ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

International Journal of Mechanical Engineering

Breast Cancer Prediction using Neural Networks

[1]Shreeanant Bharadwaj,[2]SwapnilBhalerao,[3]Abhishek Suryawanshi, [4]AnkitChahar,[5]EshitaSohani, and [6]AnishReddyBanda,

[1]Department of Computer Engineering JSPM Rajarshi Shahu College Of Engineering, Tathawade, Pune, India

[2]Department of Computer Engineering JSPM Rajarshi Shahu College Of Engineering, Tathawade, Pune, India

[3]Department Of Computer Science RV College of Engineering Bengaluru,Karnataka,India

[4]Department Of Computer Science RV College of Engineering Bengaluru, Karnataka,India

[5]B.E,Department of Electronics and Telecommunications Engineering Institute Of Engineering and Technology, DAVV Indore,Madhya Pradesh, India.

[6]Department of Computer Science Engineering Mahatma Gandhi Institute of Technology Hyderabad,India

Abstract. Breast cancer is most commonly diagnosed in women and is a major cause of increased mortality in women. Breast cancer diagnosis takes time, and because systems are limited, it is vital to design a system that can automatically diagnose breast cancer in its early stages. For the categorization of benign and malignant tumours, many Machine Learn-ing and Deep Learning Algorithms have been applied. In this world of 7 billion people and out of which 3.4 billion are women and in that 1 out of every 22 women are diagnosed with breast cancer. The dataset has been taken from the alcrase dataset, which contains 15509 datasets and 30 features which would be used in detecting the results from the algorithm ap-plied.Though this method cannot de nitively detect cancer, it can assist clinicians in determining whether or not a biopsy is necessary by giving information on whether or not the patient has breast cancer.Confusion matrix and ROC analyses were used to evaluate the de nite diagnosis for each patient as well as the data from the ANN model ndings. The main idea for the paper is to show how the algorithm made by us has an increased accuracy and can be used to accurately predict the breast cancer before the diagnosis.

Breast cancer prediction , Neural Networks , Deep Learning , Convolution Neural Network

1 Introduction

Breast cancer is~~ a type of can~~cer that develops in the cells of the breast and is a highly frequent disease in women. Breast cancer, like lung cancer, is a life-threatening illness for women. Breast cancer is classi ed into several kinds based on how the cells appear under a microscope. The two most common forms of breast cancer are (1) invasive ductal carcinoma (IDC) and (2) ductal carcinoma in situ (DCIS), the latter of which progresses slowly and has little impact on patients' everyday lives.The DCIS type accounts for a small fraction of all instances (between 20 and 80 percent) ; on the other hand, the IDC form is more haz-ardous, encircling the entire breast tissue. This is the case for the most majority of breast cancer patients (about 80 percent). Lung cancer is the most deadly Supported by organization x.

malignancy, followed by breast cancer. Breast cancer accounts for roughly 11 percent of new cancer cases, with women accounting for almost 24 percent.In the event of any cancer sign or symptom, people seek the advice of an oncologist. Mammograms, magnetic resonance imaging (MRI) of the breast, ultrasound of the breast X-ray, tissue biopsy, and other tests can help the oncologist diagnose and identify breast cancer.Arti cial neural networks are neural networks that are built on arti cial intelligence (ANN). Instead of designing a computer system to perform certain tasks, ANN trains it to execute tasks. To create an arti cial neu-ral network, several arti cial neurons are coupled in line with speci ed network architecture. The neural network's goal is to turn the inputs into meaningful outputs with a greater accuracy.

2 Literature Overview

Breast cancer is currently classi ed using immunohistochemistry (IHC), histopatho-logic features, and molecular characterisation. The two most frequent histologic subtypes of invasive breast cancer are invasive ductal carcinoma and invasive lob-ular carcinoma (80 percent to 85 percent and 10 percent to 15 percent of all cases, respectively). Other histologic cancer subtypes exist in the remaining 1 percent of invasive breast tumours[11]. Breast cancer HC characterization is essential for determining treatment options and predicting prognosis. IHC characterisation requires the expression of biomarkers such as oestrogen receptor (ER), proges-terone receptor (PR), and human epidermal growth factor receptor 2. ER and PR expression is less than 1 percent in around 75 percent of those with breast cancer who are categorised as having hormone receptor (HR){positive disease.

12 Furthermore, according to IHC, 15 to 30 percent of breast cancer patients have HER2 that has been ampli ed or overexpressed. 13 Triple-negative breast cancer refers to tumours that lack ER and PR expression as well as HER2 over-expression (TNBC). Historically, the TNM model has been used to categorise breast cancer into stages 0, 1, 2, and 3. In 2017, the 8th edition of the Amer-ican Joint Committee on Cancer's Cancer Staging Manual included prognostic biomarkers (such as histologic tumour grade, ER, PR, HER2, and multigene test-b) to breast cancer staging.

