International Journal of Engineering and Management Research e-ISSN: 2250-0758 | p-ISSN: 2394-6962
Volume-12, Issue-6, (December 2022)
www.ijemr.net https://doi.org/10.31033/ijemr.12.6.32

Early Detection and Prevention of Lungs Cancer using Artificial Neural Network

Supriya Narad[1] and Dr. K. T. V. Reddy[2]
[1]Computer Science & Engineering (AIDS), Faculty of Engineering & Technology, DMIMS (DU), Sawangi (Meghe), Wardha (MH), INDIA
[2]Computer Science & Engineering (AIDS), Faculty of Engineering & Technology, DMIMS (DU), Sawangi (Meghe), Wardha (MH), INDIA

[1]Corresponding Author: naradsupriya@gmail.com

ABSTRACT

Out of the dangerous diseases, cancer being one of the cause of death and it can be avoided if correctly detected in the early stage. The possibilities of survival will be increased if predicted and cured at early stage. To predict accurately and to provide best diagnosis, many mechanisms are developed in the field of Artificial Intelligence and machine learning. This paper provides a systematic review of different machine learning algorithms like Artificial Neural Network (ANN), Decision Trees (DT), Support Vector Machine (SVM), Random Forest (RF), Voting Classifier and Bayesian Network (BN). Survey also shows that ANN and SVM are preferred by researchers to develop the predictive models.

***Keywords*--** Cancer Prediction, Prevention, Machine Learning, Artificial Neural Network, Support Vector Machine, Decision Trees

I. INTRODUCTION

Over the past years, a research has been performed related to cancer prediction. Cancer can damages various organs like brain, skin, lungs, blood, bladder, liver, etc. Many scientists had applied various methods like evaluating and reporting in the early stage, which helps to find varieties of cancerous cells before they cause any symptoms. Many strategies are implemented for predicting cancer at early stage. For predicting and curing cancer, clinical physicians try hard using available technologies like CT scan, X-ray, etc. As machine learning technology evolved, it has become a useful tool which helps medical researchers also. Considering the Machine Learning (ML) techniques, this paper presents a review of researches that uses algorithms related to ML. These methods are used for detection and classification of tumors via x-ray and CRT images. Joseph A. Cruz [1] presents importance of ML. Machine learning technology is a part of artificial intelligence that provides statistical tools, probabilistic tools and optimization tools to learn from previous examples and to use prior training to classify new data, identify new patterns or predicts trends. Machine learning technology is used to analyze and interpret data for various applications. Machine learning methods employ Boolean logic, absolute conditional statements, conditional probabilities and optimization strategies to model data. These methods are similar with the approaches humans typically use to learn and classify.

Researcher classified techniques of Machine Learning into three categories viz. supervised learning, Un-supervised learning and Reinforcement learning [2]. Supervised learning generally resembles regression problems, such as forecasting, growth prediction and prediction using Linear Regression or Random Forest. Supervised learning also solves classification related problems such as voice recognition, digit recognition, diagnostics and fraud detection. These classification problems are implemented using algorithms, such as Support Vector Machines, Random Forest, k- Nearest Neighbor and others that are utilized in many areas. In supervised learning, two levels are operated as the training phase and the testing phase. Known labels should be available in the data sets used for the training process. Unsupervised learning deals with topics related to the reduction in dimensions used for visualization of large data, feature elaboration and which discover the secret structure. As compared to supervised learning, labels are not available in this method. Algorithms work to recognize patterns while testing data and it predicts the future values.

Another type Reinforcement Learning is based on a collection of parameters that the algorithms attempt to predict the output for a problem. Then the output becomes an input parameter, and then a new output is found once the optimum output is found. Deep Learning and Artificial Neural Networks (ANN) are used in this method.

