Prediction and Classifications of Breast Cancer using Enhanced Convolutional Neural Network Approaches

Sathishkumar J[1], Dr. Venkatasalam K[2]

[1]Assistant Professor, Department of Information Technology, Kongunadu College of Engineering and Technology (Autonomous), Tiruchirappalli, Tamil Nadu, India.

[2]Associate Professor, Department of Computer Science and Engineering, Mahendra Engineering College (Autonomous), Mahendirapuri, Namakkal, Tamil Nadu, India.

Email: sathishrad@gmail.com[1], venkispkm@gmail.com[2]

Abstract

In worldwide women mortality increases extremely every year due to breast cancer and diagnosis of the issue through prediction is very much imperative for healthy lifespan. Here precision of cancer extrapolation is an essential thing for survivability of patient with appropriate treatment. Deep learning algorithms have materialised as influential tool for predicting breast cancer in medical image processing, which leverages capabilities of artificial neural networks (ANN) that are intended to mimic an architecture and functionalities of human brain. Superior features of convolutional neural network (CNN) in deep learning for handling image-based data like, exploiting spatial information, hierarchical feature learning, parameter sharing and data augmentation are important parameters in medical image processing. In this paper CNN algorithm is incorporated for predicting breast cancer in earlier and malignant stage, the results are compared with other deep learning algorithms and our proposed algorithm is expected to give better performance in parameters like accuracy testing, image classifiers, gene sequence classifiers and malignancy detection.

Keywords: Deep Learning, Convolutional Neural Network, Breast Cancer, Malignancy Detection, Gene Sequence Classifications.

1. Introduction

Artificial Intelligence plays an important role in all aspects of regular life and it gives its presence in medical field for predicting the severity of issue and accordingly medical treatment can be taken for diagnosing. Machine learning contributes in the process of predicting breast cancer in earlier stage and helps women to get healthy life. Various classification algorithms are used in [1] like K-Nearest neighbour (KNN), Random Forest, Logistic Regression, Decision tree and Support Vector Machine (SVM) for predicting, diagnosing breast cancer and identified that SVM outstripped rest of the classifiers and produced 97.2% highest accuracy. Biosensor and several machine learning approaches are investigated in [2] for earlier prediction of breast cancer, where machine learning doesn't require any pre-programmed models, it identifies patterns based on received data and forecasts outcomes by models and biosensor quantifies organic characteristics of body fluids and tissues. Analysing breast cancer using genetic level is more expensive and instead of that histopathological image analysis is common approach that frequently used for cancer detection and analysis. In [3] histopathological image is processed using deep learning and previous works are reviewed to detect the seriousness of the disease. That provides opportunities for researchers to forecast the status of patient using RNN and LSTM methods. Most research articles focus on accuracy matrix not on confusion matrix where it fails to distinguish false negative and false positive classifications, so it recommends future research should concentrate on confusion matrix for assessing the performance. Before introducing deep learning and machine learning approaches for detecting breast tumour by building models, a public data set were used for predicting and

