ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 6 | Issue: 3 | March 2020 || Journal DOI: 10.36713/epra2013 || SJIF Impact Factor: 5.614||ISI Value: 1.188

OPTIMISING CLASSIFICATION TECHNIQUES FOR LUNG CANCER DETECTION ON CT IMAGES

T. Maria Patricia Peeris[1]
PG Student
Department of Computer Science and Engineering
Francis Xavier Engineering College
Tirunelveli, India

Prof. P. Brundha[2]
Associate Professor, Head of Department
Department of Computer Science and Engineering
Francis Xavier Engineering College
Tirunelveli, India

Article DOI: https://doi.org/10.36713/epra4214

ABSTRACT

Lungs are the most crucial organs in a human body. Since the cancer detection began, lung cancer has been the most common terminal disease amongst all type of cancers. The contribution of deep learning, especially the convolution neural networks has widely reduced the mortality rates resulting from lung cancer. The classification of Computed Tomography (CT) images has enhanced the early diagnosis of lung cancer that has enabled victims to undergo treatment at an early stage. The resolution of the CT images have been variedly used for the accuracy of the model. Besides, the detection of lumps or anomalies in the images has greatly supported early diagnosis. Classification plays a vital role in the deep learning models to sort out the input images as positive and negative based on the attribute of the model built. However, the generalisation of classifiers has reduced the accuracy of the corresponding models built. To increase the accuracy and efficiency of the deep learning model, an optimised classification technique is used to predict lung cancer from the CT images. The purpose of optimisation here will enable the model to adapt stipulated feature extraction process according to the input images fed into the network. The model will be trained for predicting purpose given any resolution of the images.

KEYWORDS: *Lung cancer, CT images, Classification techniques, Optimised Classification, Prediction*

I.INTRODUCTION

Lungs are the most crucial organs in a human body. Since the beginning of cancer, lung cancer has been the most deadliest cancer with high mortality rates around the world. Recent researches has improved the diagnosis of lung cancer and in-fact facilitated early diagnosis compared to the earlier predictive models. As a result, death rates have reduced to a great extent. Deep learning has contributed to early diagnosis of lung cancer largely and is also more accurate compared to the other machine learning models. One of the simplest and easiest deep learning technique is classification which allows the model to segregate input images. Classifiers are the parameters based on which the given input is classified as normal or abnormal. Classification of images containing cancerous cells promotes the diagnosis of cancer. Most commonly CT images are used as they are cost-efficient and easy to examine. This paper will optimise classification techniques on CT images to detect lung cancer.

In [1], Deep learning (DL) frameworks are used as they extract features deep down the hidden network with variable factors acquired during training. Most medical imaging models use Convolution Neural Networks (CNN) for prediction. CNNs are composed of various layers that enhance training of the models for accuracy through extraction, segmentation and many more. The purpose of using a CT image is the ability to isolate the infected area and categorise the same after appropriate training.

Despite the enhancement in diagnosis, the model often fall short of optimised classification techniques which can increase the efficiency of the existing model.

Firstly, this paper will highlight the existing classification techniques for lung cancer detection and identify the literature gap. Secondly, the paper will optimise the classification technique for early lung cancer detection. Finally, the paper will bring forth the future scope of the research.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 6 | Issue: 3 | March 2020 || Journal DOI: 10.36713/epra2013 || SJIF Impact Factor: 5.614||ISI Value: 1.188

II. LITERATURE REVIEW

Classification techniques can be classified into five different types namely Naive Bayes, Support Vector Machine (SVM), K-Nearest Neighbour (KNN), Decision Tree and Random Forest. Each technique is discussed in detail.

A. NAIVE BAYES

The naive bayes is a probability based technique that takes an average of the frequency of the independent variables in the neural network [5]. This algorithm requires less training compared to the other classification techniques and considers the maximum of two events dependent on each other.

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

where x and y are independent events that can occur simultaneously.

In simpler terms, x and y are two different features that can be found on the given image either separately or together. Therefore, the maximal of occurrences is considered as the prediction value of the model. This technique requires very few dataset for the model to be trained. However, the accuracy is affected if there is a zero probability. The occurrence of zero probability or no probability tricks the model to produce false values often. Therefore, the reliability of the classification technique decreases.

B. SUPPORT VECTOR MACHINE

In [3], SVMs are point-based classification techniques where the values are plotted like a graph and a pattern is recognised by the model. The test data are then categorised into the respective spaces as a result for classification. Therefore, this classification technique is useful for categorising the given input and the prediction is plotted against a scale learnt by the model. However, this classification cannot be validated and the prediction may be monotonous. The quality of a product can be predicted using this classification technique whilst the prediction of cancerous cells may be difficult.

C. K-NEAREST NEIGHBOUR

The purpose of using a KNN algorithm is to find the odd attributes among a given set in a network. The nearest neighbours are trained with respective weights for training purposes. However, in [6] the algorithm only classifies the given input but cannot used for learning. This technique only categorises similar features within the network. They do not have any parameters and therefore cannot learn from the training. Though it is easy to implement this technique, the computational cost is comparatively high as the training becomes efficient only in the presence of large set of data.

