Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 3044 – 3048
Received 05 March 2021; Accepted 01 April 2021.

Lung Cancer Detection Using Image Processing and Convolutional Neural Network

Anubha Gupta[1], V.L. Karthikeya Manda[2], B. Ida Seraphim[3]

[1]Department of Computer Science & Engineering, SRM IST, Chennai, India. E-mail: am5742@srmist.edu.in

[2]Department of Computer Science & Engineering, SRM IST, Chennai, India. E-mail: vm5752@srmist.edu.in

[3]Department of Computer Science & Engineering, SRM IST, Chennai, India. E-mail: idaserab@srmist.edu.in

ABSTRACT

Cancer is known to be one of the most dangeroushealth problems in the world and among it, lung cancer is knowntobethemostseriouscancerwiththesmallestsurvivalrate.Thelungcancerriskpopulationisalsoveryhighascomparedto other deadly diseases, for example, cardiovascular diseases.Therefore, early detection of lung cancer is a must for survival.Nowadays, a lot of research has been done using ConvolutionalNeural Networks in the medical field. Image classification is oneof the methods to detect cancer at early stages. First, the datasetsfor CT Scans are accessed from Kaggle. Images are refined withthe pre-processing method. The image dataset will be trained ontwo different models namely Manual CNN and AlexNet. Further,themodelproducingthehighestaccuracywillbechosenandtheprocessedimageswillbeusedtopredictwhetherthe CTscan image is malignant (cancerous), benign (non-cancerous) ornormal.

KEYWORDS

Lung Cancer Classification, Image Data Aug-mentation,ConvolutionalNeuralNetwork,AlexNet,Convolu-tionalLayer.

Introduction

IN THE world, many individuals are affected by man-madehazards caused due to vehicular emissions or deadly gasesemitted from factories, and many times, personal pleasure.Thisresultsinharmfulsideeffectstothebody,especiallyto the lungs. Lung malignancy is arguably, one of the mostwidespreadandaquotidiantypeofbenignamongmalesandthethird-mostamongfemales.85%ofthetotalcasesaredirectlylinkedtosmokingandconsumptionoftobaccoand also accounts for 20% of the mortality of all cancers.TraditionalwayssuchasexaminingCT-Scansbydoctorshave been used extensively nevertheless it was quite time-consuming.Thisledtotheinnovationofnext-genautonomousdetectionofcancerusingMachineLearningandBigDataoronlyimagerecognitiontechniques.Thetraditionalandpre-contemporaryapproachesfollowedinthedetectionofcancer using various Machine learning algorithms and deeplearning lead our paper to focus on augmented detection oflung cancer using convolution neural networks. The generalmethodology has been proved to be immensely helpful andappreciated across various domains. It has various advantages.The model learns to extract features by itself which helps toyield enviable results. It uses convolution of image and filtersto set invariant features which are passed on to subsequentlayer.Thisprocesscontinuestillitreachesthedesiredoutput.

The layers can be adjusted accordingly to inhibit the modelfrombeinginaccuratethroughhyper-parametertuning.Anarithmeticactionbetweentwooperationswhichresultsinanewfunctioniscalledconvolution.Theyieldoperationisresponsible togiveusa figurativeideaofthe extenttowhich the graphs of the two input functions match each other.The inputs and kernels in a machine learning algorithm arealtogetherintricatearraysofdata.

Related Work

In2019,P.Monkam,S.Qi,H.Ma,W.Gao,Y.YaoandW.Qianpublishedacomprehensiveanalysisofnumerousmethodsbased onthedevelopedConvolutionalneuralnetworkthat can classify and detect nodules in the medical images[1].ThedifferentCNNmodelssuchas:3DCNN,3DU-Net,2DCNN,2DCNNs,SVM,NaıveBayes.Thepapergives brief introduction about CNN and a detailed descriptionofdifferentmedicalimagesourcessuchas:LIDC/IDRI,LUNA16, Kaggle Data Science Bowl. It also states some oftheexistingchallengeswhichaffectsthedetectionprocessincluding the gathering of well-labelled medical dataset withlargenumberofcases.ThepapershowsthatbyapplyingCNN for the detection of nodules as well as their classificationinto malignant and benign produced exceptional results whichmakestheapproachforearlydiagnosismorepromising.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 3044 – 3048
Received 05 March 2021; Accepted 01 April 2021.

