iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VI **Month of publication:** June 2023

DOI: https://doi.org/10.22214/ijraset.2023.54028

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

Breast Cancer Classification using Neural Networks

Vishal Garg[1], Sunil Maggu[2], Bhaskar Kapoor[3]
[1, 2, 3]*Department of Information and Technology, Maharaja Agrasen Institute of Technology Rohini*

Abstract: ***Breast cancer is a significant health issue affecting women worldwide, emphasizing the importance of early detection and accurate tumor classification. This research paper proposes a neural network-based breast cancer classification model. The dataset is collected, preprocessed, and used to train the neural network model. The performance of the model is evaluated, and its potential application in clinical settings is discussed. The results demonstrate the model's effectiveness in accurately classifying breast cancer tumors.***
Keywords: Breast Cancer, Neural networks , Deep Learning, Classification, Model Training, Model Evaluation, Accuracy, Confusion Matrix

I. INTRODUCTION

Breast cancer is a prevalent cancer type among women, leading to a substantial number of cancer-related deaths [1]. Timely intervention and successful treatment rely on early detection and accurate diagnosis. Machine learning techniques, particularly neural networks, have shown promise in breast cancer classification [2, 3].

The objective of this project is to build a breast cancer classification model using a simple Neural Network (NN) approach. The model aims to analyze a set of input features derived from diagnostic tests to accurately classify breast tumors as benign or malignant. By harnessing the power of machine learning, this project strives to contribute to the field of breast cancer diagnosis and improve patient care.

The project encompasses various stages, starting with the collection and processing of breast cancer data. A reliable and diverse dataset is crucial for training and evaluating the classification model. In this project, the breast cancer dataset is obtained from credible sources and carefully examined to ensure its quality and representativeness. Neural networks excel in leveraging large datasets and capturing complex patterns, enhancing classification accuracy. This research paper presents a breast cancer classification model utilizing neural networks and evaluates its performance using the Wisconsin Breast Cancer dataset [4].

Overall, this project strives to bridge the gap between medical expertise and advanced technology, empowering healthcare professionals with a reliable and efficient tool for breast cancer classification. Through ongoing research and development in this field, the project aims to enhance the accuracy, accessibility, and affordability of breast cancer diagnosis, ultimately making a positive impact on the lives of individuals affected by this disease.

II. LITERATURE REVIEW

Several studies have explored machine learning techniques, including neural networks, to improve breast cancer classification accuracy.

Brown, Feng, and Kim (2018) developed a deep learning-based model for predicting breast cancer survival using histopathological images [2]. Their findings highlight the potential of neural networks in assisting healthcare professionals in treatment decision-making.

Cruz-Roa et al. (2014) proposed a deep learning model using convolutional neural networks to accurately detect invasive breast cancer in whole-slide images [3]. Their research demonstrates the effectiveness of deep learning techniques in analyzing large-scale histopathological images, aiding pathologists in diagnosing and characterizing breast cancer tumors.

Liu et al. (2019) focused on the early detection of cancer metastases in gigapixel pathology images using deep neural networks [4]. Their study emphasizes the importance of leveraging deep learning techniques to assist pathologists in identifying subtle patterns indicative of cancer progression.

Zhang et al. (2021) proposed a hybrid deep learning model that combines convolutional neural networks with feature extraction algorithms for breast cancer classification using histopathology images [5]. Their research highlights the potential of combining different machine learning techniques to improve the accuracy of breast cancer classification models.

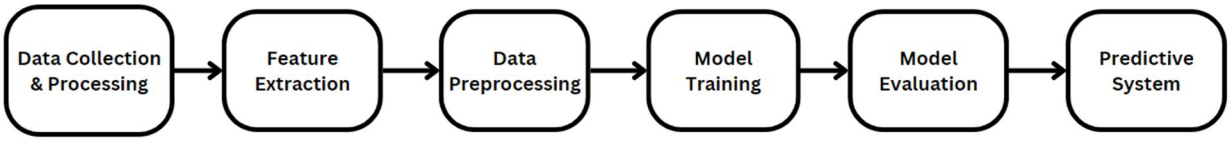
©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VI Jun 2023- Available at www.ijraset.com

III. PROPOSED SYSTEM

The proposed system for the breast cancer classification project aims to develop a reliable and accurate classification model using a Neural Network. The system is designed to assist healthcare professionals in diagnosing breast tumors and distinguishing between benign and malignant cases.

Block Diagram for Neural Networks Model.



