JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Malignant Tumor Cell Detection

Authors: Mrs. T. Veda Reddy (tvedareddycse@cvsr.ac.in), T. Sai Manogyana (18h61a05h2@cvsr.ac.in), S. P. Bala Meghana Shivani (18h61a05g7@cvsr.ac.in), S. Pavan Kumar Reddy (18h61a05h1@cvsr.ac.in)

Computer Science Department, Anurag Group of Institutions, Hyderabad

Abstract - Machine learning (ML) is widely known because of the methodology of choice in carcinoma detection and forecast modeling and its advantages in critical features detection from complex datasets. Several machine learning models are applied to predict whether the diagnosis is malignant or benign when the tumor is found. Cancer is that the second leading reason behind death globally and accounted for 8.8 million deaths in 2015. The first diagnosis and prognosis of a cancer type became a necessity in cancer research, because it can facilitate the next clinical management of patients. For better clinical decisions, it's important to accurately distinguish between benign and malignant tumors. Conventionally, statistical methods are used for the classification of high-risk and low-risk cancer, despite the complex interactions of high-dimensional medical data. To beat the drawbacks of conventional statistical methods, machine learning has emerged as a promising technique for handling high-dimensional data, with increasing application in clinical decision support. Deep Neural Networks (DNNs) are being employed here to classify the tumor as benign or malignant. The technique of cancer classification relies on several key differentiators like radius, smoothness, compactness etc. This prediction can help health workers in early detection of cancer type and supply essential diagnosis for it.

Keywords - malignant tumor, benign tumor, breast cancer, machine learning, neural networks

1. INTRODUCTION

Breast cancer is one amongst the various sorts of cancers. It's a cancer that develops within the breast cells and like most other cancers, it progresses stage-by-stage. A number of the first signs of carcinoma could include lump within the underarm or breast, itching or discharge from the nipples and alteration in skin texture of the nipple or the breast. It often needs medical imaging or lab testing for diagnosis. The possibilities of developing carcinoma increase with an increase in age.

According to the statistics of Globocan for the year 2018, carcinoma is fairly common within the younger age bracket (25 to 49 years), accounting for nearly 37.7% of all cases, which may be a pretty high number. Carcinoma peaks in the people aged 50 to 69 (which accounts for nearly 46.5% of all cases) and so starts reducing within the people of 70 years and above. It may not necessarily be reducing, but it has something to do with lifetime.

Given India's huge population, automating the method of diagnosing carcinoma accurately using machine learning techniques could be miraculously helpful in saving the lives of the many women, young or aged. It's also a standard disease in men in addition.

Digital mammography could be a major diagnosis model used throughout the globe for carcinoma detection. Computer-aided diagnosis (CAD) is widely employed in detecting numerous diseases with accurate decision. It assists healthcare professionals to analyse and conclude the stages of the cancer. CAD systems are developed in a very thanks to provide promising results and better decisions on a patient's condition which helps medical practitioners to diagnose the stages of the disease. It supports radiologists to avoid misconceptions and wrong diagnoses because of inaccurate data, lack of focus, or inexperience. The aim of the system is to create an enhanced ML model to diagnose carcinoma in its early stages with more accurate results using DNN.

DNN is followed by an analogous artificial neural network with a posh network structure that has 'n' hidden layers, which may process the input file from the previous layer. The error rate of the computer file are consistently reduced by adjusting the weights of each node, which ends up in achieve an accurate result. It helps to form a model and define its complex hierarchies in an exceedingly simple form. It supports all types of algorithms, namely supervised, unsupervised, semi-supervised, and reinforcement. So, the system didn't define any specific algorithm. The DNN generates a much better model themselves to coach the given data.

DNN is analogous to an artificial neural network but it has a fancy network structure with 'n' hidden layers, which might process the computer file from the previous layer. The error rate of the input data is consistently reduced by adjusting the weights of each node, which aids in achieving an accurate result. It helps to make a model and define its complex hierarchies during a simple form. It supports every kind of algorithms, namely supervised, unsupervised, semi-supervised, and reinforcement. So, the system need not define any specific algorithm. The DNN generates an improved model by itself to train using the given data.

