Deep Learning Application in Predicting Breast Cancer

Oluwatoyin Stella Eleja, *Student, University of Wolverhampton,*

Abstract

The incidence of breast cancer is becoming more popular and it's a great contributor to the increasing mortality rate amongst women. It is regarded as a public health menace because statistics have attributed breast cancer to most cancer-related deaths. Our datasets would be gotten from the Wisconsin Breast Cancer Dataset found in the repository of Kaggle containing 569 rows and 30 features, thereafter, pre-processed for scaling and removal of irrelevant data. The training and testing datasets are obtained and the model implementation uses convolution neural network and artificial neural network algorithms to predict breast cancer. The research will further differentiate and compare with other machine learning algorithms to determine for highest accuracy, precision and detection.

I. Introduction

Breast cancer is associated with the presence of abnormal malignant cell growth in the breast and if left undetected the cancer spreads and metastasize within the body. World cancer research fund international in it recent statistics has revealed breast cancer to be the most widely recognized cancer in women globally [1]. In a report approved by the Cancer net editorial board Jan 2022 under ASCO, gave an estimate of over two million new cases of breast cancer in the year 2020 and it now surpasses lung cancer as the most commonly diagnosed form of cancer.[2], also approximately 41,000 deaths due to breast cancer was estimated for 2018 in the USA[2]. Therefore, it is imperative to detect it early enough because it is difficult to manage the highly invasive stages of breast cancers. Breast cancer can be detected in several ways the most common ones being the mammography, tissue biopsy, Magnetic resonance imaging, ultrasound and CT of the breast etc. In biopsy[3], a fine needle aspiration is used to remove the affected cell and sent to the pathological laboratory, however taking the sample can cause excruciating pain. The use of mammography is the commonest type for medical diagnosis and it produces the 2D image projection of the affected breast[4]. In our research paper, we are applying deep learning techniques to automatically detect cancer cells early enough and to achieve a high accuracy from our Wisconsin filtered dataset which is a complex dataset. The motivation of our research is based on a thorough literature review on the applications of various deep and machine learning technique in cancer research and diving deeply into the concepts of CNN which increases predictive performance by reducing bias and variance. This model extracts features from different modalities and feed to the stacked layer [5] [6]. This technique algorithm produces more features from any hidden inputs. The rest of the paper is organized as follows; Section II depicts the Background and literature review of the proposed work. Section III presents the research aim and questions, IV describes the proposed methodology and implementation and V discusses the summary and expected outcomes and conclusion of the proposed work is found in VI.

May 11, 2022

II. Background and Literature Review

a) Concept of Deep Learning Network: Deep learning also known as deep neural network newly developed as a branch of machine learning which is advantageous in the areas of computational power, improved model structure and exponential ability to capture complex data. It has some benefits over machine learning as it can use both supervised and unsupervised algorithms to learn new features[7]. Its architecture can identify, classify, and extract features[8]. DNN contains artificial neurons in multiple layers that mimic biological neurons of the human brain. By using activation function such as relu or sigmoid, more mathematical relationship is extracted from the input data. At the moment, the CNN model is the most popular application in deep learning[8].

b) Literature Review : This section compares and reviews related studies that has been carried out utilizing some machine learning and deep learning techniques. In [9] study to provide better training and testing report in breast cancer prediction using SVM and K-Nearest Neighbours techniques; an accuracy of 98.5 and 97.14 was shown with a specificity of 95.6 and 92.3 in testing phase. Similarly, [10] study on the comparative analysis of breast cancer using data visualization and data machine learning techniques such as k-nearest neighbours, support vector machine, rotation forest, naïve Bayes, decision trees, random forest, decision trees and logistic regression model showed an accuracy of 98.1. CNN AND ANN techniques were applied to

O. Stella was with the Department of Data Science, University of Wolverhampton, UK,

increase the breast cancer prediction authenticity in this study [11] and the ANN model yielded 99.3 with CNN obtaining 97.3. Moreso, 98 accuracy was achieved with SVM in predicting breast cancer from utilizing 3D images as seen in this study[12]. The author[9], achieved 95.2 accuracy when he approached predicting breast cancer using two monitored machine learning program; K-nearest neighbours and support vector machine.