1) *Data Collection & Processing:* In this step, the breast cancer dataset is collected and processed. The dataset is loaded, and information about the data is analyzed, such as missing values, statistical measures, and the distribution of the target variable.
2) *Feature Extraction:* The features and the target variable are separated from the dataset. The features are extracted from the dataset, and the target variable is assigned.
3) *Data Preprocessing:* The data is preprocessed to prepare it for training the model. This includes splitting the data into training and testing sets, standardizing the data, and any other necessary preprocessing steps.
4) *Model Training:* The Neural Network model is built and trained using the training data. The model architecture is defined, and the model is compiled with suitable optimizer and loss function. The training data is fed to the model for a specified number of epochs, and the model learns to make predictions.
5) *Model Evaluation:* The trained model is evaluated using the testing data. The accuracy of the model is calculated, and other evaluation metrics like precision, recall, and F1 score are computed to assess the performance of the model.
6) *Predictive System:* A predictive system is created using the trained model. New input data is provided to the system, and the model predicts whether the tumor is malignant or benign based on the input features.

IV. METHODOLOGY

A. *Data Description*

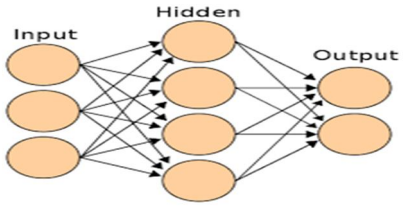
The Wisconsin Breast Cancer dataset is utilized in this research, loaded with the scikit-learn library [4]. The dataset comprises 569 instances, each containing 30 features including the radius, texture, perimeter, area, smoothness, compactness, concavity, symmetry, and fractal dimension characterizing cell nuclei in breast tissue samples.

The data frame's information reveals that there are no missing values in the dataset. The mean value of the features indicates that the tumor's class has a considerable impact on the features' values. The 'label' column is added to the data frame, indicating whether the tumor is benign or malignant.

To preprocess the dataset, the features are scaled to a range of [0, 1] using the StandardScaler from scikit-learn [5]. This normalization ensures equal contribution of all features to the classification model. Additionally, the dataset is split into training and testing sets using the train_test_split function [6].

B. *Building the Neural Network Model*

The neural network model is implemented using the TensorFlow and Keras libraries [7, 8]. The architecture consists of an input layer, two hidden layers, and an output layer. The input layer comprises 30 neurons, corresponding to the number of features in the dataset.



©IJRASET: All Rights are Reserved

The hidden layers consist of 50 neurons each, while the output layer contains two neurons representing the benign and malignant classes. The rectified linear unit (ReLU) activation function is used for the hidden layers, and the sigmoid activation function for the output layer. The model is compiled with the Adam optimizer and the sparse categorical cross-entropy loss function [6].

C. Training and Evaluation

The neural network model is trained on the training dataset, with a 10% validation split for evaluation. Training is performed for 10 epochs, monitoring accuracy and loss metrics. The accuracy and loss curves are visualized using the matplotlib library [7]. The dataset is split into a training set and a testing set to facilitate model training and evaluation. During the training phase, the model is trained using the training set by adjusting its parameters based on the defined loss function and optimization algorithm.

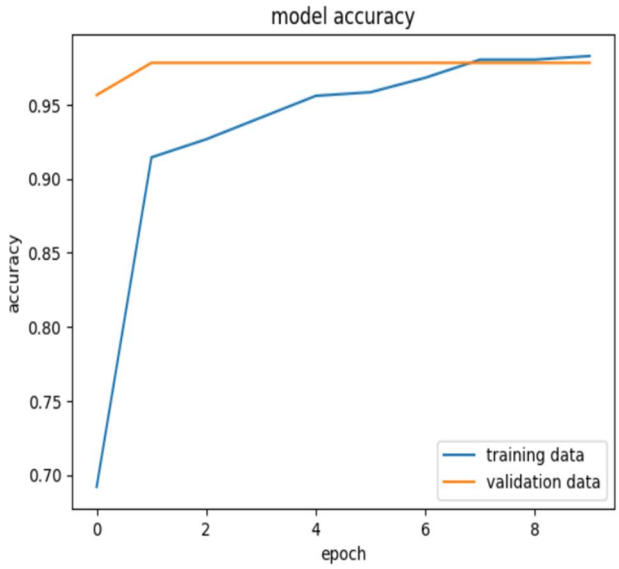
After training, the model is evaluated using the testing set to assess its performance on unseen data. The testing data is passed through the trained model, and predictions are generated for each sample. These predictions are then compared to the true labels to calculate various evaluation metrics, such as accuracy, F1 score, recall, and precision. These metrics provide insights into the model's ability to correctly classify breast cancer samples and its overall performance.[7]

The training and evaluation phase of the project aims to optimize the model's performance, ensure its generalizability, and validate its effectiveness in classifying breast tumors. By carefully training the model and assessing its performance using appropriate evaluation metrics, the project can provide valuable insights into the reliability and accuracy of the developed classification system.