D. DECISION TREE

The decision tree technique is a threshold-based method where the input data is partitioned according to the features and a differentiator is used for categorisation purpose [4]. This techniques keeps sorting the given data until the final feature is extracted and this process occurs repetitively. This classification technique is a very simple process but a small change in the data inference may lead to an entirely variant decision tree. Therefore, other classification techniques are preferred over this method when huge sets of data is used and the training given is minimal.

E. RANDOM FOREST

In [2], random forest technique uses the concept of the decision tree implementing multiple trees to extract model-based parametric features. This technique is more reliable compared to the decision tree method. However, the repetitive process make the model more complex and time consuming. The categorisation of samples and sub-set of samples into classes requires ample training and datasets which often becomes less available to the researcher.

III. LITERATURE GAP

The recent advancement in Artificial Intelligence (AI) has contributed largely to the computational literature. The evolution of classification techniques has adversely impacts the health industry in many ways. There has been an increase in the formulation of classification techniques. Each technique has its own advantages and disadvantages. As per the discussion above, naive bayes and KNN algorithms seem to be more reliable compared to the other classification techniques as they are more accurate and capable of learning. The other three classification techniques SVM, Decision Tree and Random Forest methods classify input images but do not learn from the training process. In the case of lung cancer detection, optimised classification methods will enhance the diagnosis and the accuracy of prediction. The literature fails to bridge the short comings of the techniques and they are generalised to all types of model. Here, the need for optimising classification techniques become important and more focus has to be laid on customising classification algorithms for the corresponding objective of the model. This paper will propose a customised classification technique for lung cancer detection.

IV. PROPOSED WORK

This paper proposes an optimised classification technique for lung cancer detection. Due to the learning capabilities of Naive Bayes and KNN algorithm, this paper will integrate both the classification technique into one for lung cancer detection. Convolution Neural Network is used due to the ability of the network to learn versatile datasets. The proposed process is,

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 6 | Issue: 3 | March 2020 || Journal DOI: 10.36713/epra2013 || SJIF Impact Factor: 5.614||ISI Value: 1.188

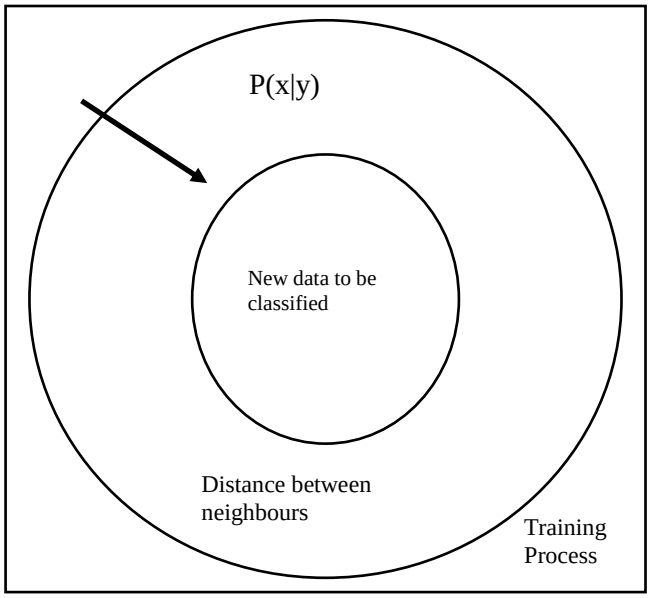

Fig 1: The integration of Naive Bayes and kNN algorithm is done for optimising lung cancer detection. The arrow depicts the direction of the process.

The network will initially classify the given input using the Naive Bayes algorithm in the first set of hidden layers. In this model, a minimum of 2 hidden layers at the beginning will help extract features from the nearest neighbours. This will help the model isolate cancerous cells when fed into the network. The Naive Bayes algorithm helps in identifying the probability of cancerous cells within the mentioned area of the CT images when fed into the network. The first two layers will be able to classify the anomalies as big and small sizes respectively. Here,

$$P(b.s) = \frac{P(b) + P(s)}{P(s)}$$

Where,
P(b) is the probability of big size cancerous cells
P(s) is the probability of small size cancerous cells

The probability of small sizes is given more importance as the main objective of this model is to detect cancer at the earliest stage possible. The kNN algorithm will then be used to classify the nearest neighbours of similar features. This way the network will be able to first classify the image and then classify the neighbours with similar abnormalities in the same image. Furthermore, once the kNN algorithm is implemented, the Naive Bayes algorithm is again executed. The repetitive process allows the quality As the training progresses, the model will be able to predict cancerous cells through Naive Bayes classification and then locate the infected neighbours through kNN algorithm. Therefore, the proposed classification technique can be termed as Naive - k Nearest Neighbour algorithm (N-kNN) classification technique. The concept of integration is depicted in the name itself. This algorithm will have both the ability to learn and classify input images. Therefore, validation may be implemented easily and the network can be updated at any point throughout the network building process.

