

Deep Neural Network Design and Implementation for Predicting Malignancy of Cancer Tumors



Javvaji Geetha Priya, Koppiseti Sri Satyanjani, K Amrendra, Kolla Bhanu Prakash

Abstract: Breast cancer is the most frequent cancer among women after lung cancer. It is second popular cause of death in the world. Breast cancer cells usually form a tumor (abnormal tissue). There are two types of breast cancer tumors: those that are non-cancerous, or 'benign', and those that are cancerous, which are 'malignant'. Early detection of cancer tumors can prevent its mortality and can take respective precautions according to the type of cancer tumor (either it is benign or malignant). The main objective of this paper is to early diagnosis of the malignancy of cancer tumors which helps to decrease the death rate and helps to give more lives. The selection of suitable deep learning technique is a challenge for the diagnosis of breast cancer. So we created model for prediction of malignancy of cancer tumors using neural network to analyze risk levels that helps in prognosis. It is useful for a doctor to predict the stage of cancer and take respective precautions

Keywords: Malignancy, benign, neural, network, cancer, random forest, classification.

I. INTRODUCTION

Breast cancer is the second most common cancer among women in the United States [1]. In the United States in 2019, there will be around 268,600 new cases of breast cancer, and an estimated of 41,760 women will die from breast cancer [1]. In recent years, incidence rates have expanded marginally by 0.3% every year [1]. Death rates from female breast cancer dropped 40% from 1989 to 2016. Since 2007 breast cancer death rates have been consistent in women younger than 50, however have kept on decreasing in older women [1]. These

decreases are believed to be the result of finding breast cancer earlier through screening and increased awareness, as well as better treatments [1]. Breast cancer cells usually form a tumor that can often be seen on an x-ray or felt as a lump. There are two types of breast cancer tumors: those that are non-cancerous or 'benign' and those that are cancerous, which are 'malignant' [1]. Malignant tumors are cancerous and aggressive because they invade and damage surrounding tissue. When a tumor is suspected to be malignant, the doctor will perform a biopsy to determine the severity or aggressiveness of the tumor [1]. A new technology is 3-D mammography also called breast tomosynthesis [2]. This procedure takes images from different angles around the breast and builds them into a 3-D-like image and used to detect breast cancer in advance [2]. In this paper we proposed a predictive model to detect whether the cancer is benign or malignant in order to prevent a breast cancer at early stages using machine learning techniques. Also estimate the validation of the models.

II. LITERATURE REVIEW

After reviewing the different literature showed that there have been several studies on the early papers like prediction of breast cancer using random forest, support vector machines [3]. Padma Priya and Sowmiya worked on the prediction of breast cancer using classification algorithms like Decision tree [4]. Shravya and pravalika implemented the models using logistic regression, K nearest neighbor. There experiments have shown that SVM is the best for predictive analysis with an accuracy of 92.7% [5]. Puneetyadav also worked on the prediction of breast cancer using decision tree model and svm. The performances of the two models have been evaluated using the statistical measures and found that support vector machine model performance is better than decision tree model on the prediction of severity of breast cancer [6]. All this papers generally predict the breast cancer using different data mining techniques

In this paper we predict a breast cancer tumor is either in benign state or malignant state using the neural network [8]. The purpose of this research is to develop prediction models that detect the cancer is benign or malignant and take respective precautions to prevent it. [9]

III. METHODOLOGY

In this paper, we are focusing on supervised learning; the area where so far deep learning had the most success is neural network [10]. Neural network helps to predict malignancy in cancer tumors very efficiently. [11]

Manuscript published on November 30, 2019.

* Correspondence Author

Javvaji Geetha Priya*, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.

Koppiseti Sri Satyanjani, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.

K Amarendra, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.

