Breast Cancer Classification using Deep Convolutional Neural Network

Breast Cancer Classification using Deep Convolutional Neural Network

Muhammad Aqeel Aslam[1,a] Aslam and Daxiang Cui[1,b]

[1]Department of Instrument Science and Engineering, School of Electronics, Information and Electrical Engineering Shanghai Jiao Tong University Shanghai, China

Corresponding Author's E-mail: [a]maqeelaslam@sjtu.edu.cn; [b]dxcui@sjtu.edu.cn

Abstract. Over the last decade, the demand for early diagnosis of breast cancer has resulted in new research avenues. According to the world health organization (WHO), a successful treatment plan can be provided to individuals suffering from breast cancer once the non-communicable disease is diagnosed at an early stage. An early diagnosis of cure disease can reduce mortality all over the world. Computer-Aided Diagnosis (CAD) tools are widely implemented to diagnose and detect different kinds of abnormalities. In the last few years, the use of the CAD system has become common to increase the accuracy in different research areas. The CAD systems have minimum human intervention and producing accurate results. In this study, we proposed a CAD technique for the diagnosis of breast cancer using a Deep Convolutional Neural Network followed by Softmax classifier. The proposed technique was tested on the Wisconsin Breast Cancer Datasets (WBCD). The proposed classifier produced an accuracy of 100% and 99.1% for two different datasets, which indicates effective diagnostic capabilities and promising results. Moreover, we test our proposed architecture with different train-test partitions.

1. Introduction

Non-communicable diseases (NCD) such as breast cancer (BC) is one of the dominant factors of mortalities around the globe. BC is the most common cancer among females. According to the report of the International Agency for Research on Cancer (IARC), 367, 900 new cases against BC were registered in China during 2018 [1]. Breast cancer is ranked 4th commonest cancer around the world according to the mortality rate. There are several factors that are multifunctional and responsible for BC. These factors include diet, family history, hormones, obesity, reproduction of abnormal cells and even radiation therapy. According to the previous studies, BC is diagnosed at the advanced stage and this will result in almost half of the patients would die [2].

The Health risk is at ramp due to the fast-growing population, as the medical images are increasing exponentially day by day, and the traditional methods for BC classification have failed to meet the increasing demands for the medical images [3]. The Computer-Aided Diagnosis (CAD) method of diagnosis has a very high demand in medical applications. Pathologists in routine examine visually and navigate the entire pathological images to analyze and identify the abnormalities inside the medical image. Moreover, the diagnosis based on the clinical diagnostic methods requires a significant amount of time to determine whether the medical image is cancerous or non-cancerous. This process is very tedious and prolonged. In addition, the human eye is less adept to subtle changes in the tissue and

each medical professional has its own unique subjective mood of reading the fatigue. This may result in different diagnostic conclusions by different doctors about the same medical image. The diagnostic procedure of medical image highly depends on the human factor, which is not free from error. A patient may face severe results due to the doctor's minor mistake. Therefore, in this paper authors have proposed a CAD system for the BC classification. A good CAD system provides low false-negative rate and low false-positive rates [4].

Machine learning (ML) is a set of tools utilized for evaluation and creation of algorithms that demonstrate regression, classification and pattern recognition. The deep learning algorithm comprises of four steps: (1) collects the data, (2) select the best suitable neural network, (3) train the neural network using input data and parameters, (4) and evaluate the neural network on test data [5]. ML has been used for decades to identify the tumor and other malignancies for decades envisages the categorization of genes responsible of cancer and express the prognostic [6]- [7].