3 Methodology

3.1 Data preparation

Data has been taken from Wisconsin breast cancer diagnostic data, which has played an important part in many researches. As the data is pretty tough to collect, the only action possible to perform the model to do the prediction was to take a data from a good source .so the dataset has been taken from the uni-versity of Wisconsin, Madison which is created by Dr william H wolberg , W Nick Street and Olvi L[12]. A digitised picture of a ne needle aspirate (FNA) of a breast mass is used to compute features. They de ne the properties of the im-age's cell nuclei. ID number and Diagnosis (M = malignant, B = benign) are two attributes. Radius, texture, perimeter, area, smoothness, compactness, concav-ity, concave points, symmetry, and fractal dimension are among the properties in the dataset. .The samples can be seen in the gure 1.

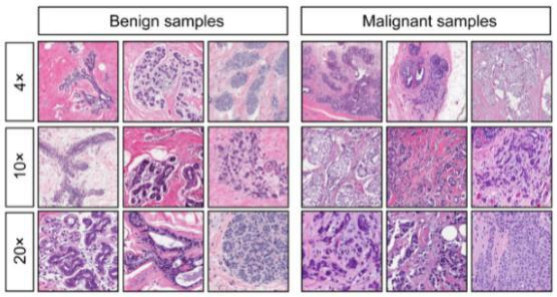

Fig. 1. Benign and Malign samples

3.2 Data Preprocessing:

A dataset is a collection of data elements called records, points, vectors, patterns, occurrences, instances, samples, observations, or entities.A range of characteris-tics that characterise data items capture the basic attributes of an item, such as the mass of a physical object or the time at which an event occurred. Features are described using words such as variables, characteristics, elds, attributes, and dimensions.. So the preprocessing is must perform the algorithm or model further which provides a base to follow the next step.So the rst step is to load the dataset which gives us the gure 2 , after this we would clean and prepare the data , in this step we will take the length of the dataset elements and take all the unique values which would help me getting more accurate results. After this we will describe the dataset which can be seen in gure 3.

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean
0	842302	M	17.99	10.38	122.80	1001.0	0.11840	0.27760
1	842517	M	20.57	17.77	132.90	1326.0	0.08474	0.07864
2	84300903	M	19.69	21.25	130.00	1203.0	0.10960	0.15990
3	84348301	M	11.42	20.38	77.58	386.1	0.14250	0.28390
4	84358402	M	20.29	14.34	135.10	1297.0	0.10030	0.13280

Fig. 2. Dataset

	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean
count	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000	569.000000
mean	0.372583	14.127292	19.289649	91.969033	654.889104	0.096360	0.104341
std	0.483918	3.524049	4.301036	24.298981	351.914129	0.014064	0.052813
min	0.000000	6.981000	9.710000	43.790000	143.500000	0.052630	0.019380
25%	0.000000	11.700000	16.170000	75.170000	420.300000	0.086370	0.064920
50%	0.000000	13.370000	18.840000	86.240000	551.100000	0.095870	0.092630
75%	1.000000	15.780000	21.800000	104.100000	782.700000	0.105300	0.130400
max	1.000000	28.110000	39.280000	188.500000	2501.000000	0.163400	0.345400

Fig. 3. Feature dataset

3.3 Feature Selection

In the actual world, it's uncommon that all the variables in a dataset are rele-vant for creating a machine learning model. Adding duplicate variables decreases the model's generalization ability and may also reduce a classi er's overall ac-curacy. Furthermore, adding more variables to a model enhances the model's total complexity. According to 'Occam's Razor's Law of Parsimony,' the optimal solution to a problem is one that requires the fewest assumptions. As a result, feature selection becomes an essential component in developing machine learning models. So basically there are mainly two features that the samples would be divided into malign and benign. And to display other features we would be using a correlation map which is a matrix which determines the relationship pairs in a table.