Kolla Bhanu Prakash, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

In order to perform the model we use a real data-set connected to the detection of breast cancer tumors.^[12] After studying case studies, extensive literature reviews shows that there are lots of factors influencing cancer.^[13] These factors are identified and taken as attributes for this study.^[14] This data was collected by the University Of Wisconsin Hospitals, Madison, Wisconsin, USA^[7]. There are 10 columns and 699 rows, the last attribute in the dataset is class of the tumor and it has two categorical values 2 indicates that tumor was found to be benign and 4 indicates malignant^[7].

Our goal is to prepare our neural network to predict if a tumor is benign or malignant, based on the attributes provided by the data.^[15] The benign and malignant classes are related to the digits 2 and 4.^[16] The last layer of our network which is output layer ranges between 0 and 1 through its sigmoid function.^[17] Moreover, neural networks generally work better when data is set in the range from 0 to 1.^[18] We will therefore change the values of the class columns to hold a 0 instead of a 2 for benign and a 1 instead of a 4 for malignant.^[19] I think it is easier to describe neural network as a mathematical function that maps a given input to a desired output.^[20] Neural network consists of an input layer, hidden layers, output layer and a set of weights and biases between each layer.^[21] By applying the neural network to the Wisconsin cancer dataset to predict the cancer tumor is benign or malignant based on the 9 attributes and exploring deeply how back propagation and gradient descent works^[8].

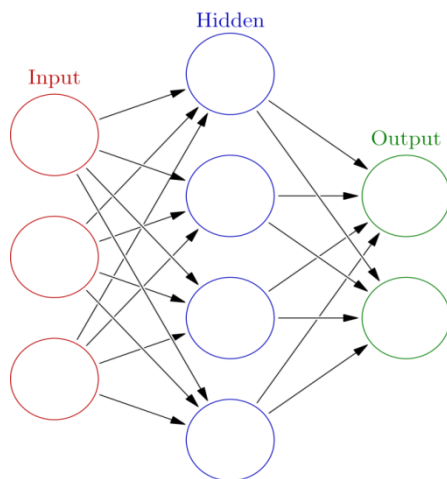


Fig .1 shows the deep neural network model with 3 neurons in the input layer, 4 in the hidden layer and 2 in the output layer.^[10]

IV. ALGORITHM

In this algorithm the network is trained by repeatedly processing the trained data set and comparing output with the actual output. If anything differs between the outputs we got from our network and the desired output then this is called as error. So after that we again back propagate the output we got from our neural network to the back layer of the neural network till first hidden layer so that again our network will assign different weight randomly to the layers .And try to generate the output with lesser error than the previous time and at every iteration our network output should be lesser than the output of previous time. So that after some iterations we should get desire output and we keep on back propagating the output from our neural network till our network trained completely.^[23] If the model sets its weights in way that it gives correct output with trained set and neglects to generate

the new samples outside the training set, in order to prevent this type of problems we generate 3 kinds of datasets.^[24] They are training set, validation set, test set.^[25] Training set is used to train the network, validation set is used to verify the quality of the training and test set used to test the network with new related data^[9].

In this model we make another variable called threshold or limit.^{[26] [27] [28]} It will control our certainty threshold; how close to 1 the output of the network should be before we choose that a tumor is malignant by default it is set to 0.5 or we can change according to our model.^{[29] [30] [31]}

V. RESULTS AND DISCUSSION

The Results obtained were based on the dataset of breast cancer taken from University Of Wisconsin Hospitals, Madison, Wisconsin, USA. We used neural network model to predict the malignancy of cancer tumors.^[32] The model is trained well on the dataset and we consider the output values to be higher than 0.5 for output to be considered malignant. We got the accuracy of trained dataset 100% shown in Fig.2 and accuracy for the validation set is 99.4% shown in the Fig.3.