II. LITERATURE REVIEW

Some techniques are widely used in cancer detection like Decision trees, Neural Network and Support Vector Machines [3]. Researcher presents a method of classification of cancer cells into Benign and Malignant by using Deep Learning Convolutional Neural Network. Author had used Deep Learning Library Keras. Datasets of Invasive Ductal Carcinoma

(IDC) were used for training and testing the model. This model helps to classify the cancerous cells into benign cell and malignant cell. IDC results to classify the cancerous cells on an average scale.

CT Scanning method has proved itself to be effective for diagnosing lungs cancer at physician's level [4]. Recently, deep learning and Convolutional Neural Networks (CNN) has been employed to automatically detect and classify pulmonary nodules which are found in medical diagnosis images. CNN can be used with RNN to get more accurate results.

Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) dataset used which consist of 1018 cases. In effect, there were 1,010 different CT scans and eight cases that were inadvertently reproduced while gathering the CT scans. Other datasets used were LUNA (Lung Nodule Analysis) 16 Datasets, NLST (National Lung Screening Trial) Datasets, Kaggle dataset and ELCAP (Early Lung Cancer Action Program) dataset. The programming language utilized for the implementation of deep learning models is python. Besides, Matlab is a high level programming language for numerical analysis and scientific works. Other platforms based on python language include Keras, Caffe, Chainer, ThensorFlow, Torch, etc.

Konstantina Kourou et. al., worked on Machine Learning Applications [5]. In this research, some techniques are used in cancer detection like ANN, Bayesian Network, Support Vector Machine and Decision Tree. This paper presents survey of ML approaches used in the modeling of Cancer progression. The predictive models used are based on supervised ML techniques. A survey based on ML applications in cancer also presented like prediction of cancer susceptibility, prediction of cancer recurrence, prediction of cancer survival. Electronic databases used are PubMed and Scopus. Among the most common applied ML algorithms, SVM and ANN classifiers were widely used. ANNs have been used for 30 years and SVMs are used recently.

Computer Aided Detection (CAD) system are already used to detect lungs cancer [6]. Dataset used is Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI). Convolutional Neural Network (CNN or ConvNet) is used for feature extraction and classification of objects in the image. CNN is a stack of different layers. A 3D- multipath VGG architecture is proposed. It is a structure of convolutional and pooling layers. U-net architecture is adapted for segmentation of lung nodes from lungs CT scan images obtained and 3D-multipath VGG architecture is used for classifying lung nodules and predicts malignancy level. This approach helps to predict the recurrence of cancer.

Yogesh Kumar, et. al., presents [7] AI based learning approaches for cancer prediction using convolutional machine and deep learning based classification is used in this paper. Survey is based upon the parameters like prediction rate, accuracy, sensitivity, specificity, dice score and detection rate. Also other parameters are area undercover, precision recall and F1 score. Clinical practices applied in the medical sector for cancer prediction are also stated like screening, chemotherapy, etc.

Anum Masood, et. al., [8] proposed system is used for detecting lung nodule and classification is done. It also serves as an image classifier method for feature extraction along with novel multilayer fusion Region Proposal Network (mLRPN). Images of 3D CT Scan are used as datasets combined with 3 other neighboring CT scan reports for each axis direction. Augmentation is recommended for reducing data over-fitting. The system is then integrated with cloud computing (Infrastructure as a Service (IaaS) by providing Virtual Machines, and Software as a Service (SaaS) by mRFCN model). Four modules are used in cloud-based mRFCN as data submission, online medical reports submission by radiologists, accessing CAD results, and physician's feedback. Computer Aided Detection (CAD) used to detect and classify pulmonary nodules. Also, Deep Convolutional Neural Network based model is proposed for early detection of lung cancer which uses 3D Spatial information. The proposed system indicates presence of nodule, location as well as outlines the possible shape of the detected nodule as benign and malignant. This model has attained the sensitivity as 98.1 % and classification accuracy as 97.91 %.

Onur Ozdemir, et. al., [9] proposed probabilistic Deep Learning System and it introduces a new Computer Aided Detection (CAD) system and diagnosis system for lungs cancer screening with low dose CT scans which produces meaningful probability assessments. Database from LUNA 16 (Lung Nodule Analysis 2016) and Kaggle Data Science used for Analysis. CADe and CADx model is introduced to detect and diagnose lung cancer using low-dose CT-scans.