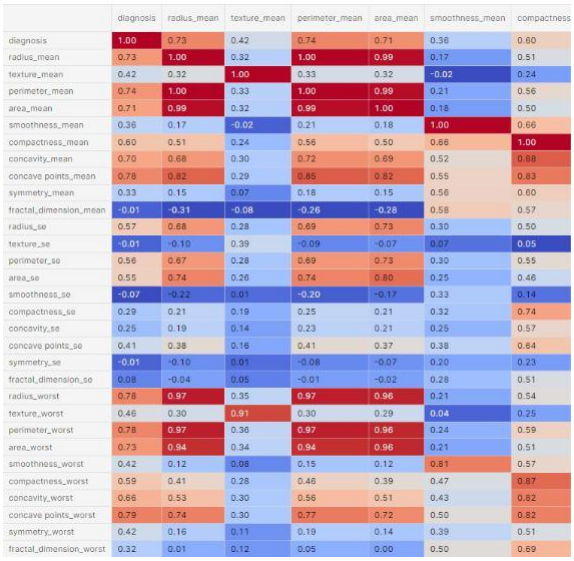

Fig. 4. corelation map

3.4 Model Architecture

The structure and function of a biological neural network are used to design ANN architecture. ANN is made up of neurons that are organized in layers, just like neurons in the brain. The input layer bu ers the incoming signal, while the output layer creates the network's output. These linked arrangements always contain two layers that are similar to all network architectures: input layer and output layer[10]. The third layer is the Hidden layer, which keeps neurons out of both the input and output layers. These neurons are concealed from anyone interacting with the system and serve as a black box for them. The system's computational and processing capability can be enhanced by adding additional hidden layers containing neurons, but the system's training phenomena become more complex at the same time. We have TensorFlow for performing the ANN, starting with that we would rst initialize the ANN using the sequential function. After that we would be adding layers to it, rstly the input layer which would have an activation function 'relu'. Activation function is used to assess whether a neural network's output is yes or no. It converts the values from -1 to 1or 0 to 1 and so on. Similar process for four hidden layers and after that for the output layer we would be using the 'sigmoid' as the activation function. Which is followed by the compiling of the ANN where we have chosen

adam optimiser , binary crossentropy as the loss function and accuracy as the metrics

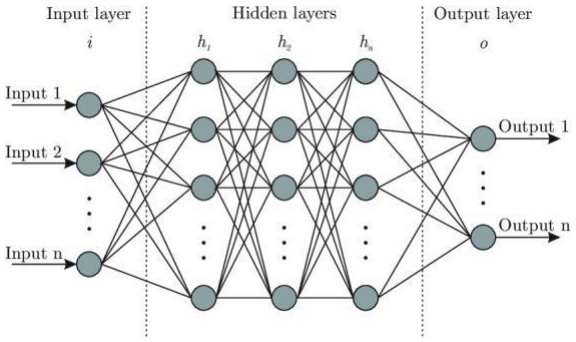

Fig. 5. ANN Architecture

4 Experimental results

Now, as we moved from collecting the data to preprocessing to feature selection to the model architecture, and it's usage, we are using 100 epochs to get better accuracy. If the epochs are less, then it means the model has not been trained well, and we need to train the model properly to get the best accuracy than the other algorithms. When the epochs are 25 we are getting an accuracy of 65

percent, gradually increasing it we are getting the accuracy of 87 percent at 50 epochs and nally increasing it to 100 we are getting an accuracy of 99 percent which is a tremendous accuracy as it predicts cancer 99 out of 100 times, which is a great boost to the health business. The epochs run accuracy at 100 epochs can be seen in gure 6. After that, a confusion matrix is important as it helps in summarizing the performance of the algorithm. Along with the performance matrix, the entire accuracy score of the algorithm is also required to get how much accurate it is.

```
Epoch 93/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0430 - accuracy: 0.9873
Epoch 94/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0435 - accuracy: 0.9870
Epoch 95/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0268 - accuracy: 0.9931
Epoch 96/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0504 - accuracy: 0.9892
Epoch 97/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0497 - accuracy: 0.9921
Epoch 98/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0516 - accuracy: 0.9905
Epoch 99/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0385 - accuracy: 0.9902
Epoch 100/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0400 - accuracy: 0.9930
```

Fig. 6. Epoch Accuracy

```
Epoch 94/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0435 - accuracy: 0.9870
Epoch 95/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0268 - accuracy: 0.9931
Epoch 96/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0504 - accuracy: 0.9802
Epoch 97/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0497 - accuracy: 0.9761
Epoch 98/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0516 - accuracy: 0.9785
Epoch 99/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0385 - accuracy: 0.9902
Epoch 100/100
23/23 [==============================] - 0s 1ms/step - loss: 0.0400 - accuracy: 0.9830
```

Fig. 7. Algorithm Accuracy

5 Conclusion

The decision support system for predicting breast cancer aids and assists physi-cians in making the best, most accurate, and quickest decisions possible, as well as lowering total treatment costs.By predicting breast cancer at an early stage, the suggested method signi cantly lowers treatment costs and enhances the quality of life.By using the Arti cial neural networks we have been able to get a whooping accuracy of 99 percent on a given dataset which is a great out-come in terms of science and invention and will help us to have fewer patients dying with breast cancer.

Future work

Furthermore, we can use this dataset to do a comparative approach to show how our algorithm used with the help the ANN tops all of the other methods out. Secondly, we can try this algorithm on a di erent dataset to know how does it perform and what problems it faces during the testing of the model.This research may aid in the development of more e ective and reliable illness prediction and diagnostic systems, which will aid in the development of a better healthcare system by decreasing overall costs, time, and death rates.