detecting the severity of the issue and it gives maximum of 99% accuracy using convolutional neural network.For creating model, data should be analysed and followed by that it should be pre-processed with machine learning algorithms for measuring precision score, recall score and accuracy score. In [4] dataset is divided into three portions like training, authenticating and testing with 50 epochs. In each epoch the performance increases gradually and reaches 99% of accuracy using the convolutional model. Logistic regression learning is proposed for classification and sequential optimization. For performing automated diagnosis recursive feature elimination is used for feature selection and CNN as classifier exemplary also various algorithms were compared in [5] for accurateness and precision. Feature selection is the process of reducing ambiguity and reduce the complication of dataset, also reduces data size and training time. Embedded methods, filter and wrapper methods are commonly used feature selection methods, where wrapper method is used by recursive feature selection and in contrast filter – based method is used for identifying largest score. Here proposed scheme distinguishes malicious and compassionate tumour with earlier rate. Pre-processing is not mandate extracting features inevitably, when complicated classifier like CNN is used and more proficient because it screens significant parameters and work extremely well on image data. 10 cross field procedure is incorporated in [6] for identifying best classifiers in machine learning approaches using voting ensemble technique. This scheme gives maximum accuracy of 98.1 and minimum error rate of 0.01 is fashioned. Parameters like sensitivity, accuracy, precision and F- measures are compared for finding best classification efficiency using voting ensemble technique and at the end it produces low error rate. Medical advancements like mammograms, histopathological imaging, Magnetic Resonance Imaging (MRI) and computed tomography are used for detection of cancer in earlier and malignant state but it is error prone and cost effective. Characteristic of deep learning is, that can analyse and extract features using high dimensional correlated data, it has been utilized in [7] for the histopathological and radiographic images, Classifying breast cancer image modalities like ultrasound, mammography, thermography and MRI using deep learning is the main aim of the research with openly offered datasets. Methods like filters and deep learning simulations were incorporated in [8] to focus prominence of micro calcification in mammogram.Here filters are used for filtering standard and unusual mammogram descriptions and that riddledimageries are taken input and given to Yolov4 deep learning model for classifying standard and unusual mammogram images.To prove significance of filter in classifying images it were implemented and classified without using filter that gives significant difference in result. Yolov4 is a deep learning model that detects an object in frame in real-time by capturing 10% accuracy and 12% frames. As number of images increases, success ratio is increased using data augmentation technique that changes brightness of images, rotates and scaling the size of an image. In [9] deep learning scheme is applied to images by converting it as one dimensional data which can increases performance of accuracy in classification. In this work deep learning is applied for both histopathological images and genetic factor dataset, which is one dimensional dataset and it uses convex hull algorithm and disseminated stochastic neighbour technique. For enlightening performance proposed method uses the dataset and its decomposition where previous method access the dataset directly for classification. CNN base classifiers, dissimilarity mode decomposition are used for data decomposition along with wavelet transformation. Hybrid CNN based cancer detection is incorporated in [10] that reveals improved performance than base classifiers and threshold value, weight factor play vital role for hybridization. Unlike conventional deep learning models, high accuracy CNN performs well, not only for large dataset, even for smaller dataset too.

2. Related Work

Earlier stage of breast cancer identified through mammogram is mainstay of clinical screening that reduces mortality greatly, still it has well known limitations and nowadays magnetic resonance and ultrasound are adjunctive screening tools that supports for enlarged risk of breast cancer. In [11] various screening tools likemammography, ultrasound and magnetic resonance are discussed based on age, risk of the disease which tool is appropriate and how it can be used. Four machine learning schemes (Naïve bayes, LR, SVM and KNN) are compared using data mining approach in [12] with high-dimensional smaller dataset and result shows that LR provides optimistic scores in almost all metrics. Followed by that SVM and KNN produces optimal results next to LR consecutively, where naïve bayes produce poor results in all aspects. Addition to the ML classification techniques like naïve bayes, LR, KNN and etc are incorporated in [13] and ensemble methods like XGBoost, Adaboost on cancer are evaluated using various presentation procedures. Even though numerous schemes are already available for breast cancer using ML algorithms, still several obstacles are lagging in accuracy but this paper addresses this and concentrateson accuracy.Optimized neural network is proposed in [14] for giving accurate diagnosis scheme for medical representatives. An improved method is suggested in this work. Input, hidden, output, and dropout layers make up the recurrent neural network (RNN) used by Keras - Tuner. Hidden layer is optimizes number of neurons, rate values of dropout layer and three feature collection method is used for importing features from databases. Since manual detection of breast cancer is a time consuming process due to laborious work, improper classification and pathologist inaccuracies, those issues are addressed in [15] by incorporating hybrid deep learning for automated detection using entire images from kaggle dataset. Proposed method in this article combines convolutional neural network and GRU to detect breast cancer automatically and validation test is done using quantitative results. Here various parameters like accuracy, sensitivity, precision and specificity that gives the results like 86.2%, 85.50%, 85.60% and 84.71% are respectively. Efficiency of the proposed results are compared with CNN-LSTM and BiLSTM machine learning approaches, then results shows that hybrid model gives better performances. For detecting cancer thermography imaging is one of the effective scheme in infrared technology, where in [16] fully automated cancer detection system is incorporated. Firstly U-Net networking scheme is used to isolate the cancer area from rest of the body by engendering affected area as noise in detection model. Secondly, two class DL model which is accomplished from scratch for classification of usual and unusual breast tissues from thermal imageries. That is used to excerpt characteristics from dataset that helps in training network and improve efficiency of classification processes.For dense-breasted women, identifying breast cancer using mammography screening fails to identify the defected tissue. Comparison is done in [17] between the performance of mammography and ultrasound plus mammography in dense-breasted women breast cancer. The result gives double the time higher than mammography in ultrasound mammography. Rapid development of deep learning involves more in medical sciences, particularly in breast cancer detection and diagnosis. In [18] screening mammograms is used for breast cancer detection consuming an endwise training model, itprofessionally influences exercise datasets which includes whole medical comment of whole image. Here for initial stages lesion annotations are needed, where as in subsequent stages image-level tags are required. Comparing to other methods, convolutional network schemes attained an excellent performance by incorporating screening mammograms. Cardinal databank for screening mammography on sovereign assessment of digitized film mammograms. Accuracy of single model is achieved with 0.88 and for four model it is achieved with 0.91 in sovereign test of digitized film mammograms from cardinal dataset. It is demonstrated that entire image of classifier accomplished using an endwise methodology on