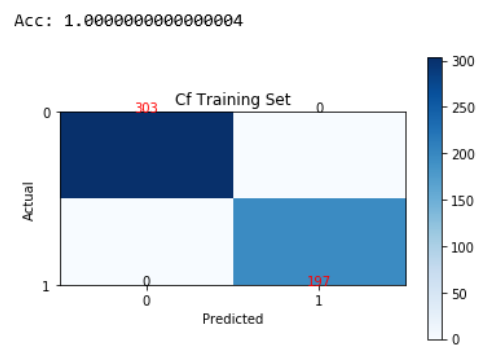


Fig .2 Training set with accuracy of 100%

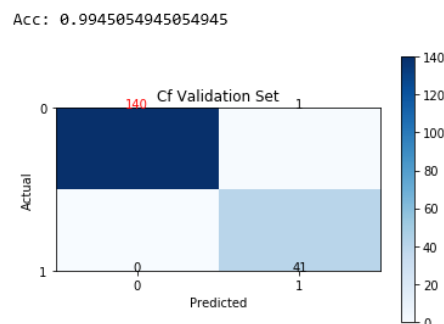


Fig.3 Validation set with accuracy of 99.4%

VI. CONCLUSIONS

Cancer is one of the dangerous disease in the world. Now a days many machine learning and data mining techniques are used to classify tumor cells such as decision trees, k-nearest neighbours etc.^[33] But in this paper we tried to classify the cancer tumor cells as benign and malignant using neural network with good accuracy.

In future, accuracy of neural network can be enhanced by increasing neurons in hidden layer in order to improve the performance of the model.