Digital mammography is a major diagnosis model used in all countries for carcinoma detection. Computer-aided diagnosis (CAD) is widely employed in detecting numerous diseases with accurate decision. It assists the healthcare professionals to research and conclude the stages of the assorted diseases. CAD systems are developed in a very thanks to provide promising results and ideal decisions on patient's condition that helps medical practitioners to diagnose the stages of the diseases. It supports radiologists to avoid misconceptions and wrong diagnosis due Digital mammography could be a major diagnosis model used throughout the globe for carcinoma detection. Computer-aided diagnosis (CAD) is widely employed in detecting numerous diseases

with accurate decision. It assists the healthcare professionals to investigate and conclude the stages of the assorted diseases. CAD systems are developed in a very thanks to provide promising results and ideal decisions on patient's condition that helps medical practitioners to diagnose the stages of the diseases. It supports radiologists to avoid misconceptions and wrong diagnosis because of inaccurate data, lack of focus, or inexperience, who uses visually screening mammogram of patients. The aim of the system is to develop a unique CAD model to diagnose a carcinoma in earlier stages with more accurate results to avoid wasting their precious lives by using DNN.

Motivation

The main motivation of this research is to provide an accurate disease diagnosis framework using neural networks. Manually, doctors have to perform number of tests in order to diagnose a particular disease which requires a lot of time, effort and money. Automated disease diagnosis system will predict tumor malignancy with high accuracy resulting in time and effort reduction.

Research Contribution

Following are some of the major contributions of the proposed research:

- Improving old manual systems
- Prediction of tumor malignancy
- Enhanced accuracy
- Improving efficiency and effectiveness

2. LITERATURE REVIEW

Neural Network is inspired by the rule of biological neural networks, which has its own input and output channels called as dendrites and axons, respectively. A typical ANN will have many processing units or elements, which forms a highly interconnected network that processes a large amount of knowledge supported the response, given by the external input of a automatic data processing system. Every single neuron during a typical neural network is termed as unit. A layer in a very neural network is taken into account as a collection of neurons in an exceedingly stack. A layer may have n number of nodes in it. A typical neural network system has single input layer and should have one or two hidden layers, which is directly connected to the output layer, which receives input from the input layer, i.e., previous layer of this node.

Classification is a data mining process that aims to divide data into classes to facilitate decision-making; it's therefore a vital task in medical field. during this paper we are going to attempt to improve the accuracy of the classification of six machines learning algorithms: Bayes Network (BN), Support Vector Machine (SVM), k-nearest neighbors algorithm (Knn), Artificial Neural Network (ANN), Decision Tree (C4.5) and Logistic Regression using feature selection techniques, for carcinoma classification and diagnosis. We examined those

methods of classification and techniques of feature selection in WEKA Tool (The Waikato Environment for Knowledge Analysis) using two databases, Wisconsin carcinoma datasets original (WBC) and diagnostic (WBCD) available in UCI machine learning repository.

Mert et al. [14] used radial basis function neural network (RBFNN) for medical data classification and independent component analysis for feature selection. the tactic selects the one feature vector randomly from 30 features. the strategy obtained the accuracy within the average of 86%. Bhattacherjee et al. used BPNN for classification. the tactic achieved 99.27% accuracy. An intelligent medical decision model was developed supported evolutionary strategy. They validated the performance of the strategy by testing on different datasets. Neural network (NN), genetic algorithm (GA), support vector machine (SVM), K-nearest neighbor (KNN), multilayer perceptron (MLP), radial basis function (RBF), probabilistic neural network (PNN), self-organizing map (SOM), and Naive Bayes (nB) are used as classifiers. Crossover and mutation techniques are applied between different algorithms. The tactic proves that the SVM classifier on WBC data set attained better recognition rate than other classifiers.

Jouni et al. [11] proposed a model supported artificial neural network with multi-layered perceptron networks and BPNN. This model learns to classify whether the results of the simulation are malignant or benign. It also includes weight adjustment factors and bias values. Bewal et al. used multilayered perceptron network with four back-propagation training algorithms like quasi-Newton, gradient descend with momentum and adaptive learning, Levenberg–Marquardt, and resilient back propagation to train the network. Steepest descent back propagation is employed to measure the performance of other neural networks. Levenberg Marquardt algorithm with MLP achieved best accuracy rate of 94.11%. SVM with recursive feature elimination (RFE) applied on Wisconsin Diagnostic carcinoma (WDBC) Data set. Principal component analysis (PCA) applied separately for the identical data set for dimensionality reduction process, and SVM is employed to classify the information set. After PCA applied, 25 features are selected. It achieved 98.58% which outperforms SVM and SVM-RFE techniques [22].