digitized mammogram. This paper proposes high accuracy of dissimilar mammography daises and clutches far-fetchedpotential for enlightening experimentaltrappings to decrease false affirmative and false destructive mammogram broadcastingconsequences. Prompt diagnosis of breast cancer detection is need of the hour of the medical sciences to reduce complications and mortality of women. Recently ML ad DL gives maximum of its contribution to detect and diagnosis of breast cancer. In [19] system is developed, here patients breast cells are reported and instant solutions are obtained for beginning and malignant stage. Wisconsin dataset is incorporated for carrying out this work from UCI repository, also five machine learning classification algorithms were investigated and performs well obtaining accuracy of classification.

3. Methodology

Foremost purpose of this proposed work is to recognize an effective deep learning approach for detecting and diagnosing breast cancer in earlier and later stages which is a subset of machine learning algorithms. Supervised learning (SL), Unsupervised Learning (UL) and Reinforcement Learning (RL) are the major types of algorithms, whereas supervised learning are in need of dataset for predicting the test data with trained dataset with external supervision. It can be categorized into regression and classification. Unsupervised learning doesn't requires any external supervision, it is trained using unlabelled dataset and it supports for solving association and clustering related problems. Reinforcement learning is completely different from above mentioned algorithms it contains one agent and it acts with an environment, according to the feedback received it takes decision. Here Q-Learning is considered as main algorithm and it supports based on action and rewards.

3.1 Deep Learning for Breast Cancer Detection

Deep learning is scope of machine learning that emphases on development and application of artificial neural networks (ANN) with several layers. It is stimulated by the construction and purpose of the humanoid brain and aims to enable computers to learn and make decisions autonomously without explicit programming. Deep learning algorithms have demonstrated promising results in predicting breast cancer risk by analysing various factors such as mammographic images, genetic data, and patient demographics. CNNs are commonly used in image analysis tasks. They learn to identify complex patterns within mammograms, such as micro calcifications, masses, and architectural distortions. By training on large datasets, CNNs can learn discriminative features indicative of breast cancer presence or absence. Such algorithms enable risk stratification, assisting clinicians in identifying high-risk individuals who may require further screening or preventive interventions. CNNs are particularly effective in analysing images and have been widely used in medical imaging tasks. They can learn complex patterns and features from mammograms or other imaging modalities to aid in breast cancer detection and diagnosis. CNNs consist of multiple layers of interconnected units, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers apply filters or kernels to the input images, which helps in capturing local patterns and features. The pooling layers reduce the spatial dimensions of the features, making the network more robust to translation and scale variations. Finally, the fully connected layers enable the network to make predictions based on the learned representations. CNNs for breast cancer prediction, a dataset of labelled images is required. This dataset typically consists of mammograms or other relevant imaging data, along with corresponding labels indicating the presence or absence of breast cancer. The CNN is trained on this dataset, where it learns to automatically extract discriminative features from the images and associate them with the cancer diagnosis. In above pictorial representation Figure 1 shows deformities of breast cancer using convolutional neural network using deep learning. Feature extraction, parameter sharing, spatial hierarchies, non-linearity and down samplings are key responsibilities of convolution layer in deep learning.
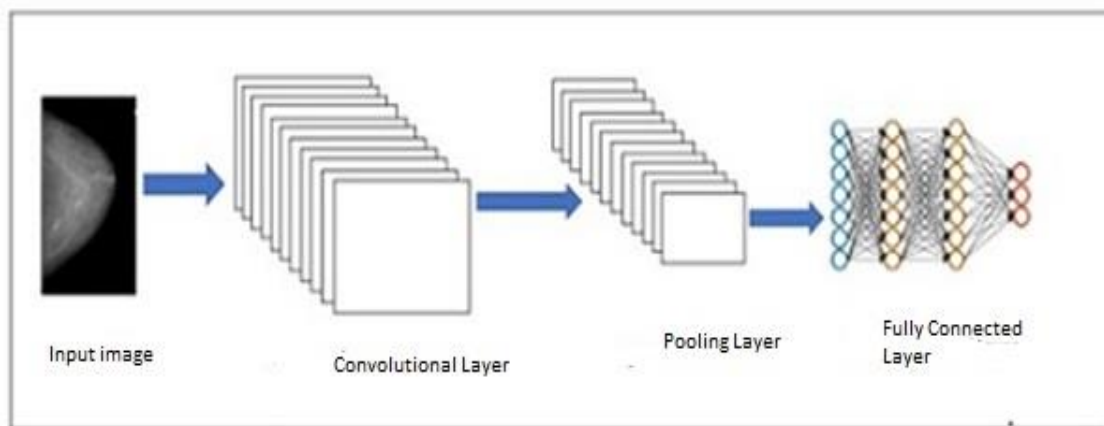
Figure 1 DLL Model for Classification of Deformities