2. Related Work

Although machine learning has resolved very complex medical issues, there are still vacuum in early cancer diagnosis. Therefore, data scientists are trying to develop new research methods, which are easy to implement and can produce much high accuracy results to diagnose cancer at an early stage. In 2019, Kadam et al., demonstrated a method for Breast Cancer Classification (BCC) using Feature Ensemble Learning. They developed Stacked Sparse Autoencoder and Softmax Regression for the detection of benign cells to malignant cells. They achieved 98.6% accuracy for the Wisconsin Breast Cancer Dataset (BCD) [8]. Quite a lot of studies have been stated in the literature and are based on diverse approaches that could permit premature cancer investigation and prediction, which includes SVM [9], Decision Tress [10], Artificial Neural Network [11], Minimum Distance Classifier [12], Fuzzy Classifier [13], Fuzzy Rough Neural Network [14], Particle swarm optimization [15-18], microRNA and biomarkers [19-21], and Deep Learning approaches. In 2017, Litejens et al., presented a survey on deep learning in medical image analysis [22]. In literature, there are numerous methods have been developed for breast cancer classification, (BCC) using machine learning and deep learning. In 2018, Jasmir et al., presented a method for the detection of BC using deep learning [23]. In 2018, Rankhlin et al., demonstrated a technique for the BCC from the Histology Image Analysis using deep convolutional neural networks [24]. In 2016, Xu et al., proposed a method for the nuclei detection on breast cancer histopathology images using Stacked Sparse Autoencoder [25]. In 2001, Baldi et al., demonstrated a detection method for breast cancer using mammography images. However, this method is not sufficiency enough and misclassified almost 15% cases of breast cancer [26]. M. Akay et al [27]., demonstrated a method for the diagnosis of breast cancer and they achieved an overall accuracy of 99.51%, sensitivity 100%, and specificity of 97.91%. They used 80-20% train + validate, and test ratio partition. They used Ls-SVM method to identify breast cancer from the WBCD. In 2007, Übeyli et al [28]., also developed a breast cancer classification model using SVM, and they achieved an overall accuracy of 99.54%, however, they did not mention the specificity and selectivity values for their results. They used 37% data for training and validation purposes, whereas, 63% data was used to examine the performance of the classifier. In 2014, Dheeba et al., demonstrated Particle Swarm Optimized Wavelet Neural Network for the breast cancer classification [29]. They achieved 93.67% accuracy, the sensitivity of 94.17% and specificity of 93.1%. However, they did not mention the ratio for the partition of data. In 2015, Meet et al., developed ICA-RBFNN based neural network to distinguish breast cancer, they achieved an overall accuracy of 90.49% [30]. In 2015, Nahato et al., developed a breast cancer classifier with an overall efficiency of 98.86% [31]. In 2016, Ahmed et al., developed deep belief networks for the breast cancer classification. They evaluated their algorithm using different training + validate and test partition of the data [32]. They achieved an overall accuracy of 99.68% for DBN-NN, which was a semi-supervised method. This is the highest diagnosis's accuracy in the literature. In machine learning, Naive Bayes, and K-Nearest Neighbor are the most widely used algorithms to classify, and regression applications [33-35]. M. Amare et al., [36] diagnosed breast cancer with an overall accuracy of 97.5% with NB classifier, and they also achieved

96.1% accuracy with K-NN classifier. In 2015, Bhuvaneswariaa [37], proposed that K-NN algorithm used to evaluate the performance of cancer classification.

3. Proposed Methodology

3.1 Dataset

We used two different datasets to evaluate the performance of the proposed Convolutional Neural Network (CNN) classifier, which is developed for the BCC using a deep convolutional neural network. The first dataset contains the six ninety-nine (699) samples. Each sample corresponds to the eleven (11) attributes. Out of these 699 samples, sixteen samples contained missing values, so the samples which contain missing values, automatically discarded from the dataset. The actual number of samples for the second dataset was six eighty-three (683). The 11th feature column contains the binary responses of each clinical sample. Figure 1 shows the distribution of the benign class and malignant class for each dataset(s).

The second dataset contains 31 parameters. Out of these 31 parameters, we remove the first column, which contains the name/identity of the patients, and this information is irrelevant for the development of deep learning techniques. Whereas, the last column indicates the response of the data, whether the patient belongs to the benign class or belongs to the malignant class. The total number of observations for this labelled dataset was five hundred sixty-nine (569). This dataset was donated by the University of California, Irvine. In this dataset, there were 357 malignant clinical samples and 212 benign clinical samples. However, in the first dataset, there were 239 benign samples and 444 malignant samples. There were nine and thirty attributes present in first and second datasets, respectively, which were used to develop and evaluate the performance of the deep CNN architecture based model for the breast cancer classification.

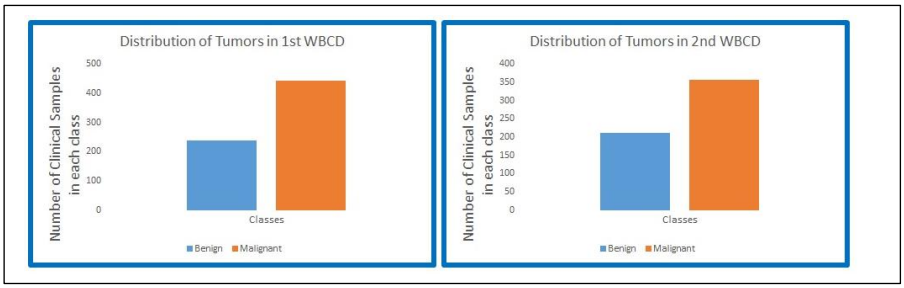

Figure 1. Distribution of Wisconsin Breast Cancer Dataset

3.2 CNN for Breast Cancer Classification

The dataset was fed as an input to the CNN in application to the breast cancer classification. After feeding the input, we trained the deep convolutional kernels in the proposed architecture of CNN. We used LeakyRELU [38] nonlinearity for the convolutional layers, which can be defined as:

$$f(x) = \begin{cases} x, if\ x > 0 \\ ax, otherwise \end{cases} \quad (1)$$

In general, the convolutional layer can be expressed as:

$$y^j = f(b^j + \sum_i k^{ij} + x^i \quad (2)$$

Here x^i, and y^i are the representing the i-th input map and j-th output map, respectively. The term b^j represents the bias parameter of the j-th map, * denotes the convolutional operation between the two functions, and b^{ij} shows the convolutional kernel used between the i and j maps.