III. Research Aim

The core aim of this research is to design a deep learning neural network model to promptly predict breast cancer diseases in human. This research would attempt to answer the series of questions:

- Is this proposed model currently the best at predicting breast cancer.
- To what extent would the proposed model accuracy predict breast cancer?
- What features are most effective in predicting the disease and to understand the general trend?
- How robust does the proposed model compare with others in breast cancer detection?

IV. Research Methodology

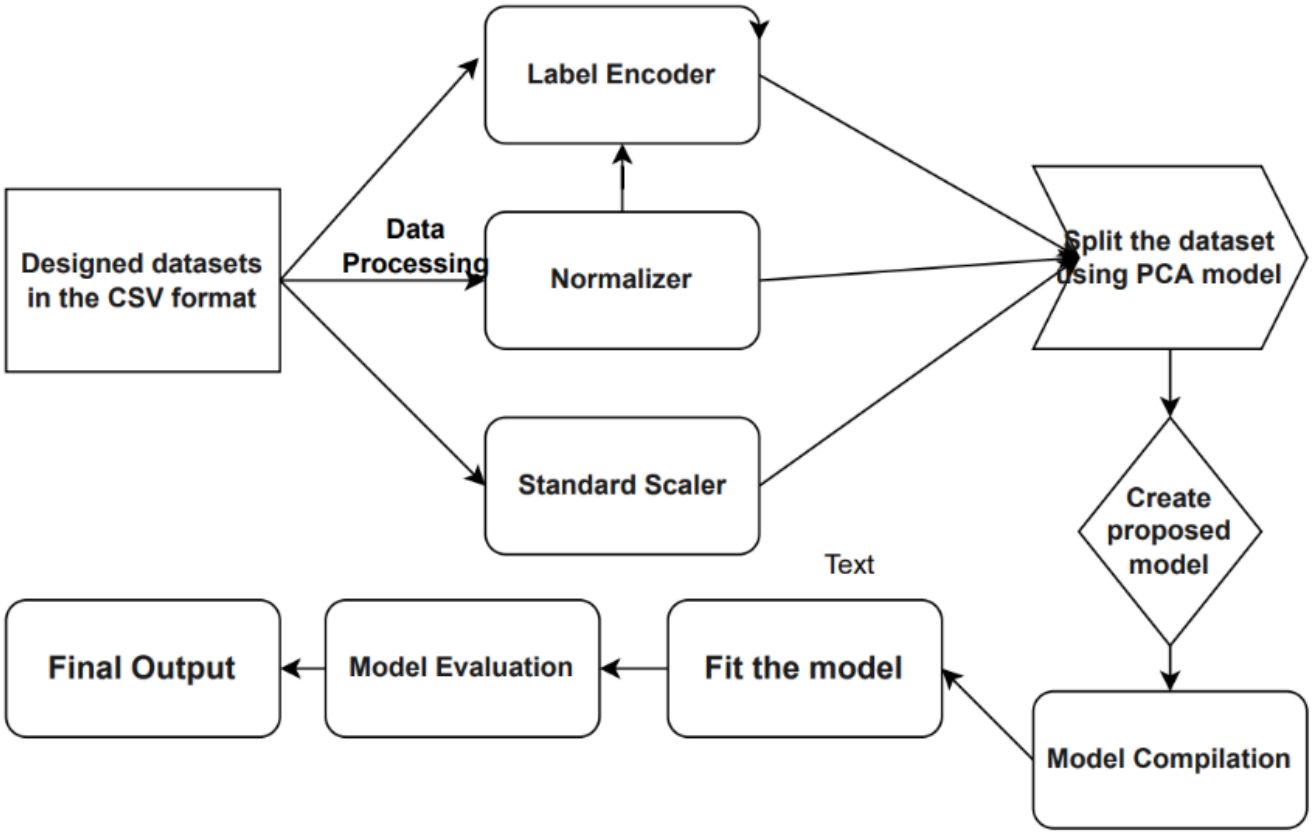

Fig. 1. The Figure 1 outlines the flow diagram of the proposed methodology

1) Data Collection: The image dataset of the proposed study that would be utilized is the Wisconsin breast cancer dataset available online for researchers and It is obtained from biopsies containing 569 rows and 30 columns[13]. Various deep learning algorithms can be used to detect breast cancer such as the artificial neural network, recurrent neural network and convolutional neural network [11]. In this study, convolutional neural network and artificial neural network would be used to predict breast cancer.

2) Pre-processing Data: Data pre-processing is a data mining model used to filter and coordinate data into useful formats[14]. The dataset needs to be tuned in a comprehensible format as it is not available according to our expectations. Data pre-processing is a proven method for solving such problems[14]. Standardization techniques are used to pre-process the BCI dataset[14]. The various standardization techniques would be implored.

1) Label encoder: This is suitable to encode the various stages of the categorical features numerically. It splits the datasets into testing and training sets and encodes labels with 0 as malignant and 1 as benign diagnosis. All the features are encoded. Neural network would be applied on the obtained dataset to make it relatively accurate.
2) Normalizer Method: This technique is useful when the data lacks even distribution. This process realigns the categorical features or attributes from 0 being the minimum value to 1 being the maximum value. Normalization process only works on the prepared label encoder's numeric dataset. To further achieve accuracy, we would apply neural network on the normalized dataset. The equation below would be used to rescale the attributes from the lowest to largest 0-1.