Convolutional layers often include pooling operations, such as max pooling or average pooling, to downsample the spatial dimensions of the feature maps. Pooling helps reduce the computational complexity of subsequent layers, provides local translational invariance, and helps the model to focus on the most salient features. Dimensionality reduction, feature selection, translation invariance robust to noise and computational efficiency are the key responsibilities of pooling layer in deep learning. Pooling layer plays a crucial role in dimensionality reduction, translation invariance, feature selection, noise robustness, and computational efficiency within deep learning models. These responsibilities contribute to the effective extraction and representation of important features, enabling the network to learn and make accurate predictions. Parameter learning, classification, representation learning and learning complex relationships are responsibilities of fully connected layers in deep learning. Fully connected layer serves as the bridge between the preceding layers, responsible for learning complex relationships, introducing non-linearity, performing representation learning, producing the final output, and adapting the model's parameters during training.

Pseudoce for breast cancer predection using deep learning model

1. *Define the CNN architecture*

2. *Initialize the CNN model*

   $output = Conv2D(input, filters, kernel\_size, activation)$

3. *Add pooling layers for downsampling*:

   $output = MaxPooling2D(input, pool\_size)$

4. *Flatten the output to prepare for fully connected layers*:

   $output = Flatten(input)$

5. *Add fully connected layers with activation functions*:

   $output = Dense(input, units, activation)$

6. Add an output layer with appropriate activation function

   $output = Dense(input, 1, activation = 'sigmoid')$

7. *Add an output layer with appropriate activation function*

$$output = Dense(input, 1, activation = 'sigmoid')$$

8. *Compile the model with a suitable loss function and optimizer*:

$$model.compile(loss = 'binary\_crossentropy', optimizer = 'adam', metrics = ['accuracy'])$$

9. *Prepare the dataset*:

10. *Train the CNN model*:

$$for\ epoch\ in\ range(num\_epochs):$$

$$model.fit(train\_images, train\_labels, batch\_size, epochs)$$

$$loss = binary\_crossentropy(predicted\_labels, actual\_labels)$$

$$weights\_update = optimizer.update\_gradients(loss)$$

11. *Repeat the steps above until convergence level of performance is achieved.*

12. *Evaluate the trained model*:

$$accuracy, precision, recall, f1\_score = model.evaluate(test\_images, test\_labels)$$

13. *Save or serialize the trained model for future use*:

$$model.save(model\_filename)$$

The above pseud code illustrates the CNN algorithm that supports for predicting the occurrence of breast cancer in earlier and later stage. Initially we should define CNN architecture and model $model = Conv2D(input, filters, kernel\_size, activation)$ and it is mandatory to add pooling layers for down sampling. After flattening the output connection layers are activated with functions $output = Dense(input, units, activation)$. Then compiled using parameters like loss, optimizer and metrics. Followed by that dataset is prepared and it is trained using CNN model. With help of trained model, the same procedure is repeated until the convergence of performance is achieved. After achieving the accuracy in result, it is saved $model.save(model\_filename)$.