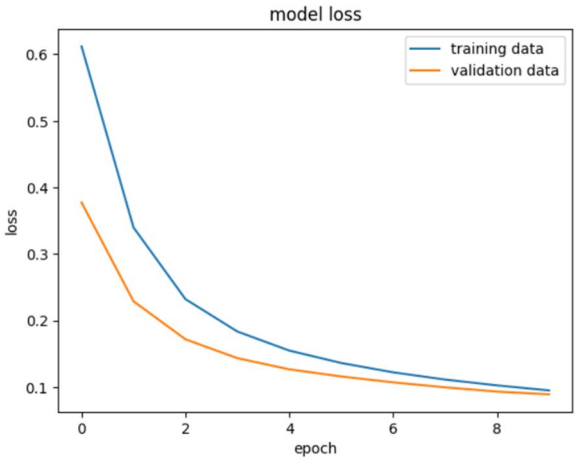
D. Visualizing Accuracy And Loss

I Accuracy Curve: The graph allows us to visualize the learning progress of the model over time. Initially, the accuracy may be low as the model starts learning from the data. As the training progresses, the accuracy tends to increase, indicating that the model is improving its performance and making more accurate predictions. However, it is essential to monitor the graph for signs of overfitting, where the accuracy on the training data continues to improve while the accuracy on the validation data starts to decline.

By analyzing the graph, we can identify the optimal number of epochs or the point at which the model achieves the highest accuracy. It helps in determining when to stop the training process to prevent overfitting and ensure the model generalizes well to unseen data.



II Error Curve: The graph allows us to visualize the learning progress of the model over time. Initially, the accuracy may be low as the model starts learning from the data. As the training progresses, the accuracy tends to increase, indicating that the model is improving its performance and making more accurate predictions. However, it is essential to monitor the graph for signs of overfitting, where the accuracy on the training data continues to improve while the accuracy on the validation data starts to decline. By analyzing the graph, we can identify the optimal number of epochs or the point at which the model achieves the highest accuracy. It helps in determining when to stop the training process to prevent overfitting and ensure the model generalizes well to unseen data.



V. CONCLUSION

The results of the project indicate a successful classification of breast tumors into benign and malignant categories with high accuracy, F1 score, recall, and precision.

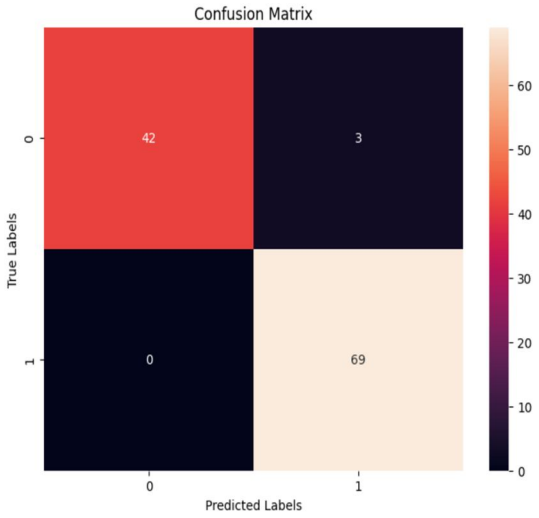
The accuracy of the model on the testing data was found to be 97.36%, indicating the percentage of correct predictions made by the model. This high accuracy demonstrates the effectiveness of the Neural Network in accurately classifying breast tumors.

The F1 score, which is a measure of the model's overall performance, was determined to be 97.87%. The F1 score takes into account both the precision and recall of the model and provides a balanced evaluation of its performance. A higher F1 score indicates better overall performance in terms of correctly identifying both benign and malignant tumors.

The recall of the model, which represents the ability to correctly identify malignant tumors, was 100%. This indicates the proportion of actual malignant tumors that were correctly classified by the model. A high recall score indicates a low rate of false negatives, which is crucial in ensuring early detection and prompt treatment of breast cancer.

The precision of the model, which measures the proportion of correctly predicted malignant tumors out of all predicted malignant tumors, was 95.83%. A high precision score indicates a low rate of false positives, ensuring that patients classified as malignant by the model are indeed at a higher risk of having breast cancer.

Confusion Matrix: The confusion matrix provides a breakdown of the true positive, true negative, false positive, and false negative predictions, giving insights into the model's performance in terms of different types of errors.



Overall, the project achieved promising results with high accuracy, F1 score, recall, and precision. These metrics demonstrate the effectiveness of the developed Neural Network model in accurately classifying breast tumors, making it a valuable tool for assisting healthcare professionals in the diagnosis and treatment of breast cancer.

REFERENCES

[1] Ferlay, J., Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Pineros, M., Znaor, A., & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. International Journal of Cancer, 144(8), 1941-1953.