Wasudeo Rahane, et. al., [10] worked on Lung Cancer Detection. Researcher used Image Processing and Machine Learning for HealthCare. A software model is proposed which detects cancer. In this model, user need to upload CT scan reports on website and query is generated at the client side and send it to the server side. For image acquisition, CT scan images are taken from ELCAP public lung image database which contains near about 200 lung images of cancer and non-cancer patient CT scan reports. Blood sample is also one of the parameter. Input image is converted into gray scale image, noise is reduced from the image, mechanism of Binarization and segmentation is implemented then feature extraction is performed, finally SVM is used to predict the pattern. Segmentation process helps to reduce unnecessary data in the image and to locate object, lines, curves, etc. in the images. SVM classifier classifies the positive samples and

negative samples in case of lung cancer images. Research found that SVM is less time consuming as compared to decision trees, KNN, ANN & Random Forests.

As per comparison and analysis of ANN and SVM, output results of SVM model are more superior and accurate to ANN model in comparison with generalization and precision. A data driven prediction model combining the discrete wavelet transform, perform preprocessing on it and then using support vector machine it can identify ground water depth forecasting. Results of the study prove that SVM performs more reliable and provide accuracy more than ANN. Implemented prediction models have lower mean error SVM in model calibration. The mean error of SVM is slightly lower than ANN. It indicates that the SVM models are superior for generalization.

Table 1: Comparison of Research Methods

Sr. No.	Researcher	Year of Research	Method/ ML ALGO	Type of Cancer	Dataset
1.	Maheep Singh, Krishan Kumar [3]	2021	Deep Learning Convolutional Neural Network	Lung Cancer	Invasive Ductal Carcinoma (IDC)
2.	Patrice Monkam, et. al. [4]	June 2021	Deep Learning Models	Lung Cancer	LIDC, IDRI, LUNA 16
3.	Weimin Huanh, et. al. [15]	2015, IEEE	Convolutional Neural Networks	Brain Tumor	MNIST dataset
4.	Ruchita Tekade and Prof. Dr. K. Rajeshwari [6]	2018, IEEE	Convolutional Neural Networks	Lung Cancer	LIDC-IDRI
5.	Yogesh Kumar, et. al [7]	Elsevier Journal, 2022	Convolutional Machine Learning and Deep Learning	Various type	Elsevier's biomedical bibliographic database EMBASE
6.	Anum Masood, et. al. [8]	2020, IEEE	Enhanced multidimensional Region-based Fully Convolutional Network (mRFCN)	Lung Cancer	CT scans Database
7.	Onur Ozdemir, et. al. [9]	2019, IEEE	Deep Learning Models	Lung Cancer	LUNA 16
8.	Wasudeo Rahane, et. al. [10]	2018, IEEE	KNN, SVM	Lung Cancer	ELCAP public lung image database
9.	Mukherjee, S., & Bohra	2020, IEEE	Convolutional Neural Networks	Lung Cancer	DICOM dataset
10.	Song, C. [17]	2022, PubMed	KNN Algorithm	Liver Cancer	UCI database
11.	Thallam, C., Peruboyina, A., Raju, S. S. T., & Sampath, N [14]	2020, IEEE	KNN, ANN, SVM, RF	Lung Cancer	data.world datasets

ANN is proved to be more effective in extracting the hidden non-linear input-output relationships than traditional algorithms. SVM is a machine learning theory based algorithm. SVM does not have a theory based predetermined structure, while the training samples are judged by their contributions. M. Divyavani, Kalpana Govindswami [12], have worked on Breast Cancer Dataset. Researcher have used SVM and ANN combined with feature selection and both models were tested on popular Kaggle Wisconsin Diagnosis Breast Cancer Dataset. A comparison study of SVM and ANN showed that ANN has got improved accuracy than SVM and classification accuracy of 99% and 98% respectively.