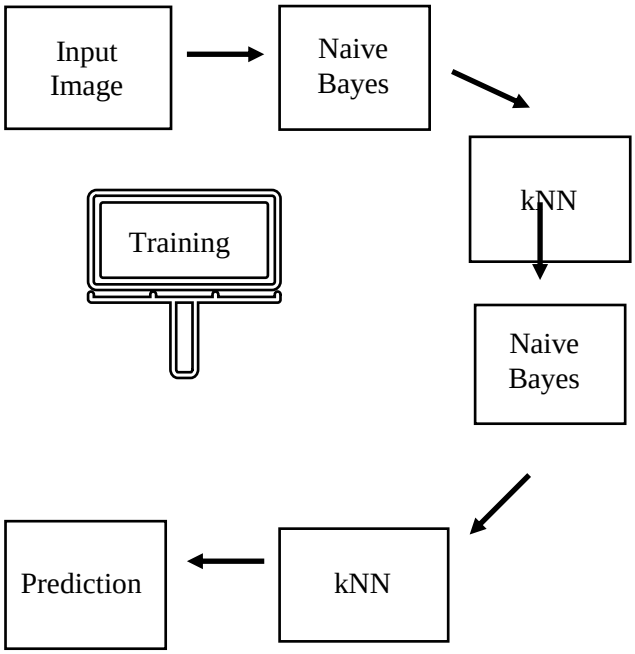

Fig 2: The CNN training cycle of the optimised model which involves Naive Bayes and kNN classification techniques.

V. RESULTS AND CONCLUSION

This paper aims at optimising classification technique for lung cancer detection to improve the efficiency and accuracy of the prediction. Here, we integrate kNN algorithm and Naive Bayes algorithm as N-kNN technique to optimise lung cancer detection using CT images. As a result, the accuracy of the prediction is high and the quality is also improved. Firstly, the integration of two classification algorithms will increase the efficiency of the model by first classifying and then predicting cancerous cells. This way the model will be able to predict cancer from the CT images extracting features based on the parameters with which the model was trained initially. Secondly, the use of two algorithms will be able to assure the prediction of the model to be more than satisfactory as they will study the probability of both the big and small sized anomalies and thereby chance of missing any infected area may not be possible. Finally, the ability of the network to learn will be able to detect cancer as training and testing phases takes place. In the learning space, validation of the network will also be possible.

ISSN (Online): 2455-3662
EPRA International Journal of Multidisciplinary Research (IJMR) - Peer Reviewed Journal
Volume: 6 | Issue: 3 | March 2020 || Journal DOI: 10.36713/epra2013 || SJIF Impact Factor: 5.614||ISI Value: 1.188

REFERENCES

1. *D. Riquelme and M. Akhloufi, "Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans", AI, vol. 1, no. 1, pp. 28-67, 2020. Available: 10.3390/ai1010003 [Accessed 17 March 2020].*
2. *K. Roy et al., "A Comparative study of Lung Cancer detection using supervised neural network", 2019 International Conference on Opto-Electronics and Applied Optics (Optronix), 2019. Available: https://ieeexplore.ieee.org/abstract/document/8862326. [Accessed 18 March 2020].*
3. *M. Munir Prince, A. Hasan and F. Shah, "An Efficient Ensemble Method for Cancer Detection", 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), 2019. Available: https://www.researchgate.net/profile/Faisal_Shah3/publication 338071772_An_Efficient_Ensemble_Method_for_Cancer_Detection/links/5e004d064585159aa492c3dc/An-Efficient-Ensemble-Method-for-Cancer-Detection.pdf. [Accessed 18 March 2020].*
4. *M. Prabukumar, L. Agilandeeswari and K. Ganesan, "An intelligent lung cancer diagnosis system using cuckoo search optimization and support vector machine classifier", Journal of Ambient Intelligence and Humanized Computing, vol. 10, no. 1, pp. 267-293, 2017. Available: https://www.researchgate.net/profile/Manoharan_Prabukumar/publication/321885115_An_intelligent_lung_cancer_diagnosis_system_using_cuckoo_search_optimization_and_support_vector_machine_classifier/links/5b5af068a6fdccf0b2f99d27/An-intelligent-lung-cancer-diagnosis-system-using-cuckoo-search-optimization-and-support-vector-machine-classifier.pdf. [Accessed 18 March 2020].*
5. *M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification", International Journal of Intelligent Systems and Applications in Engineering, pp. 88-91, 2019. [Accessed 18 March 2020].*
6. *M. Toğaçar, B. Ergen and Z. Cömert, "Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks", Biocybernetics and Biomedical Engineering, vol. 40, no. 1, pp. 23-39, 2020. Available: https://www.sciencedirect.com/science/article/pii/S0208521619304759. [Accessed 19 March 2020].*