International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 5 ǁ May 2023 ǁ PP. 419-424

Breast Cancer Prediction UsingMachine Learning

Dr. M. Suresh Kumar[1], M. Madhumitha[2], R.S. Nivea[3], D. Sabaripriya[4]
[1]*Head of the Department, Department of IT*
[2,3,4] *Department of Information of Technology*
[1,2,3,4] *Sri Ramakrishna Institute of Technology*

Abstract: *Breast cancer is one of the most common diseases in women worldwide accounting for the majority of new cancer cases and cancer- related deaths. Early detection is essential for enhancing breast cancer outcomes and survival. Therefore, the creation of a website that can predict breast cancer is paramount in potentially saving the lives of women. This paper tells about the prediction of breast cancer using the Support Vector Machine (SVM) algorithm for text data that results in either benign or malignant (benign means not cancerous and malignant means cancerous) and the Convolutional Neural Network (CNN) algorithm for image data that results in cancerous or non-cancerous.*

Keywords*: Breast Cancer, Benign, Malignant, Support Vector Machine, Convolutional Neural Network.*

Date of Submission: 02-05-2023 Date of acceptance: 14-05-2023

I. INTRODUCTION

The most typical type of cancer among women is breast cancer. Breast cells that develop a typically are cancerous. To overcome this problem this system will classify the patient's tumor into benign and malignant using classification algorithm. If common malignancies are treated early, death can be avoided. Early detection of breast cancer can result in successful therapy and reduces the death rate.

Machine Learning (ML) is a field of study which gives computers a capability to learn without being explicitly programmed. machine adapts to the user based on the data. Machine learning is the subset of Artificial Intelligence which uses different algorithms for building mathematical model, predicting results and so on. Machine learning algorithms can be broadly classified into three categories: supervised learning, unsupervised learning, and reinforcement learning.

Deep Learning, which is basically a neural network with three or more layers, is a subset of Machine Learning. Neural networks are a set of algorithms modelled after the structure and functioning of the human brain, and they are capable of learning to recognize patterns in data. In deep learning, these neural networks are typically composed of multiple layers of interconnected nodes, or neurons, which allow the network to learn increasingly complex representation of input data. By adjusting the weights and biases of the connections between neurons during the training process, the network can learn to identify patterns and make predictions about new data.

II. LITERATURE SURVEY

Gaurav Singh (2020) The main goal of this study is to develop a model for breast cancer prediction utilizing different machine learning classification methods, such as k Nearest Neighbor (kNN), Support Vector Machine (SVM), Logistic Regression (LR), and Gaussian Naive Bayes (NB) [1]. Additionally, evaluate and compare the accuracy, precision, recall, f1Score, and Jaccard index performance of the various classifiers. The implementation phase dataset will be divided into two phases: training and testing. Since the breast cancer dataset is publicly accessible on the UCI Machine Learning Repository, 80% of it will be used for training and 20% for testing before the machine learning algorithms are applied. In terms of all parameters, k Nearest Neighbors performed significantly better.

Muktevi Srivenkatesh (2020) In order to diagnose or raise awareness of breast cancer, this research suggests a prediction model to determine whether a person has the disease or not [4]. An accurate model to predict breast cancer disease is presented by comparing the accuracies of applying rules to the individual findings of the SVM, RF, NB, and LRon the dataset gathered in an area.

M. A. Mohammed, B. Al-Khateeb, A. N. Rashid, D. A. Ibrahim, M. K. A. Ghani, and S. A. Mostafa. This paper presents an effort to automate characterization of breast cancer from ultrasound images using multi-fractal dimensions and back propagation neural networks[5]. In this study, a total of 184 breast ultrasound images (72 abnormal (tumour cases) and 112 normal cases) were examined. Various setups were employed to achieve a decent balance between positive and negative rates of the diagnosed cases. The obtained results manifested in high rates of precision (82.04%), sensitivity (79.39%), and specificity (84.75%).

U. K. Kumar, M. B. S. Nikhil, and K. Sumangali, The main objective of this paper is to compare the results of supervised learning classification algorithms and combination of these algorithms using voting classifier technique [7]. Voting is one of the ensemble approaches where we can combine multiple models for the better classification. The dataset is taken from Wisconsin University database.

III. EXISTING SYSTEM

The existing system of the project has compared some widely used machine learning algorithms such as Random Forest, Decision Tree, K-Nearest Neighbor, Logistic Regression to the dataset for predicting breast cancer.

Breast Cancer is the most frequently diagnosed disease among women and its risk is getting higher year by year.