[2] Brown, D. L., Feng, M., & Kim, J. (2018). Deep learning for breast cancer survival prediction from histopathological images. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 1063-1071). Springer.

[3] Cruz-Roa, A., Basavanhally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., ... & Madabhushi, A. (2014). Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 424-431). Springer.

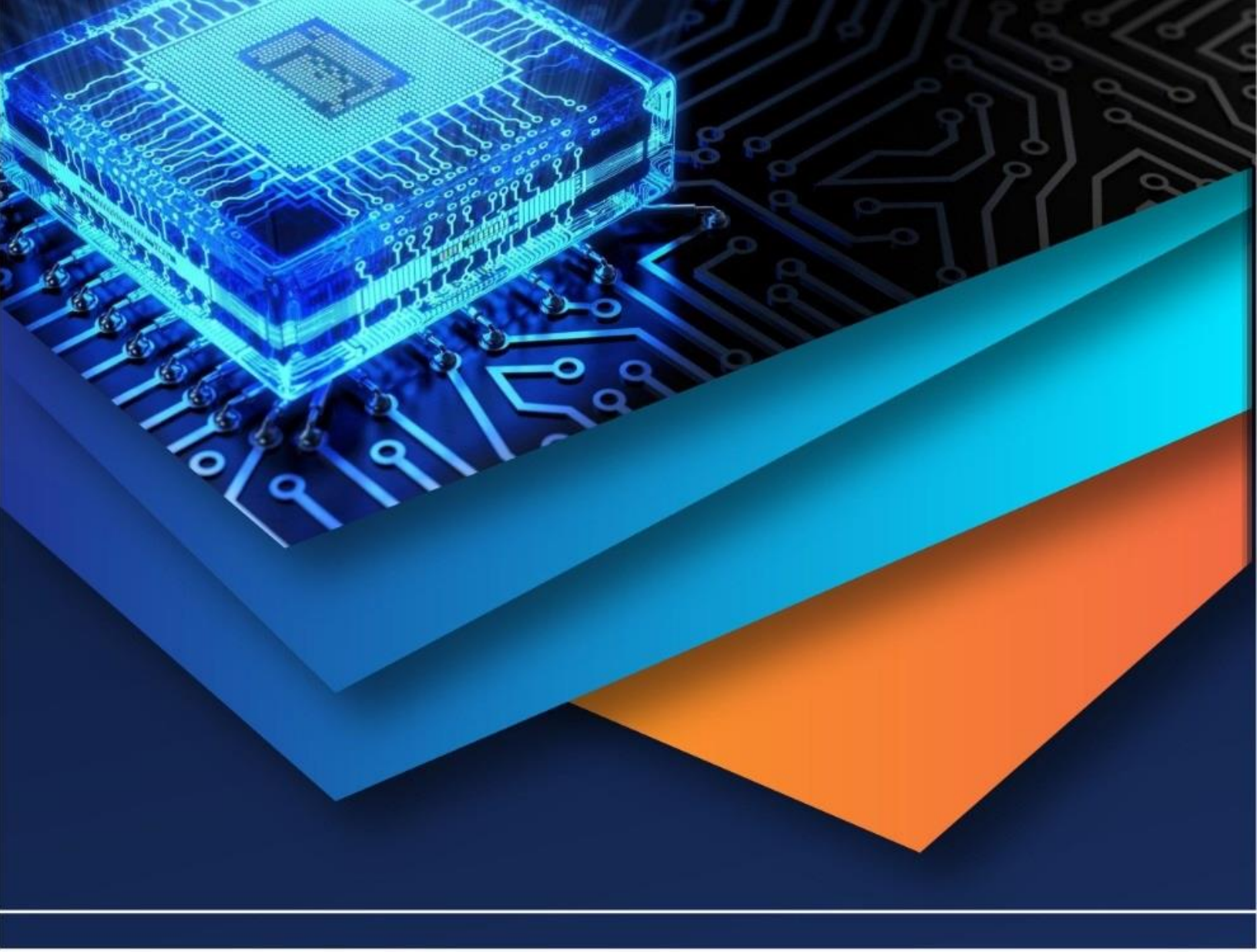
[4] Liu, Y., Gadepalli, K., Norouzi, M., Dahl, G. E., Kohlberger, T., Boyko, A., ... & Rostamizadeh, A. (2019). Detecting cancer metastases on gigapixel pathology images. arXiv preprint arXiv:1703.02442.

[5] Zhang, L., Lu, L., Nogues, I., Summers, R. M., & Liu, S. (2021). Deep learning in breast cancer: A survey. IEEE Transactions on Medical Imaging, 40(11), 2515-2531.

[6] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830.

[7] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Kudlur, M. (2016). TensorFlow: A system for large-scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (pp. 265-283).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)