$$\frac{x_1}{\sqrt{x_1^2 + y_1^2 + z_1^2}} \tag{1}$$

3) Standard Scaler: this method of standardization is useful for Gaussian distribution datasets to rescale different categorical features so they have 0 as mean number and 1 as standard deviation. Following scaling the dataset, accuracy is tested to obtain the predict models. The equation to be utilized to achieve the scaler method is seen below.

$$\frac{x_i - mean(x)}{stdev(x)} \tag{2}$$

4) Principal Component Analysis: in other to achieve a multidimensional subspace category, principal component analysis would be carried out on the standardized and non-standardized data. The subspace feature ensures the axes variance. The variance enumerates the information size that relates to all the principal component. The principal component analysis contains the Eigen values which identifies the magnitude of the recent space, and vectors that predicts the direction of the recent features. The Eigen break-down on the covariance matrix E would be done using the PCA.
5) Model Implementation: in this stage some parameters in the convolution neural network will be used to obtain the trained model. The number of input, number of neurons, activation function and number of epochs would be considered[14]. Neural network algorithm is applied on a biological method. To arrive at the final output, some additional unit of bias and some less conspicuous layers are integrated. The activation effect is extracted then, then the final output. Two final algorithms would be integrated to give a high accuracy and prediction.
    a) A convolution neural network: this algorithm can obtain images and video recognition, classify images, analysis, segmentation, and assign biases and weights to the image object. The parameters that would be used are Number of neurons, 36 Con Layer, 64-Con layer 2, 30 Number of inputs, 50 Number of epochs and 50 activation function ReLU Sigmoid [15].
    b) Artificial neural network: models the neurons within a biological fluid and it is utilized to solve health challenges in the area of appropriate classification and detection using peculiar features like the adaptive self-learning ability, distributed storage and parallelism. The parameters to be used are 15 Number of neurons, 30 number of epochs and 50 activation function ReLU sigmoid.