Any kernel function in SVM applied to data set without feature selection process will increase computational time of the system. GA is employed to pick best features from data set. The results show that, for little scale datasets, linear kernel with bagging ensembles and RBF with boosting ensembles outperforms than other classifiers. The information set is split into 90–10% splits supported k-fold cross-validation. GA+RBFSVM achieved accuracy of 98.00 and 99.52%, respectively, for tiny and huge datasets. Nayak et al. [16] proposed a system which has adaptive resonance theory (ART-1) network for classification, and it's compared with PSO-MLP and PSO-BBO algorithms which prove that ART is best than other two classifiers. They split the info set into 70–30 for training and testing the dataset.

Onan [18] proposed a completely unique classification model supported fuzzy-rough nearest neighbor method. This method consists of three phases, namely instance selection, feature selection, and

classification. Fuzzy rough instance selection method is employed for data point selection with weak gamma evaluator to get rid of erroneous and ambiguous instances. Consistency-based feature selection method is employed in conjunction with re-ranking algorithm to efficiently look for possible enumerations in search space. Fuzzy-rough nearest neighbor method is employed for the classification process. This tactic performed better than other fuzzy approaches.

Cancer is one among the foremost dangerous diseases to humans, and yet no permanent cure has been developed for it. Carcinoma is one among the most common cancer types. In accordance with the National Breast Cancer foundation, in 2020 alone, over 276,000 new cases of invasive carcinoma and quite 48,000 non-invasive cases were diagnosed within the US. to place these figures in perspective, 64% of those cases are diagnosed early within the disease's cycle, giving patients a 99% chance of survival. AI and machine learning are used effectively in detection and treatment of several dangerous diseases, helping in early diagnosis and treatment, and thus increasing the patient's chance of survival. Deep learning has been designed to research the foremost important features affecting detection and treatment of significant diseases. as an example, carcinoma is detected using genes or histopathological imaging. Analysis at the genetic level is extremely expensive, so histopathological imaging is that the commonest approach accustomed detect carcinoma. During this research work, we systematically reviewed previous work done on detection and treatment of carcinoma using genetic sequencing or histopathological imaging with the assistance of deep learning and machine learning. We also provide recommendations to researchers who will add this field.

Breast cancer remains one amongst the top diseases that result in thousands of death in women once a year. Sub-field of Computer Science, AI has been utilized for early, rapid, and accurate diagnosis of breast tumors. The target of this paper was to review recent studies for classifying these tumors. Machine learning algorithms like Support Vector Machine (SVM), K-Nearest Neighbour (K-NN), and Random Forest (RF) were used to classify medical images into malignant and benign. Moreover, deep learning has been employed recently for the identical purpose. Among them, Convolutional Neural Network (CNN) is one among the most popular techniques. The results showed that the SVM achieved high accuracy, about 97%, therefore, the researchers utilized various functions for this algorithm and added more features like bagging and boosting to extend its efficacy. Additionally, deep learning obtained high accuracy using CNN which is more than 98%.

Schmidhuber [21] provides a summary of deep learning in neural networks. The strategy proves that the deep learning algorithms reduced the error rate and increased the accuracy in terms of training of algorithm. Abdel and Eldeib Breast Cancer Classification Using Deep Neural Networks 231 applied deep belief network (DBN) for WBC data set and achieved 99.68% accuracy. The input set was divided into train-test split of 54.945.1%. DBN follows unsupervised path and back-propagation network to follow supervised path. This system was created using BPNN with Levenberg Marquardt learning function. Here, the weights are initialized with DBN path. This technique provided a promising result and performed much better than other classifiers.

This motivated the use of deep learning concepts for medical data classification. Deep learning reduces the error rate and improves the accuracy rate.

3. PROPOSED APPROACH

The proposed approach used to complete this research is started by retrieving the dataset present in scikit-learn library. After verifying the dataset, next step is pre-processing in the form of Data cleaning, Data Transformation. Later a Deep Neural Network algorithm is applied on the data. After applying algorithms and techniques we analyze results and discuss about conclusion.

A. Data Collection

The dataset is collected from Kaggle, which is an open online source and is associated with multitudes of diseases and covers a large source of databases, domain theories and data generators which are utilized by the researchers

B. Data Preprocessing

Preprocessing of data is presented in an intelligible presentation by turning raw data into fathomable context for a purpose. It involves data transformation through Normalization. As a result, data quality is improved resulting in usefulness of data.