Chaohong Song and Xinran Li, worked on Cost-Sensitive KNN Algorithm and performed Cancer Prediction algorithm Based on Entropy Analysis [17]. They have used database of the patients with liver cancer and patients with ovarian cancer. Researcher has implemented KNN algorithm for improving the accuracy of cancer recognition. Research was carried out by following two aspects determined by feature extraction and algorithm. In feature extraction, the information provided by the entropy value was considered, and in algorithm, cost-sensitive learning technology was used to improve the original KNN algorithm for imbalanced

sample data.

Thallam, C., Peruboyina, A., Raju, S. S. T., & Sampath, N. [14], worked on Early Stage Lung Cancer Prediction Using Various Machine Learning Techniques. Raw data is transformed into training data which is capable of decision making in the system. This study has collected data from data.world and got 25 features. Different algorithms like KNN, ANN, SVM, Decision Tree, are used for training the dataset and result was predicted.

III. METHODOLOGY

Different machine learning algorithms are Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tress (DT), Random Forest (RF), K-Nearest Neighbor (KNN), Voting Classifier and Bayesian Network (BN).

1. Artificial Neural Network:

An artificial neural network (ANN) defined as a model of reasoning based on the human brain [12]. Over the past few decades, ANNs have been implemented increasingly by more and more researchers, and become an active research area. A typical ANN model is made up of a hierarchy of layers as input, hidden and output layers shown in Figure 1. Layers consist of interconnected neurons which contain an activation function for nonlinear transformation.

Artificial Neural Networks (ANN) consists of neurons and weights and it is assigned to inter neuron connections and it helps in storing the acquired knowledge [13]. ANN is a nonlinear and parallel adaptive system that is used to

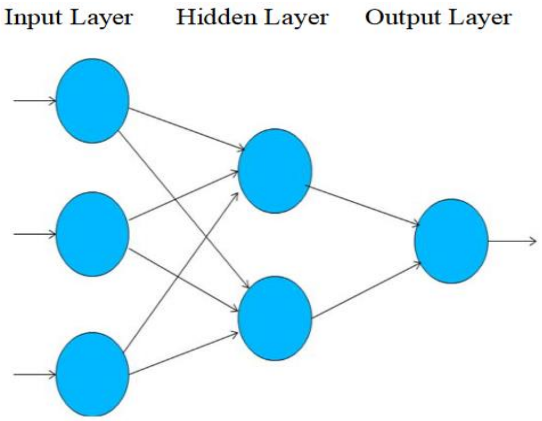

Figure 1: Artificial Neural Network

model relationships between inputs and outputs. The output is decided by I/O characteristics while the overall working of ANN is determined by its structure and the training algorithm implemented on NN. Advantages of ANN are adaptive interaction between different elements, self-organization, real time operation, parallel computations, and Fault Tolerance. ANN helps inhandling critical problems and is used in robotics, pattern recognition, medicine, manufacturing, and optimization; signal processing, system modeling and identification.

2. Support Vector Machine:

It is a supervised learning model that can be used for both classification as well as regression analysis. SVM helps in reducing the rate of misclassification and thus gives good results. It plays out a few emphases to locate an ideal hyper plane. It isolates the various classes of a dataset in a multidimensional space. The hyperplane is framed by the support vectors that are nearest to the edge. If it is framed with the greatest edge, then it is supposed to be the ideal hyperplane, that is, maximum marginal hyperplane (MMH). The output is then anticipated by plotting the test set data points on the obtained hyperplane. The workflow of the algorithm involves an SVC classifier to train the model and the model is fitted to the training set and prediction is done using the test set.