After the convolutional layer, the max-pooling layer was introduced. In this layer, each neuron presents in the output map y^i pools over an s * s non-overlapping region in the input map x^i. Generally, the max-pooling layer is described as:

$$y_j^i = \max_{0 \le m \le s} \{x_{j.s+m}^i\} \quad (3)$$

The convolutional layers and max-pooling layers are fully connected and which are then connected to the Softmax classifier. The Softmax classifier contains the number of output classes equal to the number of the outputs. In our prosed architecture, we used ***tanh*** as a non-linear protocol in the connectivity of all the layers. The Softmax function operates as a function of squashing, and the k-dimensional input vector (dataset) is re-normalized, produces in the range of [1,2] of real values. Mathematically it can be shown as:

$$\sigma(z)_j = \frac{e^{zj}}{\sum_{k=1}^{K} e^{zk}} \text{ , } for\ j = 1,\ldots,\text{K} \quad (4)$$

Two types of error produced during the process of development of the ML technique. Training error is induced during the training phase of the neural network, whereas, the generalization error is produced during the testing phases of the proposed classifier. In deep learning, the training is often affected by the overfitting and under-fitting phenomenon. To reduce these issues in our proposed architecture for BCC, we applied batch normalization after each layer [39]. The dropout was introduced after the first fully connected layer [40]. The complete proposed architecture for the breast cancer classification can be viewed in figure 2.

3.3 CNN Training

In the proposed architecture we have two classes, which are benign class and malignant class. The following weighted loss function was used to train the proposed CNN classifier.

$$£(\omega, x_n, y_n) = -\frac{1}{N} \sum_{n=1}^{N} \alpha_n \sum_{k=1}^{K} t_{kn} \ln y_{kn} \quad (5)$$

Here x_n is the input vector, y_n is the prediction generated by the classifier for the n^{th} clinical input data, and t_n is the actual response of the n^{th} clinical sample. K represents the number of classes, and N is the total number of clinical samples.

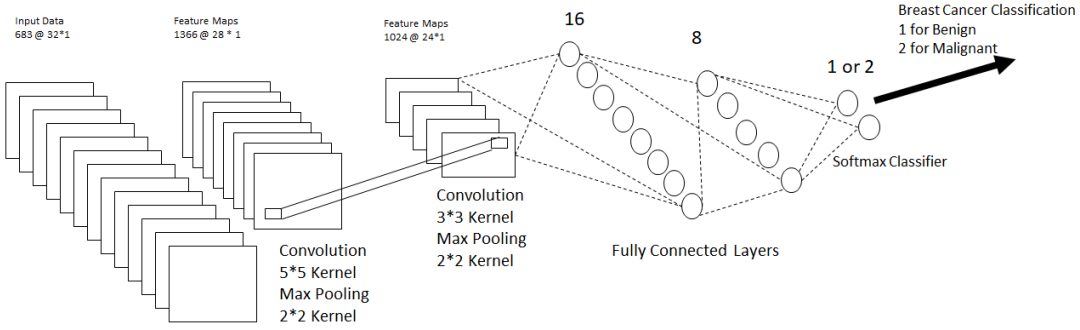

Figure 2. Proposed CNN Architecture for Breast Cancer Classification

3.4 Performance Evaluation of Proposed Architecture

The performance of proposed architecture on two datasets for breast cancer classification shown in table 1. The CNN classifier is evaluated by the metrics shown in (6) - (9), respectively. The performance of breast cancer classifier suing deep CNN architecture was expressed in terms of True Positive Rate (TPR) or recall, True Negative Rate (TNR), False Positive Rate (FPR), and precision.

True Positive (TP) is defined as the clinical samples, which were correctly diagnosed as benign by the developed classifier. True Negative (TN) is defined as the clinical samples, where the proposed classifier has correctly diagnosed the malignant clinical data. False-negative and false-positive are those instances, where the proposed architecture has misclassified the data, either into the benign class or malignant class. It indicates the error produced by the classifier. A good classifier is one that can diagnose all the samples correctly. However, if a model predicts correctly true negative samples but

unable to identify the true positive samples, such a model cannot be used in the clinics due to the uncertainty of the classifier. Therefore, it is very necessary that the developed classifier should have high accuracy, sensitivity, selectivity, and sensitivity values and such a model is not perfect for clinical usage.