IV. PROPOSED SYSTEM

In this study, the Support Vector Machine (SVM) algorithm, which provides greater accuracy for text data than other algorithms, and the Convolutional Neural Network (CNN) algorithm, which provides greater accuracy for image data, are used to build the breast cancer prediction model.The Wisconsin Dataset and Breast Histopathology Images dataset were used in this project.

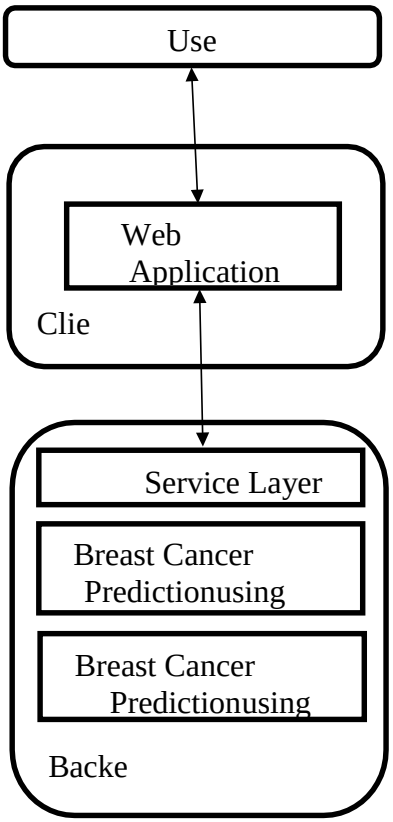

Fig 1: Flow Diagram

A. DATA COLLECTION

Data collection is the process of gathering data relevant to the project.

A collection of records is called a data set. Data sets can hold information such as medical records or insurance records, to be used by a program running on the system. The Wisconsin Dataset were used which has characteristics such as clump thickness, uniform cell size, uniform cell shape, marginal adhesion, single epithelial size, naked nuclei, bland chromatin, normal nucleoli, and mitosis for text data. Breast Histopathology Images dataset were used for image data.

B. PREPROCESSING

The process of getting a raw material that is suitable for a machine learning model is called data preprocessing. The data must be accurate and in the right format while developing a machine learning project in order to improve the performance of the model.

C. MODEL BUILDING

Convolutional Neural Networks (CNNs) are a type of neural network that are commonly used in image classification tasks. The architecture of a CNN consists of several layers, each with a specific function. The input layer receives the raw image data, and the subsequent convolutional layer applies a set of filters to extract local features. The ReLU activation function is then applied to the output of the convolutional layer to introduce non-linearity. The pooling layer is used to down sample the feature maps produced by the convolutional layer, and the dropout layer helps to prevent overfitting. The fully connected layer applies a set of weights to classify the input image, and the softmax activation function is used to obtain the probability distribution over the classes. The loss function measures the difference between the predicted probabilities and the true labels, and the optimization algorithm updates the weights of the model to minimize the loss. The output layer gives the final prediction for the input image based on the probability distribution over the classes. Overall, the architecture of a CNN is designed to extract useful features from the input image and classify it accurately.

Input Layer:

- The first layer of the CNN architecture isthe input layer, which receives the raw image data as input.
- The input layer has a fixed size that matches the size of the input image.

Convolutional Layer:

- The convolutional layer applies a set of filters or kernels to the input image to extract local features.
- The filters are learned during the trainingprocess and are used to detect specific patterns such as edges, corners, or textures

ReLU Activation:

- The output of the convolutional layer is passed through a Rectified Linear Unit (ReLU) activation function.
- The ReLU function adds non-linearity to themodel and helps to remove negative values from the output.

Pooling Layer:

- The pooling layer down samples the feature maps produced by the convolutional layer.
- The most common pooling operation is maxpooling, which takes the maximum value in each region of the feature map.

Dropout Layer:

- The dropout layer randomly drops out some of the neurons during training to preventoverfitting.
- The dropout layer helps to reduce the coadaptation between neurons and forces the model to learn more robust features.

Fully Connected Layer:

- The fully connected layer takes the output of the convolutional and pooling layers and applies a set of weights to classify the input image.
- The output of the fully connected layer is a probability distribution over the classes.

Softmax Activation:

- The softmax activation function is applied to the output of the fully connected layer to obtain the probability distribution over the classes.
- The softmax function ensures that the probabilities sum up to one and are between 0 and 1.

Loss Function:

- The loss function is used to measure thedifference between the predicted probabilities and the true labels.
- The most common loss function for classification problems is categorical cross entropy.