In Feb 2020, H. Guo along with some senior memberIEEEproposedamethodcalledKAMP-Netforprediction[2].KAMP-Net makes use of collective framework which consistsof features that are extracted using clinical knowledge andfeatureexposedbytheDual-Streamnetwork(DSN)forimproved prediction. It does not completely rely on automaticfeature extraction. It is shown that KAMP-Net can achievesuperiorperformanceincomparedtoothermethods.Thepaperintroduces Low-Dose CT (LDCT) due to which accuracy ofthe lung cancer detection and its diagnosis increases and thecancer deaths reduced as compared to X-ray. Paper introducesanovelapproachwhichcombinestheCNNandSVMmodeltogetthefinaloutput.

Leukocytes which exist in marrowbone makes up approximately 1% of all erythrocytes. Untrammelled increase of thewhite erythrocyte will lead to the risk of hematologic cancer.InanotherpaperbyD.Kumaret.,proposedamodelthateliminates the likelihood of fallacy during the hand-operatedby implementing deep learning strategy, convolutional neuralnetwork [3]. The structure is trained on cell data (images). Theimages are extracted to draw out the best attribute and thenpre-processed.The archetype with the enhanced Convolutionalneural network (CNN) substructure is trained to predict thetype of malignancy existing in the cells. The overall accuracyrecorded"97.2%"issuperiorthanmachinelearningtechniquesuchasSVM.R.Y.Bhaleraoetal.,proposedasystemth at works with Convolution Neural Network.The proposedpaper shows end to end process of image processing, startingfrom image data extraction from different sources like LIDC,Kaggle, LUNA16 [4]. The image pre-processing was executedin MATLAB. The pre-processed images were sent to the CNNmodel and then trained for prediction. The paper proposedCNN method because it provides a superior production com-pared to machine learning design like Naıve Bayes and SVM.Overall precision of 94.34% was achieved from the proposedarchetype.

M. B. Khumancha and his group of researchers, introducedauniquemethodforthedetectionofLungcancer[5].Themethodology involves working with two datasets, data1 anddata2.Data1whichisaccessedfromLIDC/IDRIusedfordetection of nodule whereas Data2,accessed from Kaggle usedfor Cancer detection. The paper developed two Convolutionalmodule,first CNN module,which detect the nodule present indatasets and used the model for the second procedure whichconsist of detection of malignant symptoms in the dataset.Theaccuracywascalculatedbasedontruecountandfoundtobe 90.78% whereas the precision calculated to be 89.24%.Zheng,S.,Guo,J.,Cui,X.,Veldhuis,R.N.J.,Oudkerk,M.,& van Ooijen, P. M. A. published a paper which focuses onautomatic detection of pulmonary nodule using ConvolutionalNeural Network and also make use of MIP images[6].Thepaper provides a system which will increase the probability ofdetecting the nodule. It introduced MIP- Maximum intensityprojectionimageswhichenhancethenotingofpulmonarynodules in magnetic resonance imaging assessment with computedtomography(CT)scans.Insteadofthenormaldatasetof CT scan, this proposes to introduce MIP images which aresenttotheCNNmodel.TheMIPimageswhicharetaken,their thickness was high and because of which the accuracylevel is increased in the detection of a lung nodule. Thicknessisdirectlyrelatedtoaccuracyofthesystem.

The detection of lung cancer at early stage is crucial formortality,fortheearlydetection,InanotherpaperbyR.Tekadeand K. Rajeswari worked on detecting the malignancy levelU-Net[7].Respective paper mainly focuses on introducing aCAD system for the detection of the nodule. As nowadaysthere is a lot of CT data present, CAD is used to utilize theeffective clinical support it provides. The raw CT dataset isextracted and then pre-processing is executed on the dataset.ForsegmentationROImaskisused.Allthepre-processeddatais then sent to the CAD system, which performs individualnoduledetection.Thefalse-positivereductionisperformedandnoduleclassificationisdone.Aftertheclassification,theresultwillbeeithercancerousornoncancerous.

Yutong Xie et al., introduced MV-KBC model which isolated benign from malignant nodules on chest CT by takingintoconsiderationtheimageofthenoduleonnineviewplanes. An adaptive weighting scheme was applied for noduleheterogeneity which would be able to train the model in amulti-channel manner [8]. In MV-KBC, the archetype trainitself3-Dlungnoduleattributebydecayinga3-Dnodule.The results show that the model is more accurate than currentapproaches on the LIDC-IDRI dataset. When the prediction iscompleted it is shown that model is more accurate than thenormal machine learning model outlook on LIDC dataset. Itprovides knowledge about the MV-KBC archetype to detachcancerous from non-cancerous in CT scans. The total 91%accuracyisshownbyusingtheLIDCdataset.

Implementation

The model will adopt CNN (convolution neural networks)which has proven to yield better, efficient, accurate, and enviable results correlated to other machine learning algorithmslike Naive Bayes, Support Vector Machine (SVM), RandomForest,andothersuchalgorithms.Thefeatureextractionis automatically done by the CNN algorithm based on theinformationprovidedtothemodelwhichinourcaseisaset of CT-Scan images and an output tag. For training, theconvolutionallayersdefinethefeaturesandparameters.