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$FPR = \frac{FP}{FP+TN} \tag{8}$$

$$F-measure = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{9}$$

We also draw the Receiver Operating Characteristic (ROC) curves and precision-recall curves to evaluate the performance of breast cancer classification provided for different datasets.

4. Results and Discussion

4.1 Quantitative Analysis

Receiver Operating Characteristics (ROC) curves are an essential tool to visualize the performance of the developed classifier. It uses a graphical plot to demonstrate the diagnostic capability of the proposed neural network to distinguish the malignant breast cancer samples from the benign samples. This tool allowed us to select the best possible optimal neural network model for the BC classification. The ROC curves also illustrate how a considerable neural network-based model is skilled in distinguishing between classes. The total area under the ROC curve is 1, which indicates there is no misclassification in any class, and such models possess the perfect classification. The higher the Area Under Curve (AUC) values, the better the performance of the neural network. Figure 3 shows the Receiver Operating Characteristics (ROC) curves for the first dataset, using three different train + validate, and test partition data. Figure 3(a) is indicating maximum area under the curve, which shows that the proposed model has achieved 100% accuracy when tested on the test data. Whereas, Figure 3(b) indicates more area under the curve as compared to Figure 3(c), but less area under the curve when compared to Figure 3(a). Figure 3(c) is representing the least area under the curve, which indicates the performance of the classifier is affected by the misclassification. The proposed model has achieved minimum accuracy while testing at 57.54 - 42.46 % in three different sets of training + validation and test partition data. This result indicates that data divided into 73.3–26.7 % (train + validate, test) has a better in-sample average performance for the breast cancer dataset.

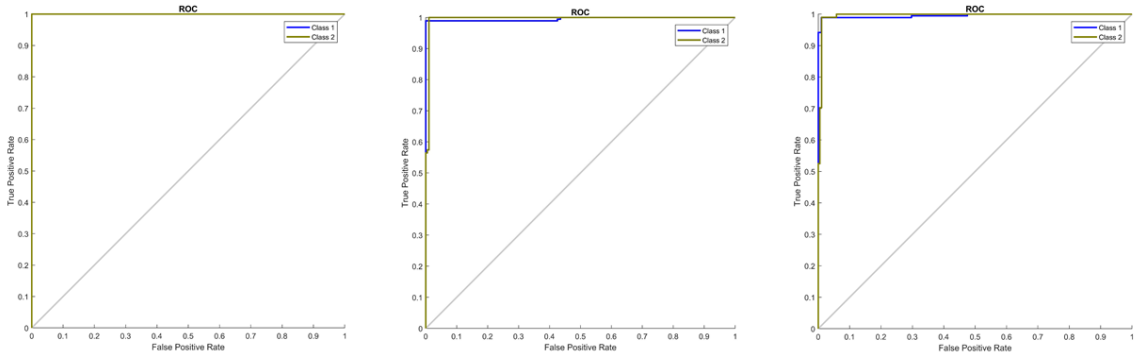

Figure 3. Receiver Operating Characteristics (FOC) Curve for 683 samples (1st Dataset) (A) 73.3 – 26.7 (%) Train + validate to test partition (B) 64.42 - 35.58 (%) Train + validate to test partition (C) 57.54 - 42.46 (%)Train + validate to test partition

Figure 4 represents the ROC curves for the second dataset. Figure 4(a) indicates the maximum area under the curve, while Figure 4(c) is showing the minimum area under the curve. Although, there is only a minute difference in the AUC values of all three data portioned datasets. However, the data

divided into 73.3-26.7% data has performed slightly better than the other two data partitions done for this study. This also indicates that the performance of the proposed model depends on the number of training samples, as the number of training samples increases, the performance of the classifier on the test increases. When the number of training samples is decreasing, the performance of the classifier is also deteriorating.

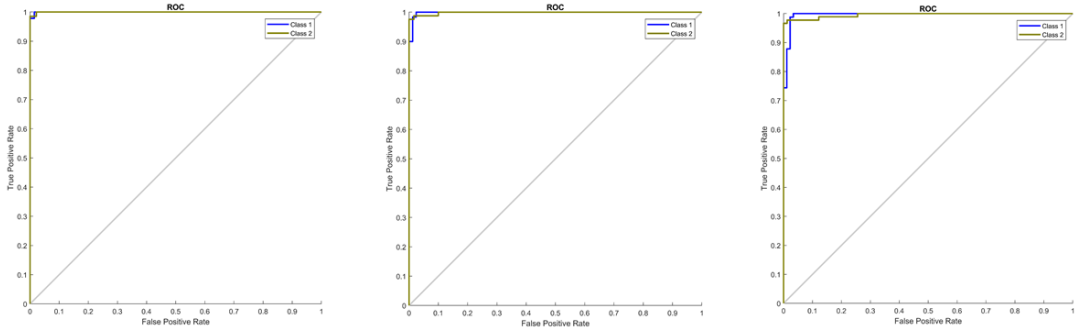

Figure 4. Receiver Operating Characteristics (FOC) Curve for 569 samples (2nd Dataset) (A) 80 – 20 (%) Train + validate to test partition (B) 75 – 25 (%) Train + validate to test partition (C) 70 - 30(%) Train + validate to test partition