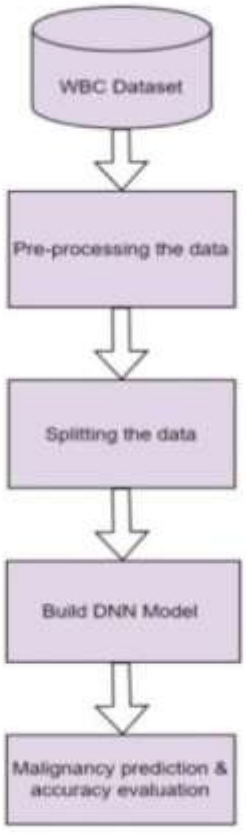

Figure 1: Flow Chart of Proposed Approach

C. Classification and Prediction

The system divides the dataset into 80–20 train-test split for experimental purpose. Splitting of the dataset is done randomly. After partitioning, a training set of data is initially applied to the classifier. This deep neural network classifier has a network structure with four input nodes, out of which three are hidden layers with 10, 20, and 10 hidden nodes. In addition, it has an output layer with a single node that predicts malignancy. We define activation function Rectified Linear unit (ReLU) for hidden layer and for the output layer, we define sigmoid function as the activation function. We apply Adam optimizer along with a learning rate of 0.0001. The number of epochs to back propagate from output is set to 86. Since this network has multiple layers with a huge number of inner nodes, computationally it is expensive but provides promising results after training the model.

DEEP NEURAL NETWORK:

Deep neural networks follow the structure of a typical artificial neural network with a complex network model. It helps us to create a model and define its complex hierarchies in a simple form. It has 'n' hidden layers and processes the data from the previous layer called as the input layer, and after every epoch, error rate of the input data will be gradually reduced by adjusting the weights of every node, back-propagating the network and continues till reaches better results.

Algorithm of Deep Neural Network

1. Define a neural network with an input layer having n input nodes.

2. Initialize the number of hidden layers needed to train the data.

3. Define the learning rate and bias value for every node. The weight will be randomly selected in initial forward propagation.

4. Define the activation function

5. Define the number of epochs to back propagate the value from the output node.

6. Train the network for given set of training data.

7. After the network is trained, pass the test data to the trained network to find the classification rate of the model.

8. Train the network until the number of epochs is completed (or) expected output is achieved.

9. Calculate the accuracy of the model using evaluation metrics

D. Trained Model

Model is trained using Deep Neural Network and test data is applied to it test the accuracy.

E. Experimental results

The performance of a model is estimated through confusion matrix. The confusion matrix helps to find the classified and misclassified rate of the system. Effectiveness and performance of a system can be measured by calculating the accuracy.

Deep neural network: After applying the DNN algorithm, we achieved an accuracy of 98.62 %.

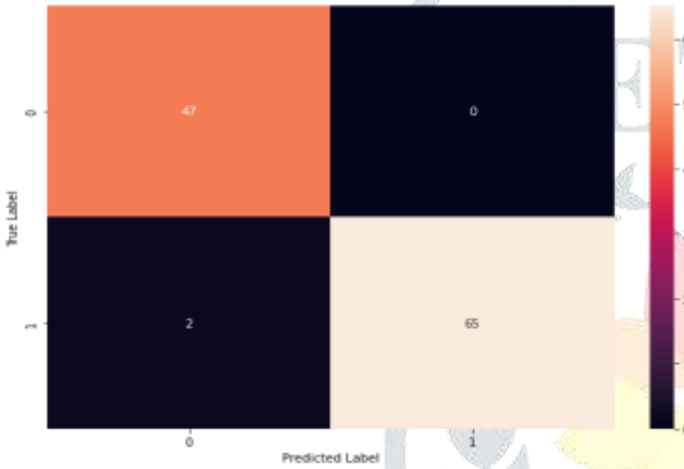

Figure 2: Confusion matrix

4. CONCLUSION & FUTURE WORK

As breast cancer is one of the leading causes of death in women with high mortality rate, early detection can immensely help in giving a better prognosis for patients by helping medical professionals place more emphasis on early care and better treatment plans instead of on diagnosis/detection of cancer.

Experimental results proved that the proposed DNN is quite better than the existing methods. It is ensured that the proposed algorithm is advantageous in terms of accuracy and efficiency.