Optimization Algorithm:

- The optimization algorithm is used toupdate the weights of the model during training to minimize the loss function.
- The most common optimization algorithm is Stochastic Gradient Descent(SGD) with momentum.

Output Layer:

- The output layer gives the final predictionfor the input image based on the probability distribution over the classes.
- The output layer can have one or more neurons depending on the number of classes.

After several iterations and optimization techniques, I have implemented a Base CNN model that has shown promising results in image classification tasks. By tuning the hyper parameters and optimizing the architecture of thenetwork, I was able to achieve higher accuracy and better performance compared to previous versions of the model. The final version of the model, named as Final CNN model, incorporates several improvements, such as adding more layers and using different activation functions. I have also experimented with various techniques, such as data augmentation and regularization, to improve the robustness of the model. Overall, the Final CNN model has demonstrated superior accuracy and is suitable for a wide range of image classification tasks.

D. TRAINING

The process of training a model to make accurate predictions on new, unseen data is called training. In Machine Learning project the datasets are split into training and testing set.

The training set is used to evaluate its performance.Typically, the training set is much larger than the testing set because the model needs to be trained ona significant amount of data to learn the fundamental patterns in the data. The common splitis to use 70% of the data for Training and remaining 30% for testing.

Data which is used to train algorithm is called Training data and the data which is used to test aMachine Learning algorithm is called Testing data.

E. PREDICTION

Prediction is the final output of a project after trainedon a dataset using an algorithm.

V. RESULT

Fig: 2 Home Page

Fig:3 The Result shows that Patient has no BreastCancer.

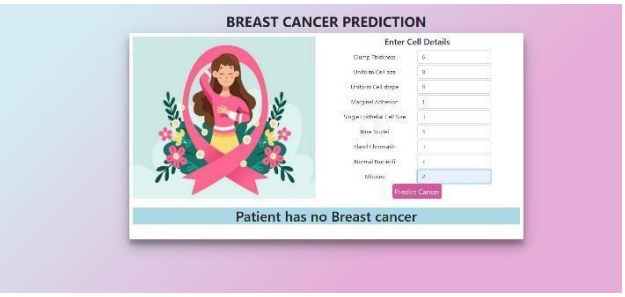

Fig: 4 Patient Details

Fig: 5 The Result shows that Patient has BreastCancer.

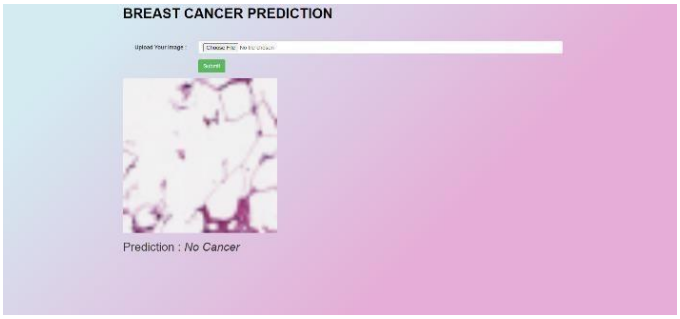

Fig: 6 The Result shows that No Cancer.

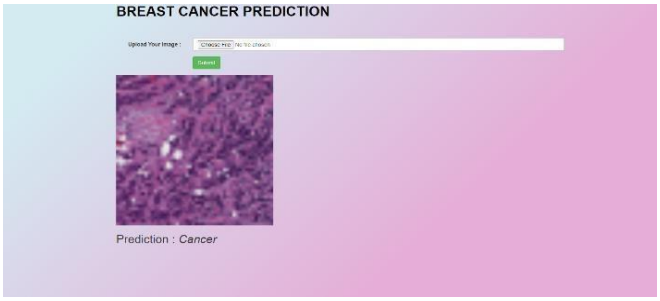

Fig:7 The Result shows that Cancer.