4.2 Qualitative Analysis

Figure 5 and figure 6 represent the confusion matrices for the test data, using two different datasets as described in the previous sections. Figure 5(A) shows the result obtained from 73.3-26.7% data. In this experiment, 183 samples were selected as the test data, out of 683 samples, other samples were used for the training + validation purpose. Out of 183 samples, 115 samples belong to the malignant class and 68 samples belong to the benign class. The proposed classifier accurately distinguished all the benign and malignant samples, respectively. This model produced an overall accuracy of 100%, with a precision 100%, recall 100%, and the F-measure value also 100%.

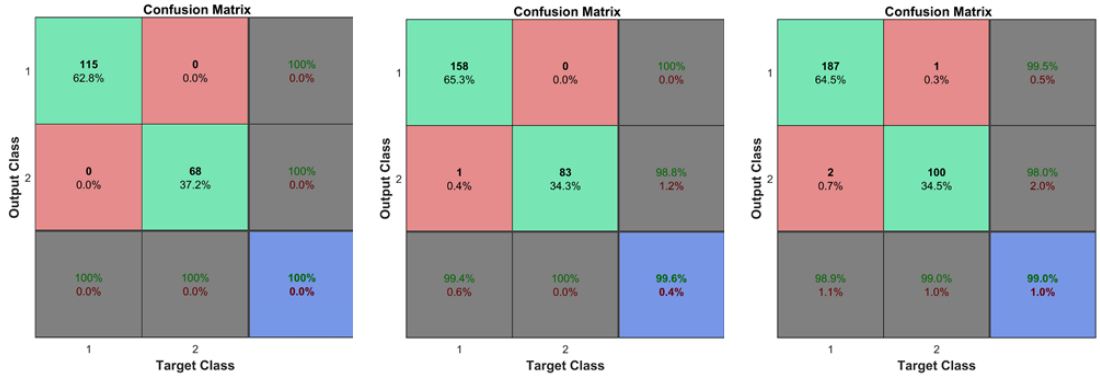

Figure 5 Confusion Matrices for 683 samples (1st Dataset) (A) 73.3 – 26.7 (%) Train + validate to test partition (B) 65 – 35 (%) Train + validate to test partition (C) 58 – 42 (%)Train + validate to test partition

In the second experiment, there were 242 samples used to test the performance of the classifier. Out of these 242 samples, 159 samples belong to the malignant class and 83 samples belong to the benign class. In this experiment, the proposed classifier classified all the benign samples, but one sample from the malignant class was misclassified. This model produced an overall accuracy of 99.6%. For this experiment, we obtained 99.37% value for precision, 100%, and 99.68% for the recall and F-measure parameters, respectively.

In the third experiment, we used 290 samples to evaluate the performance of the proposed classifier. Out of these 290 samples, 189 specimens belong to the malignant class and 101 samples belong to the

benign class. However, this model yields an overall accuracy of 99%. This model misclassified two and one samples from benign and malignant classes, respectively. For this experiment, the developed classifier has achieved 98.94%, 99.46%, and 99.19% for the precision, recall, and F-measure parameters, respectively.