In future these techniques can also be applied on real time medical datasets and also can be used in the form of ensembles i.e., combinations of multiple techniques. This would result in further increase of accuracy and performance.

5. REFERENCES

[1]Details on breast cancer found on the website: http://www.breastcancer.org/symptoms/understand_bc/what_is_bc.

[2] Y.S. Hotko, Male breast cancer: clinical presentation, diagnosis, treatment, Exp. Oncol. 35 (4) (2013) 303–310.

[3]https://www.biospectrumindia.com/views/21/15300/statistical-analysisof-breast-cancer-in-india.html.

[4] S. Malvia, S.A. Bagadi, U.S. Dubey, S. Saxena, Epidemiology of breast cancer in Indian women, Asia Pac. J. Clin. Oncol. 13 (4) (2017) 289–295.

[5] V. Anji Reddy, Badal Soni, Breast cancer identification and diagnosis techniques, in: Machine Learning for Intelligent Decision Making, Springer, 2020.

[6] Qiao Pan, Yuanyuan Zhang, Dehua Chen, Guangwei Xu, CharacterBased Convolutional Grid Neural Network for Breast Cancer Classification, IEEE, 2017, p. 31.

[7] SanaUllah Khan, Naveed Islam, Zahoor Jan, Ikram Ud Din, Joel J.P.C. Rodrigues, A novel deep learning based framework for the detection and classification of breast cancer using transfer learning, in: Pattern Recognition Letters, Elsevier, 2019.

[8] Qinghua Huang, Yongdong Chen, Longzhong Liu, Dacheng Tao, Xuelong Li, On combining biclustering mining and adaboost for breast tumor classification, IEEE Trans. Knowl. Data Eng. 32 (4) (2020) 728–738.

[9] Shweta Kharya, SunitaSoni, Weighted naive bayes classifier: A predictive model for breast cancer detection, Int. J. Comput. Appl. 133 (9) (2016) 32–37.

[10] R.D.H. Devi, M.I. Devi, Outlier detection algorithm combined with decision tree classifier for early diagnosis of breast cancer, Int. J. Adv. Engg. Tech./Vol. VII/Issue II/April-June 93 (2016) 98.

[11] Jouni, H., Issa, M., Harb, A., Jacquemod, G., & Leduc, Y. (2016). Neural network architecture for breast cancer detection and classification. In: IEEE International Multidisciplinary Conference on Engineering Technology (IMCET), pp. 37–41.

[12] Karabatak, M., & Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications, 36(2), 3465–3469.

[13] Kiyan, T., & Yildirim, T. (2004). Breast cancer diagnosis using statistical neural networks. Journal of Electrical and Electronics Engineering, 4(2), 1149–1153.

[14] Mert, A., Kılıc¸, N., Bilgili, E., & Akan, A. (2015). Breast cancer detection with reduced feature set. Computational and Mathematical Methods in Medicine.240 S. Karthik et al.

[15] Nahato, K. B., Harichandran, K. N., & Arputharaj, K. (2015). Knowledge mining from clinical datasets using rough sets and backpropagation neural network. Computational and Mathematical Methods in Medicine.

[16] Nayak, T., Dash, T., Rao, D. C., & Sahu, P. K. (2016). Evolutionary neural networks versus adaptive resonance theory net for breast cancer diagnosis. In: Proceedings of the International Conference on Informatics and Analytics (ACM), p. 97.

[17] Nilashi, M., Ibrahim, O., Ahmadi, H., & Shahmoradi, L. (2017). A knowledge-based system for breast cancer classification using fuzzy logic method. Telematics and Informatics, 34(4),133–144.

[18] Onan, A. (2015). A fuzzy-rough nearest neighbor classifier combined with consistency-based subset evaluation and instance selection for automated diagnosis of breast cancer. Expert Systems with Applications, 42(20), 6844–6852.

[19] Paulin, F., & Santhakumaran, A. (2011). Classification of breast cancer by comparing back propagation training algorithms. International Journal on Computer Science and Engineering, 3(1), 327–332.

[20] Prevention Control: Center for Diseases Control and Prevention (2014): https://www.cdc.gov/cancer/breast/index.htm.

[21] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. Neural Networks,61, 85–117.

[22] Yin, Z., Fei, Z., Yang, C., & Chen, A. (2016). A novel svm-rfe based biomedical data processing approach: Basic and beyond. In: IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, pp. 7143–7148