancerRes,28(1): 103,2009.

[14] A. K. Santra, and C. JosephineChristy,"GeneticAlgorithmandConfusionMatrixforDocumentClustering",IJCSIInternationalJournal of Computer Science Issues,Vol.9,Issue1,No.2,2012.