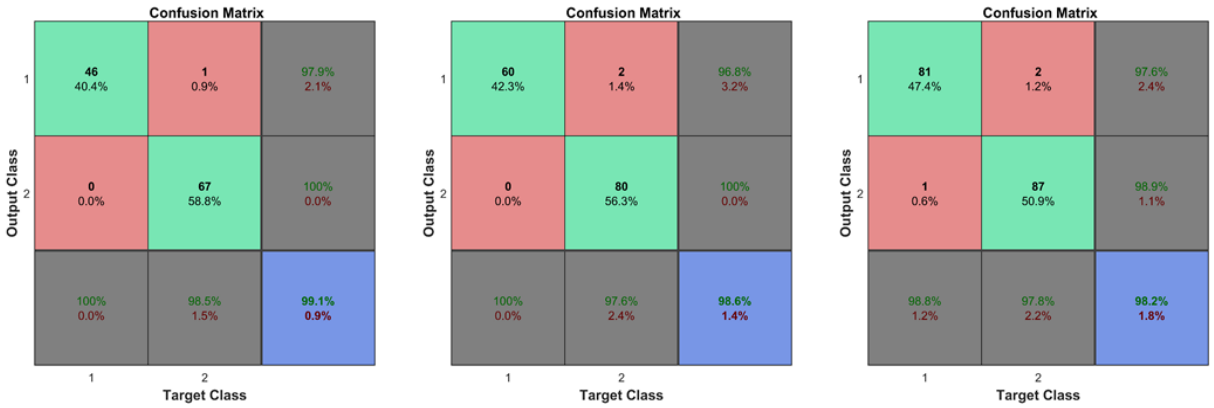

Figure 6. Confusion Matrices for 569 samples (2nd Dataset)
(A) 80 – 20 (%) Train + validate to test partition (B) 75 – 25 (%) Train + validate to test partition (C) 70 - 30(%)Train + validate to test

For the second dataset, we used 30 features in each sample to identify benign tumour patients from malignant tumour patients. In this dataset, we performed several experiments, and the best of the three are presented here. For 80-20% data, there were 114 samples in the test data. Out of these 114 samples, 46 and 68 samples were selected from the benign and malignant classes, respectively. The developed neural network misclassified one sample from the malignant samples and correctly diagnosed all the benign samples. This model produced an overall accuracy of 99.1%, with a precision of 100%, recall 97.87%, and F-measure score was 98.92%.

For the second experiment, we used 75-25% data to visualize the performance of the CNN classifier. In this experiment, 142 samples were used. Out of these 142 samples, there were 60 benign samples and 82 malignant samples. The model misclassified two samples from the malignant class and correctly diagnosed all the benign samples. This model produced an overall accuracy of 98.6%, with a precision value of 100%, a recall value of 96.78%, and the F-measure value was 98.36%. In the third experiment, we split the data into 70-30% to evaluate the performance of the classifier. In this experiment, we have 82 benign samples and 89 malignant samples. The classifier misclassified one and two samples from benign and malignant classes, respectively. For this experiment, the model produced an overall accuracy of 98.2%, with precision 98.78%, recall 97.59%, and F-measure 98.18%.

Table 1 and 2 summarizes the performance of the proposed deep CNN for breast cancer classification using WBCD.

Table 1. Performance Evaluation of proposed Deep CNN and Comparison with previously studies (1st Dataset 683 samples)

Method Proposed	Train + validate to Test partition (%)	Accuracy (%)	Sensitivity (%)	Selectivity (%)
[25] K-Nearest Neighbour Na ïve Bayes Classifier	-	97.5 96.1	-	-
[29] PSOWNN	-	93.67	94.17	92.11
[30] ICA-RBFN	-	90.49	-	-
Our proposed method	73.3 - 26.7	100	100	100

Table 2. Performance Evaluation of proposed Deep CNN and Comparison with previously studies (2nd Dataset 569 samples)

Method Proposed	Train + validate to Test partition (%)	Accuracy (%)	Sensitivity (%)	Selectivity (%)
[1] Stacked Sparse Autoencoder Ensemble Learning	-	98.6	-	-
Our proposed method	80 – 20	99.1	100	97.8

The different parameters which were used to evaluate the performance of the developed neural network is shown in Table1. The cross-entropy illustrates the loss function in the proposed classifier. From the following table, we can see that cross-entropy loss is decreasing as the accuracy of the developed model is increasing, and cross-entropy loss function also determined us in selecting the best-developed model for the breast cancer classification. From Figure 3(A), we conclude that the model has zero log loss function value. In general, the value of cross-entropy lies between 0 and 1. In this study, we got the maximum cross-entropy function values equals to $3.57*10^{-5}$ and $1.34 * 10^{-2}$, and these values were attained by using 73.3 – 26.7 %, and 80 – 20 % data, for 1st and 2nd dataset, respectively.

5. Conclusions

In this work, we have proposed an automatic prognosis system for the detection of breast cancer using deep convolutional neural network (DCNN). The pre-training phase was carried out using the Convolutional layer, max-pooling layer, and fully connected layers, and this pre-training phase was followed by a classification layer to distinguish the benign samples from the malignant samples. For repeatability and reliability, the performance of the proposed DCNN was tested on two WDBC datasets using several train + validate to test partition data. From the experiments on the specified datasets, deep CNN outperforms previously published studies. We have achieved 100% accuracy, with 100% F-measure, and 100% Recall values for the 1st dataset, when we used 73.3% data for the training + validation, and 26.7% data was used to evaluate the performance of the proposed neural network architecture. For the second dataset, we have achieved an overall accuracy of 99.1%, with 98.92% F-measure score, and 97.87% Recall values. The second dataset contains 30 features, by changing the number of filters and filter size, we may get more accuracy for the second dataset. The achieved results for this study indicate the classifier performance over other state of art methodologies. Although, CNN's learning process still needs substantial effort on legacy hardware. Consequently, a CAD system based on CNN using FPGA or other commercial hardware is still a challenging task, and the hardware implementation of such neural networks can assist the medical professionals in the early diagnosis of breast cancer. In the future, we are looking to develop a single chip-based neural networks to diagnose the abnormalities of diabetic retinopathy, EEG, and cardiac arrhythmia.