3.2 Data Pre-Processing

Predicting breast cancer using convolutional neural network involves medial image analysis such as mammograms and prediction is a complex task that requires extensive data pre-processing through various steps. Those steps are resizing, normalization, grayscale conversion, image filtering and image enhancement.

$$Resized\ Image = Resize(Image, width, height) \qquad (1)$$

$$Normalized\ Image = (Image - Min) / (Max - Min) \qquad (2)$$

$$Grayscale\ Image = 0.2989 * Red + 0.5870 * Green + 0.1140 * Blue \qquad (3)$$

$$Output(x, y) = \Sigma\ \Sigma\ \big(Input(i, j) * Filter(x - i, y - j)\big) \qquad (4)$$

$$Output(x, y) = CDF(Input(x, y)) - MinCDF) * (L - 1) / (M * N - MinCDF) \qquad (5)$$

In the above (1) image is resized using parameters like image, height and its weight, in (2) resized image is normalized with its min and max values, in (3) normalized image is converted as grayscale image. In (4) and (5) calculates the pre-processed image.

3.3 Feature Extraction

Using CNN for image processing feature extraction involves mining relevant and discriminative features of images. CNN learns hierarchical representation of visual data and patterns of images. For detecting special features of images CNN consists of multiple filters and it supports for computing element wise estimation. Complex patterns are enabled by introducing activation functions, here CNN includes rectified linear unit that replaces negative values with zero. Spatial dimensions are reduced by pooling layers and it supports for extracting more relevant information. Local Response Normalization (LRN) layers used to normalize responses within local region and promotes against adjacent sectors. Feature maps are visualized to understand patterns a specific layer, which helps to interpret network behaviour.

3.4 Image Classification

Image classifications are done in breast cancer prediction using CNN deep learning model by considering various parameters like dataset preparation, data pre-processing, model architecture, training, validation and hyper parameter tuning, interpretation and deployment, and testing are considered carefully for various influences like data quality, architecture selection and interpretation techniques. This technique analyses the model's predictions and investigate the learned features to gain insights into its decision-making process. The interpretation of mammography images can be improved by using interpretability approaches like gradient-based class activation maps (CAM) or occlusion sensitivity to pinpoint key areas that are crucial to the model's categorization. Once the model's performance has been validated, it can be used to classify breast cancer on fresh, previously unveiled mammography pictures.

4. Results and Discussion

In this paper for predicting breast cancer of patient in earlier and later stage we incorporated four deep learning algorithms CNN, LSTM, RNN and GAN. We considered parameters like accuracy testing, gene sequence classifiers, images data classifiers

and malignancy detection, comparing with above deep learning algorithms convolutional neural network performs well in all parameters mentioned above. Because CNN is having ability to capture spatial information, relevant feature from raw data is extracted automatically, CNNs were created primarily to take use of spatial hierarchy of visual data. Local structures and slight anomalies in mammography pictures are frequently used to detect breast cancer.
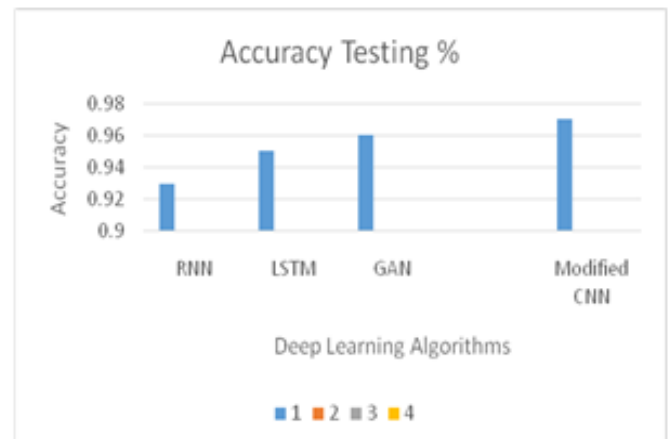

Figure 2 Accuracy Testing Vs Deep Learning Algorithms

Accuracy testing in predicting breast cancer using various deep learning algorithms are done in this paper and our results in Figure 2 shows that modified CNN gives more accuracy than other algorithms. Results 0.93%, 0.95%, 0.96% and 0.97% are given by the RNN, LSTM, GAN and modified CNN respectively.
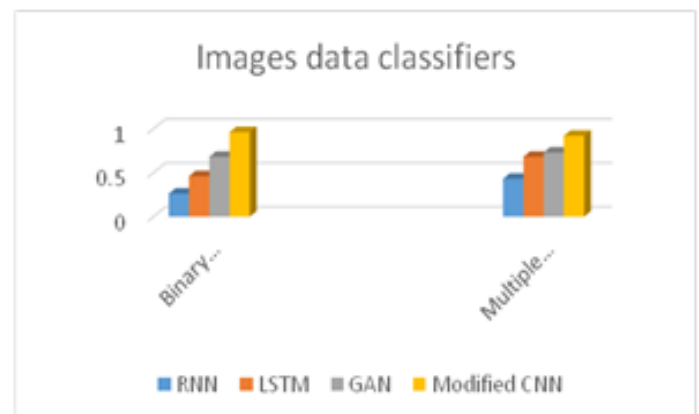