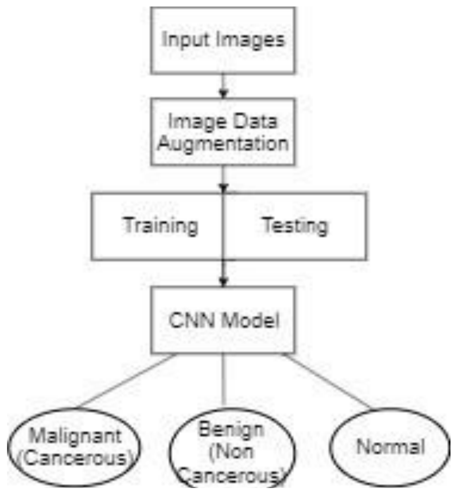

Fig. 1. Flow Chart Diagram

A. Image Data Augmentation

Data augmentation is a technique which is used to increasethe number of sample images in a dataset in order to reduceclass imbalance.This technique is used to increase the numberofsamplesofeachimageinthedatasettopreventthemodelfrombeingundertrained.Thediversityofthetrainingsetcan beincreasedbyapplyingseveraldifferenttransformationtechniquestoourimagedatasetsuchasflipping,rotating,stretchingt heimage.

B. Convolutional Neural Network (Manual CNN)

TheConvolutionNeuralNetwork(CNN)bandscharacterizethe specifications for training. The veracity of ConvolutionNeural Network (CNN) based operations can be upgraded byenhancing the nature of input data and by being contingentupon exceptional training. The Convolution Neural Networkexemplified also plays a dominant aspect in enhancing results.Additional bands suggest exceptional training. Few importantpartsofaConvolutionalNeuralNetworkareasfollows:

- Input Layer- The data which is pushed into the networksiscalledtheinputlayerwhichisaboxlikearrayofpixels
- Convolutional Layer- The main responsibility the convolutional layer has is, extract the highlight from the imagewhich is pushed into the layer from the input layer. It isoneofthemostsupremepartsofCNN.Theconvolutionallayer comprises kernels arranged in series which have toexecute convolution. The initial layers extract lessenedfeatures from the input and as the profundity increases,higherandintricate-levelcharacteristicsareextracted
- Activation Layer- Intermittence in the system is introduced by the activation layer which supports the learningof complex data. The basic function for this model isReLu which endows the pace of how the CNNs are beingtrained by gradient supervision which is constant at allnetworklayers.
- Max-PoolingLayer-Inthislayerdegradationofthedimensionality is done. It allows assumptions for the areawheremax-poolinghastobedone.MPLisappliedonthe initial sector by applying a max filter on the non-intersecting region. Attribute extraction is implementedby coupling various steps which are alike, comprising ofcascading layers specifically Convolutional, Activation,and Max-pooling. The proposed CNN model will predictwhetherthe inputimageisamalignant(cancerous),benign(non-cancerous)ornormal.

C. AlexNet Deep Learning Model

AlexNet is a complex and a successfully pre-trained modelwhich is trained over ImageNet dataset containing 15 millionlabelledimagescategorizedunder21,000classes.AlexNetcontains 5 convolutional layers and 3 fully connected layerswhichtotalsto8totallayers.Itoriginallyperformson2GPUs.However, researchers nowadays tend to use only one GPU fortheimplementationofAlexNet.

- Relu-Nonlinearity: AlexNet makes use of the ReLufunc-tioninsteadofthetanhfunction.ReLufunctionmakesthemodelmoretimeefficient.
- Numerous GPUs: AlexNet allows to put one half of theneuronsinoneGPUandotherhalfinanotherGPUowing to the large dataset. This allows to train largermodelsandcutdownonprocessingtime.Arigorouscomparison will be established between the two models:ManualCNNandAlexNet.Themodelwiththehighestaccuracyandnegligiblelosswillbechosenforthefurth erclassificationoftheinputimageintomalignant,benignornormalEquations.