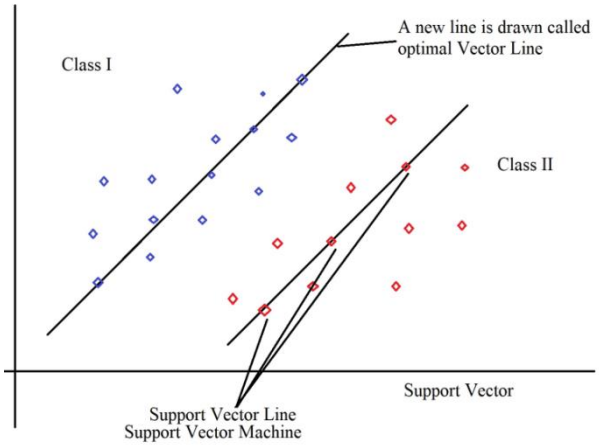

Figure 2: Support Vector Machine

As shown in Figure 2, a straight line will cover maximum data points and points through which line is drawn is called as support vector.

3. K- Nearest Neighbor (KNN)

It is a non- parametric, supervised learning classifier. The model discovers k closest neighbors of a specific component vector out of the absolute N training vectors accessible. It calculates the difference between the observed data points and actual data points using a similarity measure. Euclidean distance is the similarity measure, that is, the metric used is 'minkowski'.

After computing the similarity measure between the new case and the actual data, it classifies a new case by majority voting of the class labels of the first k closest neighbors, that is, the first k similarity values computed. The workflow of the algorithm involves a K Neighbors Classifier to train the model and the model is fitted to the training set and prediction is done using the test set. For a good practice the estimation of k should not be a multiple of the number of classes. It is likewise critical to take

note of that the estimation of k ought to be odd while considering a 2 class issue to stay away from a tie between the classes. [17] For a fixed K, when the sample distribution is imbalanced, K-NN algorithm may be more precisely inclined to judge the test samples into the category with more samples, thus reducing the accuracy of classification.

4. Decision Tree:

Data Mining is the huge domain to study about extracting patterns, classifying huge and uncertain data, where data is of the heterogeneous forms like text, audio, video, etc. Sometimes the data provided may be unclean, incomplete, noisy, damaged, etc.

Classification and data pre-processing is one of the techniques to handle with this type of data. Decision tree can give the approximate solution to the data which is used in data mining and machine learning [18]. There are different algorithms which can be used like neural network, k-nearest neighbor, etc. From these algorithms Decision trees is one of the most useful and powerful algorithm in data mining. It is able to handle the number of input data like numerical and alphabetical and this is the benefit of it. This algorithm can process the data which contains some missing values and errors. This type of contents may vary in number of platform and various packages of datasets. By using the decision rules, decision tree are used to extract a data from large amount of available datasets. Decision tree helps simply to classify the data which can be easily stored and further it can be easily classified again. The following example shows working of simple decision tree algorithm.

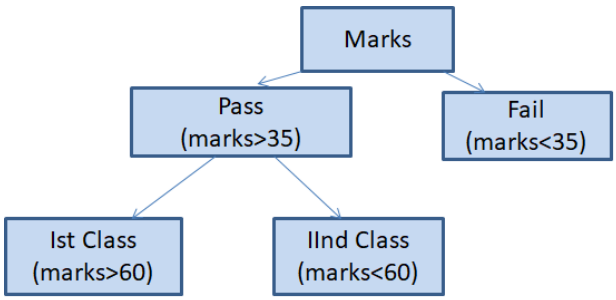

Figure 3: Decision Tree

IV. CONCLUSION

This study has analyzed various Machine Learning tools for predicting various cancer types. Methods like KNN, ANN, SVM, Decision Tree are discussed in the study. It is found that SVM is more accurate. The focus of this article is to review, analyze and categorize methodologies of different types of cancer and uncover existing limitations. The review has presented four types of cancers like lung cancer, breast cancer, brain tumor, liver cancer. The primary intention of this research is to present an intelligent background to new researchers who wish to begin their research activity in this field.

REFERENCES

[1] Joseph A. Cruz & David S. Wishart. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*.

[2] Dakhaz Abdullah & Nawzat Ahmed. (2021). A review of most recent Lung cancer detection techniques using machine learning. *Research Gate*.