REFERENCES

1. Kavuri, M. & Prakash, K.B. 2019, "Performance comparison of detection, recognition and tracking rates of the different algorithms", International Journal of Advanced Computer Science and Applications, vol. 10, no. 6, pp. 153-158
2. Kolla, B.P., Dorairangaswamy, M.A. & Rajaraman, A. 2010, "A neuron model for documents containing multilingual Indian texts", 2010 International Conference on Computer and Communication Technology, ICCCT-2010, pp. 451.
3. Kolla, B.P. & Raman, A.R. 2019, Data Engineered Content Extraction Studies for Indian Web Pages, Advances in Intelligent Systems and Computing, 711, pp. 505-512.
4. Naga Pawan, Y.V.R. & Prakash, K.B. 2019, "Variants of particle swarm optimization and onus of acceleration coefficients", International Journal of Engineering and Advanced Technology, vol. 8, no. 5, pp. 1527-1538.
5. Pradeep Kumar, V. & Prakash, K.B. 2019, "QoS aware resource provisioning in federated cloud and analyzing maximum resource utilization in agent based model", International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 8, pp. 2689-2697.
6. Prakash, K.B. 2018, "Information extraction in current Indian web documents", International Journal of Engineering and Technology(UAE), vol. 7, no. 2, pp. 68-71.
7. Prakash, K.B. 2017, "Content extraction studies using total distance algorithm", Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016, pp. 673.
8. Prakash, K.B. 2015, "Mining issues in traditional indian web documents", Indian Journal of Science and Technology, vol. 8, no. 32, pp. 1-11.
9. Prakash, K.B., Ananthan, T.V. & Rajavarman, V.N. 2014, "Neural network framework for multilingual web documents", Proceedings of 2014 International Conference on Contemporary Computing and Informatics, IC3I 2014, pp. 392.
10. Prakash, K.B. & DoraiRangaswamy, M.A. 2019, Content extraction studies for multilingual unstructured web documents, Advances in Intelligent Systems and Computing, 749, pp. 653-664.
11. Prakash, K.B. & DoraiRangaswamy, M.A. 2016, "Content extraction studies using neural network and attribute generation", Indian Journal of Science and Technology, vol. 9, no. 22, pp. 1-10.
12. Prakash, K.B., DoraiRangaswamy, M.A. & Ananthan, T.V. 2014, "Feature extraction studies in a heterogeneous web world", International Journal of Applied Engineering Research, vol. 9, no. 22, pp. 16571-16579.
13. Prakash, K.B., DoraiRangaswamy, M.A., Ananthan, T.V. & Rajavarman, V.N. 2015, "Information extraction in unstructured multilingual web documents", Indian Journal of Science and Technology, vol. 8, no. 16.
14. Vijay Muni, T., Srikanth, K.S., Venkatesh, N., Sumedha, K.L., " A high performance hybrid MPPT control scheme for a grid connected PV system based three level NPCMLI", International Journal of Engineering and Technology(UAE), 7(2), pp. 41-43
15. T. Vijay Muni, G Sai Sri Vidya, N Rini Susan, "Dynamic Modeling of Hybrid Power System with MPPT under Fast Varying of Solar Radiation", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 1 (2017), pp.:530-537.
16. M Srikanth, T. Vijay Muni, M Vishnu Vardhan, D Somesh, "Design and Simulation of PV-Wind Hybrid Energy System", Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 04-Special Issue, 2018, pp: 999-1005
17. S Ilahi, M Ramaiah, T Vijay Muni, K Naidu, " Study the Performance of Solar PV Array under Partial Shadow using DC- DC Converter", Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 04-Special Issue, 2018, pp: 1006-1014.
18. S Moulali, T Vijay Muni, Y Balasubrahmanyam, S Kesav,"A Flying Capacitor Multilevel Topology for PV System with APOD and POD Pulse Width Modulation", Jour of Adv Research in Dynamical & Control Systems, Vol. 10, 02-Special Issue, 2018, pp: 96-101.
19. T Vijay Muni, S V N L Lalitha, "Fast Acting MPPT Controller for Solar PV with Energy Management for DC Microgrid", International Journal of Engineering and Advanced Technology (JJEAT), Volume 8, Issue 5, pp-1539-1544.
20. T Vijay Muni, S V N L Lalitha, "Power Management Strategy in Solar PV System with Battery Protection Scheme", International Journal of Innovative Technology and Exploring Engineering, Volume 8, Issue 6, pp-960-964.
21. Prakash, K.B., DoraiRangaswamy, M.A. & Raman, A.R. 2010, "Text studies towards multi-lingual content mining for web communication", Proceedings of the 2nd International Conference on Trendz in Information Sciences and Computing, TISC-2010, pp. 28.
22. Prakash, K.B., Kumar, K.S. & Rao, S.U.M. 2017, "Content extraction issues in online web education", Proceedings of the 2016 2nd International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2016, pp. 680.
23. Prakash, K.B. & Rajaraman, A. 2016, "Mining of Bilingual Indian Web Documents", Procedia Computer Science, 89, pp. 514-520.
24. Prakash, K.B., Rajaraman, A. & Lakshmi, M. 2017, "Complexities in developing multilingual on-line courses in the Indian context", Proceedings of the 2017 International Conference On Big Data Analytics and Computational Intelligence, ICBDACI 2017, pp. 339.
25. Prakash, K.B., Rajaraman, A., Perumal, T. & Kolla, P. 2016, "Foundations to frontiers of big data analytics", Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics, IC3I 2016, pp. 242.
26. Prakash, K.B. & Rangaswamy, M.A.D. 2016, "Content extraction of biological datasets using soft computing techniques", Journal of Medical Imaging and Health Informatics, vol. 6, no. 4, pp. 932-936.
27. Prakash, K.B., Rangaswamy, M.A.D. & Raja Raman, A. 2012, ANN for multi-lingual regional web communication, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7667 LNCS(PART 5), pp. 473-478.
28. Prakash, K.B., Rangaswamy, M.A.D. & Raman, A.R. 2013, "Attribute based content mining for regional web documents", IET Seminar Digest, pp. 368.
29. Prakash, K.B., Rangaswamy, M.A.D. & Raman, A.R. 2012, Statistical interpretation for mining hybrid regional web documents, Communications in Computer and Information Science, 292 CCIS, pp. 503-512.
30. Ismail, M., Prakash, K.B. & Rao, M.N. 2018, "Collaborative filtering-based recommendation of online social voting", International Journal of Engineering and Technology(UAE), vol. 7, no. 3, pp. 1504-1507.