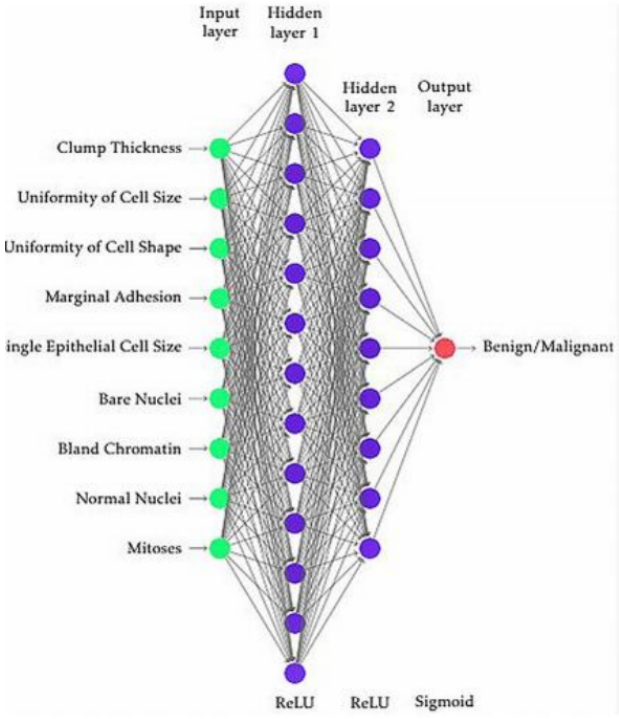

Fig. 2. The System Architecture of ANN [16]

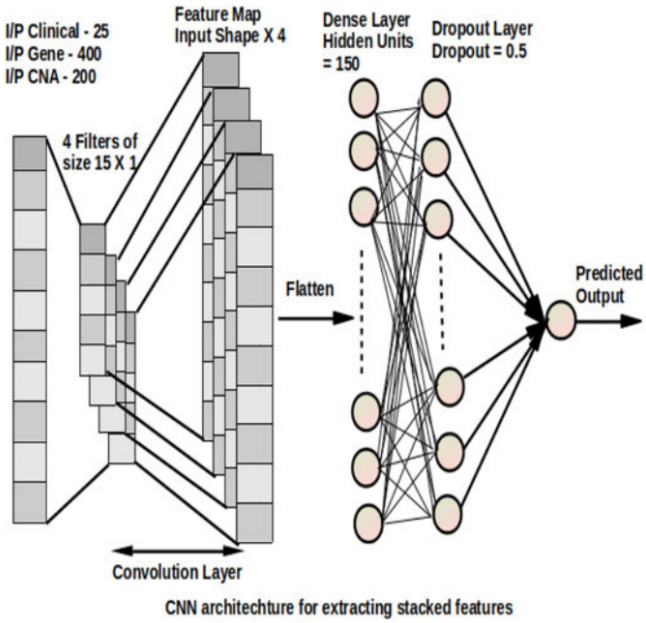

Fig. 3. The System Architecture of CNN [17]

V. Summary of Expected Outcome

In this proposed research, a model is expected to be delivered based on a deep learning technique known as the convolutional and artificial neural network. These models would be utilized to detect and predict the accuracy of breast cancer diseases. In this proposed research, a model is expected to be delivered based on a deep learning technique known as the convolutional and artificial neural network. These models would be utilized to detect and predict the accuracy of breast cancer diseases.

VI. Conclusion

The proposed research would develop deep learning methods such as convolutional neural network and artificial neural network to classify images dataset and predict breast cancer. By using ANN the percentage accuracy of the disease would be achieved and further research also hopes to compare with other machine learning models to determine highest precision, accuracy and detection. Challenges such as difficulty to separate area of interest from the complex dataset, different disease may manifest with similar symptoms and over-fitting of the model may occur in the course of the research. It is hoped that in the future, this model would be trained and improved to provide a 100% accuracy in detecting breast cancer.