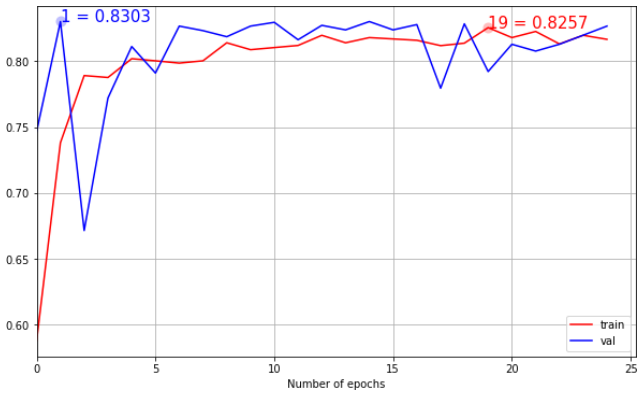

Fig: 8 Base CNN epochs

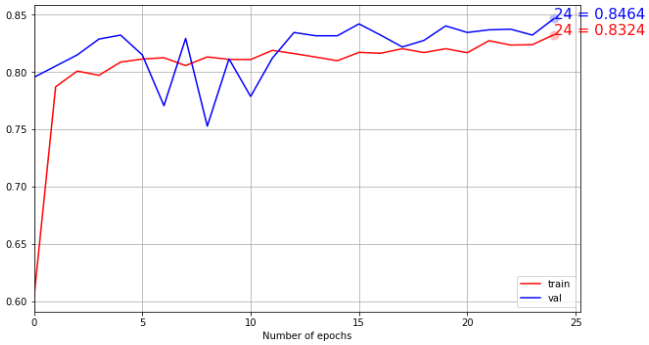

Fig: 9 Base CNN epochs

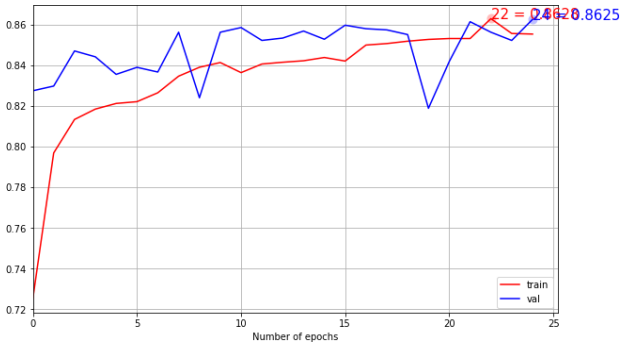

Fig: 10 Final CNN epochs

VI. CONCLUSION

Breast cancer is one of the major cause of death among women worldwide. Early detection and accurate diagnosis are crucial in improving the survival rate of patients. Machine learning and deep learning have emerged as promising techniques for breast cancer prediction and diagnosis. In recent years, numerous studies have been conducted to investigate the effectiveness of these techniques in breast cancer prediction. These studies have shown that machine learning and deep learning can achieve high accuracy rates in breast cancer diagnosis and prediction.

REFERENCES

[1]. Gaurav Singh, "BreastCancer Prediction UsingMachine Learning", International Journal of Scientific Research in omputer Science, Engineering and Information Technology, Volume 6,Issue 4, pp 278- 284,2020.

[2]. Ch. Shravya, K. Pravalika, Shaik Subhani, "Prediction of Breast Cancer Using Supervised Machine Learning Techniques", International Journal of Innovative Technology and Exploring Engineering(IJITEE), Volume-8 Issue- 6, April 2019.

[3]. Ritu Madhan, Shivani Kalariya, "Design and Development of ProstheticBrassieres for Breast CancerPatients", Journal ofScientific Research Instituteof Science, Banaras Hindu University, Varanasi, India,Volume 65, Issue 4, 2021.

[4]. Muktevi Srivenkatesh,"Prediction of Breast Cancer Disease usingMLAlgorithms", InternationalJournal of (IJITEE) ISSN: 2278-3075, Volume-9 Issue-4, 2020.

[5]. M. A. Mohammed, B. Al-Khateeb, A .N .Rashid, , D. A. Ibrahim, M. K. A. Ghani, and S. A. Mostafa, "Neural networkand multi-fractal dimension features for breast cancer classification from ultrasound images", Computers & Electrical Engineering, vol.70, pp.871- 882, 2018.

[6]. U. K. Kumar, M. B. S. Nikhil, and K.Sumangali, "Predictionof breast cancer usingvoting classifier technique," inProceedings of the 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM),pp. 108- 114, 2017.

[7]. Noreen Fatima, Li Liu, Sha Hong, Haroon Ahmed,"Prediction of Breast Cancer, Comparative Review of Machine Learning Techniques, and Their Analysis", IEEE Access , Volume 8 ,2020.

[8]. M. A. Mohammed, M. K. Abd Ghani, and N. Arunkumar, "Decisionsupport system for nasopharyngeal carcinomadiscrimination fromendoscopic images using artificial neural network," The Journal of Supercomputing, vol. 76,pp. 1086– 1104,2020.

[9]. Ch. Shravya, K. Pravalika, Shaik Subhani, "Prediction of Breast Cancer Using Supervised Machine Learning Techniques", International Journal of InnovativeTechnology and Exploring Engineering (IJITEE),Volume-8 Issue-6, April 2019.