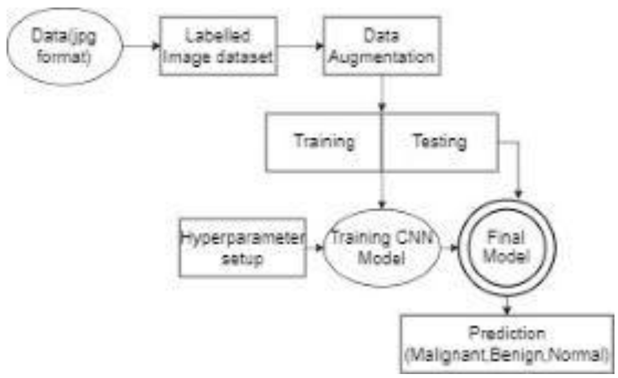

Fig. 2. Architecture Diagram

Result Discussion

ManualCNNrecordedatrainingaccuracyof90.47%onthe given lung cancer image dataset. AlexNet ArchitectureCNNrecordedatrainingaccuracyofonlyabout50.79%.The observations conclude that Manual CNN performed muchbetterthanAlexNetCNNonthegivendataset.

Conclusion

Thepaperaimstoselectthebestmodelforthedetectionof lung cancer with the help of deep learning, namely CNN.The CT images can be extracted from various sources likeKaggle. Image augmentation will be done to provide a betterqualityofdatatoourmodel.CNNmodelistrainedinordertopredictcancer.DeepLearningisacontemporarywingofArtif icialIntelligenceexplorationwhichwillmaneuverimprovedoperationinConvolutionNeuralNetwork-basedsystems. The projected strategy will also take into considerationthe transforming capacity and time law of the cancer detectionmechanism for competency. The proposed system works onCNN model which aims to increase the accuracy and will alsoconsiderthetimedelayandprocessingpoweroftheprocessof detection of cancer for increased efficiency. Accuracy andloss is hence calculated and the best model having highestaccuracyandminimallossischosenforprediction.
In the future, several different knowledgebased models canbe trained on a larger dataset with effective hyperparametertuning and the model with optimal accuracy can be used forprediction.

References

[1] Monkam, P., Qi, S., Ma, H., Gao, W., Yao, Y., & Qian, W. (2019). Detection and classification of pulmonary nodules using convolutional neural networks: a survey. *IEEE Access*, *7*, 78075-78091. http://doi.org/10.1109/ACCESS.2019.2920980

[2] Guo, H., Kruger, U., Wang, G., Kalra, M.K., & Yan, P. (2019). Knowledge-based analysis for mortality

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 3044 – 3048
Received 05 March 2021; Accepted 01 April 2021.

prediction from CT images. *IEEE journal of biomedical and health informatics*, *24*(2), 457-464.http://doi.org/10.1109/JBHI.2019.2946066

[3] Kumar, D., Jain, N., Khurana, A., Mittal, S., Satapathy, S.C., Senkerik, R., & Hemanth, J.D. (2020). Automatic Detection of White Blood Cancer from Bone Marrow Microscopic Images Using Convolutional Neural Networks. *IEEE Access*, *8*, 142521-142531.http://doi.org/10.1109/ACCESS.2020.3012292

[4] Bhalerao, R.Y., Jani, H.P., Gaitonde, R.K., & Raut, V. (2019). A novel approach for detection of lung cancer using digital image processing and convolution neural networks. *In IEEE5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 577-583.

http://doi.org/10.1109/ICACCS.2019.8728348.

[5] Khumancha, M.B., Barai, A., & Rao, C.R. (2019). Lung Cancer Detection from Computed Tomography (CT) Scans using Convolutional Neural Network. *In IEEE10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1-7.

http://doi.org/10.1109/ICCCNT45670.2019.8944824

[6] Zheng, S., Guo, J., Cui, X., Veldhuis, R.N., Oudkerk, M., & Van Ooijen, P.M. (2019). Automatic pulmonary nodule detection in CT scans using convolutional neural networks based on maximum intensity projection. *IEEE transactions on medical imaging*, *39*(3), 797-805.

http://doi.org/ 10.1109/tmi.2019.2935553

[7] Tekade, R., & Rajeswari, K. (2018). Lung cancer detection and classification using deep learning. *In IEEEFourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 1-5. http://doi.org/10.1109/iccubea.2018.8697352

[8] Xie, Y., Xia, Y., Zhang, J., Song, Y., Feng, D., Fulham, M., & Cai, W. (2018). Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest CT. *IEEE transactions on medical imaging*, *38*(4), 991-1004.http://doi.org/10.1109/TMI.2018.2876510

[9] Da Silva, G.L.F., Valente, T.L.A., Silva, A.C., De Paiva, A.C., &Gattass, M. (2018). Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. *Computer methods and programs in biomedicine*, *162*, 109-118.

[10] Qin, C., Yao, D., Shi, Y., & Song, Z. (2018). Computer-aided detection in chest radiography based on artificial intelligence: a survey. *Biomedical engineering online*, *17*(1), 1-23.

[11] Yamashita, R., Nishio, M., Do, R.K.G., &Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, *9*(4), 611-629.

[12] Ronneberger, O., Fischer, P., &Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical image computing and computer-assisted intervention*, 234-241.