References

[1] The Global Cancer Observatory, China, International Agency for Research on Cancer, 2018, https://gco.iarc.fr/today/data/factsheets/populations/160-china-fact-sheets.pdf. Accessed: 06 June 2019.

[2] L.A. Altonen, R. Saalovra, P. Kristo, F. Canzian, A. Hemminki, P Peltomaki, R. Chadwik, A. De La Chapelle, "Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease", *N Engl J Med*, vol. 337, pp. 1481-1487, 1998.

[3] B. Liu, K. Yao, M. Huang, J. Zhang, Y. Li and R. Li, "Gastric Pathology Image Recognition Based on Deep Residual Networks," *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, Tokyo, 2018, pp. 408-412. doi: 10.1109/COMPSAC.2018.10267.

[4] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, "Gene selection for cancer classification using support vector machines", *Machine Learning*, vol. 46, pp. 389-422, 2002.

[5] S. Gokhale, "Ultrasound characterization of breast masses", *The Indian journal of radiology &imaging*, vol. 19, pp. 242-249, 2009.
[6] J. Tang, R. M. Rangayyan, J. Xu, I. E. Naqa and Y. Yang, "Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 2, pp. 236-251, March 2009. doi: 10.1109/TITB.2008.2009441.
[7] A. Jemal, R.S., E. Ward, Y. Hao, J. Xu, T. Murray, M.J. Thun, "Cancer statistics", CA Cancer J Clin, Mar-Apr;58(2):71-96. 2008, doi: 10.3322/CA.2007.0010.
[8] Kadam, V.J., Jadhav, S.M. & Vijayakumar, K. J Med Syst (2019) 43: 263. https://doi.org/10.1007/s10916-019-1397-z
[9] Zheng, B., Yoon, S.W., and Lam, S.S., Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Syst. Appl.* 41(4):1476–1482, 2014.
[10] Salama, G.I., Abdelhalim, M.B., and Abd-elghany Zeid, M., Breast cancer diagnosis on three different datasets using multi-classifiers. *Int. J. Comput. Inf. Technol.* 1(Issue 01):2277–0764, 2012.
[11] Jafari-Marandi, R., Davarzani, S., Gharibdousti, M.S., and Smith, B.K., An optimum ANN-based breast cancer diagnosis: Bridging gaps between ANN learning and decision-making goals. *Appl. Soft Comput.* 72:108–120, 2018.
[12] Guo, H., and Nandi, A.K.: Breast cancer diagnosis using genetic programming generated feature. 2005 IEEE Workshop on Machine Learning for Signal Processing, Mystic, CT. , pp. 215–220, 2005.
[13] Lim, C.K., and Chan, C.S., A weighted inference engine based on interval-valued fuzzy relational theory. *Expert Syst. Appl.* 42:3410–3419, 2015.
[14] Zhao, J.Y., and Zhang, Z.L.: Fuzzy rough neural network and its application to feature selection. In: *The Fourth International Workshop on Advanced Computational Intelligence, Wuhan*. pp 684–687, 2011.
[15] Prasad, Y., Biswas, K.K., and Jain, C.K.: SVM classifier based feature selection using GA, ACO and PSO for siRNA design. In: Tan, Y., Shi, Y., and Tan, K.C (Eds.) , Advances in Swarm Intelligence. ICSI 2010. Lecture Notes in Computer Science, Vol. 6146. Springer, Berlin, 2010.
[16] Sheikhpour, R., Sarram, M.A., and Sheikhpour, R., Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer. *Appl. Soft Comput.* 40:113–131, 2016.
[17] Xue, B., Zhang, M., and Browne, W.N.: New fitness functions in binary particle swarm optimisation for feature selection. In: WCCI 2012 IEEE World Congress on Computational Intelligence June, 10–15, 2012 - Brisbane, Australia, 2012.
[18] Xue, B., Zhang, M., and Browne, W.N., Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms. *Appl. Soft Comput.* 18:261–276, 2014.
[19] Fortunato O, BoeriM, Verri C, Conte D, Mensah M, Suatoni P, et al. Assessment of circulating microRNAs in plasma of lung cancer patients, *Molecules,* vol. 19, pp. 3038 – 3054, 2014
[20] Heneghan HM, Miller N, Kerin MJ. MiRNAs as biomarkers and therapeutic targets in cancer, *Curr Opin Pharmacol*, vol. 10, pp. 543-550, 2010.
[21] Zen K, Zhang CY. Circulating micro RNAs: a novel class of biomarkers to diagnose and monitor human cancers, *Med Res Rev*, vol. 32, pp. 326-348, 2012.
[22] Litjens, G. et al. A survey on deep learning in medical image analysis. *Med. Image Anal.* Vol. 42, pp: 60–88, 2017. https://doi.org/10.1016/j.media.2017.07.005
[23] Jasmir *et al*., "Breast Cancer Classification Using Deep Learning," *2018 International Conference on Electrical Engineering and Computer Science (ICECOS)*, Pangkal Pinang, pp: 237-242, 2018. doi: 10.1109/ICECOS.2018.8605180