References

1. Akay MF. Support vector machines combined with feature selection for breast cancer diagnosis. Expert systems with applications. 2009 Mar 1;36(2):3240-7.
2. Marcano-Cede~no A, Quintanilla-Dom nguez J, Andina D. WBCD breast cancer database classi cation applying arti cial metaplasticity neural network. Expert Systems with Applications. 2011 Aug 1;38(8):9573-9.
3. Polat K, Gunes S. Breast cancer diagnosis using least square support vector ma-chine. Digital Signal Processing. 2007 Jul 1;17(4):694- 701.
4. Kaya Y, Uyar M. A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease. Applied Soft Computing. 2013 Aug 1;13(8):3429-38.
5. Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A, Feuer EJ, Thun MJ. Cancer statistics, 2005. CA: a cancer journal for clinicians. 2005 Jan 1;55(1):10-30.
6. Yeh WC, Chang WW, Chung YY. A new hybrid approach for mining breast cancer pattern using discrete particle swarm optimization and statistical method. Expert Systems with Applications. 2009 May 1;36(4):8204-11.
7. Nahato KB, Harichandran KN, Arputharaj K. Knowledge mining from clinical datasets using rough sets and backpropagation neural network. Computational and mathematical methods in medicine. 2015;2015.
8. Liu L, Deng M. An evolutionary arti cial neural network approach for breast cancer diagnosis. InKnowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on 2010 Jan 9 (pp. 593-596). IEEE.

9. Chen HL, Yang B, Liu J, Liu DY. A support vector machine classi er with rough set-based feature selection for breast cancer diagnosis. Expert Systems with Applications. 2011 Jul 1;38(7):9014-22.

10. Saritas, Ismail. (2011). Prediction of Breast Cancer Using Arti cial Neural Net-works. Journal of medical systems. 36. 2901-7. 10.1007/s10916-011-9768-0.

11. Overview of Breast Cancer and Implications of Overtreatment of Early-Stage Breast Cancer: An Indian Perspective Gouri Shankar Bhattacharyya, Dinesh C. Doval, Chirag J. Desai, Harit Chaturvedi, Sanjay Sharma, and S.P. Somashekhar JCO Global Oncology 2020 :6, 789-798

12. Lin, YL., Xu, DZ., Li, XB. et al. Consensuses and controversies on pseudomyxoma peritonei: a review of the published consensus statements and guidelines. Orphanet J Rare Dis 16, 85 (2021). https://doi.org/10.1186/s13023-021-01723-6

13. Chang-Min Kim, Roy C. Park, Ellen J. Hong, "Breast Mass Classi cation Using eLFA Algorithm Based on CRNN Deep Learning Model", Access IEEE, vol. 8, pp. 197312-197323, 2020.

14. X. Zhang and Y. Sun, "Breast cancer risk prediction model based on C5.0 algorithm for postmenopausal women," 2018 International Conference on Se-curity, Pattern Analysis, and Cybernetics (SPAC), 2018, pp. 321-325, doi: 10.1109/SPAC46244.2018.8965528.

15. S. Modi and M. H. Bohara, "Facial Emotion Recognition using Convolution Neural Network," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1339-1344, doi: 10.1109/ICICCS51141.2021.9432156.

16. K. S. Bhangu, J. K. Sandhu and L. Sapra, "Improving diagnostic accuracy for breast cancer using prediction-based approaches," 2020 Sixth International Con-ference on Parallel, Distributed and Grid Computing (PDGC), 2020, pp. 438-441, doi: 10.1109/PDGC50313.2020.9315815.

17. A. Bharat, N. Pooja and R. A. Reddy, "Using Machine Learning algorithms for breast cancer risk prediction and diagnosis," 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), 2018, pp. 1-4, doi: 10.1109/CIMCA.2018.8739696.

18. P. Singhal and S. Pareek, "Arti cial Neural Network for Prediction of Breast Can-cer," 2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on, 2018, pp. 464-468, doi: 10.1109/I-SMAC.2018.8653700.

19. Rashmi G D, A. Lekha and N. Bawane, "Analysis of e ciency of classi cation and prediction algorithms (Na ve Bayes) for Breast Cancer dataset," 2015 Inter-national Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT), 2015, pp. 108-113, doi: 10.1109/ERECT.2015.7498997.

20. R. Shen, Y. Yang and F. Shao, "Intelligent Breast Cancer Prediction Model Using Data Mining Techniques," 2014 Sixth International Conference on In-telligent Human-Machine Systems and Cybernetics, 2014, pp. 384-387, doi: 10.1109/IHMSC.2014.100.