[3] Maheep Singh & Krishan Kumar. (2021). *Cancer detection using convolutional neural network*.

[4] Patrice Monkam, et. al. (2019). Detection and classification of pulmonary nodules using convolutional neural networks: A survey. *IEEE*.

[5] Konstantina Kourou et. al. (2015). Machine Learning Applications in cancer prognosis and prediction. *Elsevier: Computational and Structural Biotechnology Journal, 13*(2015), 8–17.

[6] Ruchita Tekade & Prof. Dr. K. Rajeshwari. (2018). Lungs cancer detection and classification using deep learning. *IEEE*.

[7] Yogesh Kumar, et. al. (2022). A systematic review of artificial intelligence techniques in cancer prediction and diagnosis. *Archives of Computational Methods in Engineering, 29*, 2043–2070.

[8] Anum Masood, et. al. (2020). Automated decision support system for lung cancer detection and classification via Enhanced RFCN with multilayer fusion RPN. *IEEE Transactions on Industrial Informatics*.

[9] Onur Ozdemir, et. al. (2019). A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose ct scans. *IEEE Transactions on Medical Imaging*. DOI: 10.1109/TMI.2019.2947595.

[10] Wasudeo Rahane, et. al. (2018). Lung cancer detection using image processing and machine learning healthcare. *IEEE*.

[11] Mehmood ul Hasan, et. al. (2019). Comparative analysis of svm, ann and cnn for classifying vegetation species using hyperspectral thermal infrared data. *ISPRS Remote Sensing and Spatial Information Sciences, XLII-2/W13*.

[12] M. Divyavani & Kalpana Govindswami. (2020). An analysis on SVM and ANN using breast cancer dataset. *Aegaeum Journal, 8*(12).

[13] Narad, S. K., & Chavan, P. V. (2015). Neural network based group authentication using (n, n) secret sharing scheme. *International Conference on Advances in Computer Engineering and Applications*. DOI: 10.1109/icacea.2015.7164739.

[14] Thallam, C., Peruboyina, A., Raju, S. S. T. & Sampath, N. (2020). Early stage lung cancer prediction using various machine learning

techniques. *4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. DOI: 10.1109/iceca49313.2020.929757.

[15] Yuehao Pan & Weimin Huang, et. al. (2015). Brain tumor grading based on neural networks and convolutional neural networks. *IEEE*.

[16] Mukherjee, S. & Bohra, S. U. (2020). Lung cancer disease diagnosis using machine learning approach. *3rd International Conference on Intelligent Sustainable Systems (ICISS)*. DOI: 10.1109/iciss49785.2020.931590.

[17] Song, C. & Li, X. (2022). Cost sensitive KNN algorithm for cancer prediction based on entropy analysis. *Entropy*, *24*, 253. Available at: https://doi.org/10.3390/e24020253.

[18] Somvanshi, M. & Chavan, P. (2016). A review of machine learning techniques using decision tree and support vector machine. *IEEE Sponsored, International Conference on Computing Communication Control and Automation (ICCUBEA)*. DOI: 10.1109/iccubea.2016.7860040.

[19] R. A. Welikala et al. (2020). Automated detection and classification of oral lesions using deep learning for early detection of oral cancer. In: *IEEE Access*, *8*, pp. 132677-132693. DOI: 10.1109/ACCESS.2020.3010180.

[20] E. P. Papageorgiou, B. E. Boser & M. Anwar. (2020). Chip-scale angle-selective imager for in vivo microscopic cancer detection. In: *IEEE Transactions on Biomedical Circuits and Systems*, *14*(1), pp. 91-103. DOI: 10.1109/TBCAS.2019.2959278.

[21] R. Elakkiya, V. Subramaniyaswamy, V. Vijayakumar & A. Mahanti. (2022). Cervical Cancer diagnostics healthcare system using hybrid object detection adversarial networks. In: *IEEE Journal of Biomedical and Health Informatics*, *26*(4), pp. 1464-1471. DOI: 10.1109/JBHI.2021.3094311.