REFERENCES

[1] "Breast cancer statistics World Cancer Research Fund International," WCRF International. https://www.wcrf.org/cancer-trends/breast-cancer-statistics/s/breast cancer-statistics, accessed on: Aug. 25, 2017

[2] Breast cancer statistics. [Online]. Available: http://www.wcrf.org/int/cancer-facts-figures/data-specific-ancers/breast cancer-statistics, accessed on: Aug. 25, 2017

[3] A. M. Ahmad, G. M. Khan, S. A. Mahmud, and J. F. Miller, "Breast cancer detection using cartesian genetic programming evolved artificial neural networks," Proceedings of the fourteenth international conference on Genetic and evolutionary computation conference - GECCO '12, 2012.

[4] F. M. Gody Rania Ahmed AbulSeoud, Amr M., "Genetic Algorithm and Firefly Algorithm in aHybrid Approach for Breast Cancer Diagnosis," International Journal of Computer Trends and Technology, Accessed: May 18, 2022. [Online]. Available: https://www.ijcttjournal.org/archives/ijctt-v32p111

[5] M. Tan, S.-P. Yuan, and Y.-X. Su, "Content-based similar document image retrieval using fusion of CNN features," arXiv:1703.08013 [cs], Aug. 2017, Accessed: May 18, 2022. [Online]. Available: https://arxiv.org/abs/1703.08013

[6] C. Szegedy et al., "Going Deeper with Convolutions," arXiv.org, 2014. https://arxiv.org/abs/1409.4842

[7] L. S. Yashaswini, H. U. Vani, H. N. Sinchana, and N. Kumar, "Smart automated irrigation system with disease prediction," IEEE Xplore, Sep. 01, 2017. https://ieeexplore.ieee.org/abstract/document/8392329 (accessed May 18, 2022).

[8] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," Procedia Computer Science, vol. 132, pp. 679–688, 2018, doi: 10.1016/j.procs.2018.05.069.

[9] Md. M. Islam, H. Iqbal, Md. R. Haque, and Md. K. Hasan, "Prediction of breast cancer using support vector machine and K-Nearest neighbours," 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), Dec. 2017, doi: 10.1109/r10-htc.2017.8288944.

[10] "Development of Intelligent Breast Cancer Prediction using Extreme Learning Machine in Java," International Journal of Computing, Communication and Instrumentation Engineering, vol. 3, no. 1, Jan. 2016, doi: 10.15242/ijccie.er0116114.

[11] M. Tiwari, R. Bharuka, P. Shah, and R. Lokare, "Breast Cancer Prediction Using Deep Learning and Machine Learning Techniques," papers.ssrn.com, Mar. 22, 2020.

[12] S. Nayak and D. Gope, "Comparison of supervised learning algorithms for RF-based breast cancer detection,". IEEE Xplore, Jun. 01, 2017.

[13] "Breast Cancer Wisconsin (Diagnostic) Data Set," www.kaggle.com. https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data

[14] N. Khuriwal and N. Mishra, "Breast Cancer Detection From Histopathological Images Using Deep Learning," 2018. Accessed: May 18, 2022. [Online]. Available: http://www.kresttechnology.com/krest-academic-projects/krest-mtech-projects/ECE/M-TECH

[15] M. Tiwari, R. Bharuka, P. Shah, and R. Lokare, "Breast Cancer Prediction Using Deep Learning and Machine Learning Techniques," papers.ssrn.com, Mar. 22, 2020.

[16] "Proceedings of the 3rd International Conference on Integrated Intelligent Computing Communication Security (ICIIC 2021) — Atlantis Press," www.atlantis-press.com. https://www.atlantis-press.com/proceedings/iciic-21 .

[17] N. Arya and S. Saha, "Multi-Modal Classification for Human Breast Cancer Prognosis Prediction: Proposal of Deep-Learning Based Stacked Ensemble Model," IEEE/ACM transactions on computational biology and bioinformatics, vol. 19, no. 2, pp. 1032–1041, Mar. 2022, doi: 10.1109/TCBB.2020.3018467.