Figure 3 Image Classifiers vs Deep Learning Algorithms

The models used in Convolutional Neural Networks (CNNs) for classifying images into various groups or categories are known as image classifiers that consists of binary and multiple classification and it is shown in Figure 3 Because they have the capacity to automatically learn and extract pertinent characteristics from unprocessed picture data, CNNs are highly suited for image classification applications.
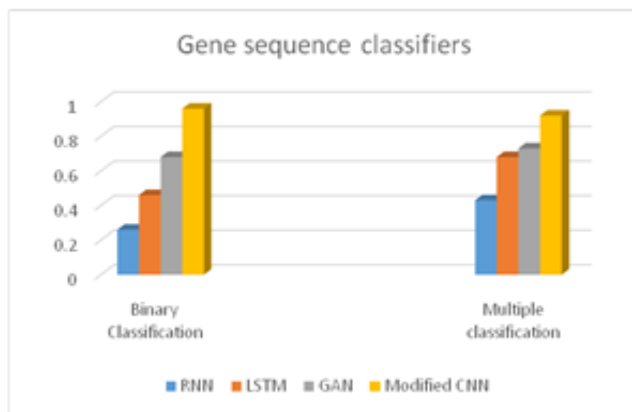

Figure 4 Gene Sequence Classifiers vs Deep Learning Algorithms

Deep learning gene sequence classifiers have produced encouraging results in a number of bioinformatics tasks. Gene sequence classification and analysis have both been effectively used to deep learning models including recurrent neural networks (RNNs), convolutional neural networks (CNNs), and their derivatives. Figure 4 shows that CNN performs well in both binary and multiple classifications.
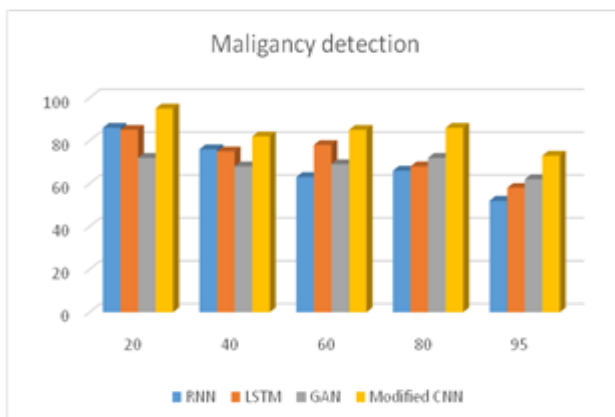

Figure 5 Malignancy Detection vs Deep Learning Algorithms

CNN have become a potent tool for medical image processing, especially when used to detect breast cancer malignancy. In order to accurately detect malignancy in mammography pictures, CNNs have special capacities for learning and extracting pertinent information. In Figure 5 it is identified that comparing to all algorithms CNN gives significant performance and it clarifies that when the number of images get increased detection of malignancy get reduced, still CNN performs well in all number of images.

Conclusion

The primary challenges in breast cancer detection and diagnosis an issue in earlier and malignant stage includes scheming examination pipelines, identifying tissue development, scheming preclinical schemes, handling complex image dataset, earlier detection and inculcating clinical schemes for prediction. For addressing these issues various machine learning and deep learning algorithms are introduced and, in this paper, we proposed CNN for predicting breast cancer in earlier and malignancy stage. Our result gives significant improvement in parameter like accuracy testing, image classifications, gene sequence classifications and malignancy detection with 0.97%, 0.92% in binary and multiple classifications, 0.94% in binary and multiple classifications and 94% respectively. Integrating multi model data, handling imbalanced datasets, incorporating clinical and genomic data are the future directions in breast cancer predictions and extensions of this work.

References

[1]. Naji, M. A., El Filali, S., Aarika, K., Benlahmar, E. H., Abdelouhahid, R. A., &Debauche, O. (2021). Machine learning algorithms for breast cancer prediction and diagnosis. Procedia Computer Science, 191, 487-492.