[24] A. Rakhlin, A. Shvets, V. Iglovikov, A. A. Kalinin, A. Campilho, F. Karray, B. ter Haar Romeny, "Deep convolutional neural networks for breast cancer histology image analysis" in Image Analysis and Recognition, Cham:Springer International Publishing, pp. 737-744, 2018.
[25] J. Xu *et al*., "Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images," in *IEEE Transactions on Medical Imaging*, vol. 35, no. 1, pp. 119-130, Jan.2016. doi: 10.1109/TMI.2015.2458702.
[26] P. Baldi, S.R.B, Bioinformatics: The machine learning approach, Second Edition, The MIT Press, Cambridge, Massachusetts, London, England 2001.
[27] MF. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis", *Expert Syst Appl*, vol. 36(2), pp. 3240-3247, March 2009.
[28] Übeyli, E.D. Implementing automated diagnostic systems for breast cancer detection. *Expert Systems With Applications, vol. 33*, pp. 1054–1062, 2007.
[29] Dheeba, J., Singh, N.A., & Selvi, S.T,. Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *Journal of Biomedical Informatics, vol. 49*, pp. 45–52, 2014.
[30] Mert,A., Kılıç,N.Z.,Bilgili,E.,&Akan, A, Breast cancer detection with reduced feature set. *Computational and Mathematical Methods in Medicine*, pp. 1–11. 2015.
[31] Nahato, K.B.,Nehemiah,H.K.,&Kannan, A, .Knowledge mining from clinical datasets using rough sets and back propagation neural network. *Computational and Mathematical Methods in Medicine*, pp. 1–13, 2015.
[32] Ahmed M. Abdel Zaher, AymanM.E ldeib, Breast cancer classification using deep belief networks, Expert Systems With Applications vol. 46, pp. 139–144, 2016.
[33] Prabhakar S.K., Rajaguru H. Performance Analysis of Breast Cancer Classification with Softmax Discriminant Classifier and Linear Discriminant Analysis. In: Maglaveras N., Chouvarda I., de Carvalho P. (eds) Precision Medicine Powered by pHealth and Connected Health. ICBHI 2017. IFMBE Proceedings, vol 66. Springer, Singapore, 2018. doi: https://doi.org/10.1007/978-981-10-7419-6_33.
[34] J. S. Snchez, R.A.M., J. M. Sotoca, "An analysis of how training data complexity affects the nearest neighbor classifiers", *Pattern Analysis and Applications*, vol. 10, no. 3, pp. 189-201, August 2007.
[35] M. R. Al-Hadidi, A. Alarabeyyat and M. Alhanahnah, "Breast Cancer Detection Using K-Nearest Neighbor Machine Learning Algorithm," *9th International Conference on Developments in eSystems Engineering (DeSE)*, Liverpool, 2016, pp. 35-39. 2016. doi: 10.1109/DeSE.2016.8.
[36] M. Amrane, S. Oukid, I. Gagaoua and T. Ensarİ, "Breast cancer classification using machine learning," *2018 Electric Electronics, Computer Science, Biomedical Engineering' Meeting (EBBT)*, pp. 1-4, 2018. . doi: 10.1109/EBBT.2018.8391453
[37] P. Bhuvaneswariaa, B. Therese, "Detection of Cancer in Lung with K-NN Classification Using Genetic Algorithm", *Procedia Materials Science*, vol. 10, pp. 433-440, 2015.
[38] A. L. Maas , A. Y. Hannun and A. Y. Ng , *Proc. ICML* , 2013,
[39] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In ICML, 2015.
[40] N. Srivastava , G. Hinton , A. Krizhevsky , I. Sutskever and R. Salakhutdinov , Dropout: A Simple Way to Prevent Neural Networks from Overfitting, *J. Mach. Learn. Res.* vol. 15, pp: 1929 —1958, 2014