[2]. Amethiya, Y., Pipariya, P., Patel, S., & Shah, M. (2022). Comparative analysis of breast cancer detection using machine learning and biosensors. Intelligent Medicine, 2(2), 69-81.

[3]. Nassif, A. B., Talib, M. A., Nasir, Q., Afadar, Y., &Elgendy, O. (2022). Breast cancer

detection using artificial intelligence techniques: A systematic literature review. Artificial Intelligence in Medicine, 102276.

[4]. Khan, M. M., Tazin, T., ZunaidHussain, M., Mostakim, M., Rehman, T., Singh, S., &Alomeir, O. (2022). Breast Tumor Detection Using Robust and Efficient Machine Learning and Convolutional Neural Network Approaches. Computational Intelligence and Neuroscience, 2022.

[5]. Bhise, S., Gadekar, S., Gaur, A. S., Bepari, S., &Deepmala Kale, D. S. A. (2021). Breast cancer detection using machine learning techniques. Int. J. Eng. Res. Technol, 10(7).

[6]. Naji, M. A., El Filali, S., Bouhlal, M., Benlahmar, E. H., Abdelouhahid, R. A., &Debauche, O. (2021). Breast Cancer Prediction and Diagnosis through a New Approach based on Majority Voting Ensemble Classifier. Procedia Computer Science, 191, 481-486.

[7]. Dar, R. A., Rasool, M., & Assad, A. (2022). Breast cancer detection using deep learning: Datasets, methods, and challenges ahead. Computers in Biology and Medicine, 106073.

[8]. Yurdusev, A. A., Adem, K., &Hekim, M. (2023). Detection and classification of microcalcifications in mammograms images using difference filter and Yolov4 deep learning model. Biomedical Signal Processing and Control, 80, 104360.

[9]. Das, A., Mohanty, M. N., Mallick, P. K., Tiwari, P., Muhammad, K., & Zhu, H. (2021). Breast cancer detection using an ensemble deep learning method. Biomedical Signal Processing and Control, 70, 103009.

[10]. Sahu, A., Das, P. K., &Meher, S. (2023). High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets. Biomedical Signal Processing and Control, 80, 104292.

[11]. Lee, C. H., Dershaw, D. D., Kopans, D., Evans, P., Monsees, B., Monticciolo, D.,&Burhenne, L. W. (2010). Breast cancer screening with imaging: recommendations from the Society of Breast Imaging and the ACR on the use of mammography, breast MRI, breast ultrasound, and other technologies for the detection of clinically occult breast cancer. Journal of the American college of radiology, 7(1), 18-27.

[12]. Magboo, V. P. C., &Magboo, M. S. A. (2021). Machine learning classifiers on breast cancer recurrences. Procedia Computer Science, 192, 2742-2752.

[13]. Nemade, V., &Fegade, V. (2023). Machine Learning Techniques for Breast Cancer Prediction. Procedia Computer Science, 218, 1314-1320.

[14]. Saleh, H., Alyami, H., &Alosaimi, W. (2022). Predicting breast cancer based on optimized deep learning approach. Computational Intelligence and Neuroscience, 2022.

[15]. Wang, X., Ahmad, I., Javeed, D., Zaidi, S. A., Alotaibi, F. M., Ghoneim, M. E., ...&Eldin, E. T. (2022). Intelligent Hybrid Deep Learning Model for Breast Cancer Detection. Electronics, 11(17), 2767.

[16]. Mohamed, E. A., Rashed, E. A., Gaber, T., &Karam, O. (2022). Deep learning model for fully automated breast cancer detection system from thermograms. PloS one, 17(1), e0262349.

[17]. Kelly, K. M., Dean, J., Comulada, W. S., & Lee, S. J. (2010). Breast cancer detection using automated whole breast ultrasound and mammography in radiographically dense breasts. European radiology, 20, 734-742.

[18]. Shen, L., Margolies, L. R., Rothstein, J. H., Fluder, E., McBride, R., &Sieh, W. (2019). Deep learning to improve breast cancer detection on screening mammography. Scientific reports, 9(1), 12495.

[19]. Mehta, D., Mohite, A., Shinde, V., Khatri, R., &Dokare, I. (2022). Detection of Breast Cancer using Machine Learning Algorithms. Available at SSRN 4108758.