A Proposed Convolutional Neural Network for Breast Cancer Diagnoses

Noor Kareem KADHIM [1], *Belal AL-KHATEEB* [2], *Huda Wadah AHMED* [1]

[1]Iraqi Commission for Computer and Informatics, Informatics Institute for Postgraduate Studies, UOITC 1st Building, Al-Nidhal Street, 10066 Baghdad, Iraq
[2]Department of Computer Science, College of Computer Science and Information Technology, University of Anbar, CSIT 1st Building, Al-Ceramic Street, 31001 Ramadi, Iraq

dp202120639@iips.icci.edu.iq, belal-alkhateeb@uoanbar.edu.iq, dr.hudawadah@iips.icci.edu.iq

DOI: 10.15598/aeee.v21i1.4658

Article history: Received Aug 04, 2022; Revised Oct 03, 2022; Accepted Nov 23, 2022; Published Mar 31, 2023.

Abstract. *Breast cancer is the second greatest cause of death in women worldwide, however, early detection may result in life prolongation or even complete recovery. Breast cancer can be classified by physicians into two types: benign tumors, and malignant tumors, all of which are fatal if not treated early. Several machine-learning algorithms have been developed to help physicians make diagnostic choices, concretely a convolutional neural network is presented in this paper. The proposed system is divided into several fundamental steps. The proposed classifier is trained to distinguish between incoming tumors using a dataset of 780 images. To evaluate the classifier's performance accuracy, precision, recall, and F1-score are used. In the testing stage, the proposed method achieved an overall classification accuracy of 93 %, 93 % precision, 93 % recall, and 93 % F1-score.*

Keywords

Breast cancer, machine learning, deep learning, convolutional neural network.

1. Introduction

Breast cancer is a primary cause of death worldwide with multiple, difficult-to-prevent causes, particularly in developed nations [1]. Mammography is a popular and common screening method that is widely used thanks to that it is easy-to-use and low-cost. However, reliable mammography diagnosis requires the use of a skilled expert radiologist who can spot abnormalities [2].

Small aberrant formations, such as microcalcifications with diameters smaller than 1 mm, might be difficult to detect in some circumstances. Furthermore, dense breast formations may be difficult to differentiate from lesions of similar structure and contrast [1].

Consequently, there has always been a demand for a smart tool that gives mammography analysis comparable to that of an expert, as early detection of small size lesions is likely to improve patient survival [2].

Recently, two new technological advancements have sparked a revolution in the creation of improved healthcare for the community. The first is a big data, that means a large quantity of data can be readily shared, analyzed, and integrated to cover a wide variety of variables. The cancer imaging archive is one possible example in the area of medical imaging, where thousands of pictures from various modalities are provided with nearly complete access. The second one is deep learning, which is a subset of machine learning that uses hierarchical architectures to learn high-level abstractions from data. Deep learning is a novel method that has been widely applied in traditional artificial intelligence areas, such as computer vision and many others. Deep learning's current boom may be linked to three primary factors: much increased chip processing capacities (e.g. GPU units), significantly decreased computer hardware prices, and a significant breakthrough in machine learning methods [3] and [4]. Deep neural networks with many layers may extract many previously unattainable characteristics. Convolutional neural networks have had a significant influence on image analysis and comprehension, particularly in picture classification, segmentation, and analysis [5]. Several deep learning models have previously been created for the detection and identification of breast cancer using digital mammography data [6].

2. Literature Review

This section discusses the literature related to detection and classification of breast cancer based on deep learning, which is a subset of machine learning. It can be said that deep learning has become a semi-pioneer in many scientific, security, and other fields, including healthcare because it has strong capabilities and requires big amounts of information; these are some studies related to this topic:

Inspired on the U-net structure, authors in [7] suggested a new network architecture for the efficient and early diagnosis of breast cancer. Results demonstrated a high rate of sensitivity and specificity, indicating the potential clinical utility of the suggested method, when using Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) dataset. The proposed method can improve the clinical performance of breast cancer diagnosis, particularly in the early stages. However, one of the major limitations of their research is that the best network design for digital mammography and its sensitivity are still unknown.

The study in [8] proposed a deep learning method for breast cancer diagnosing based on BreAst Cancer Histology images (BACH) dataset biopsy microscope images. Multiple Deep Convolution networks, including Vgg16, Alexnet, Inception, Resnet50, Resnet101, and Densenet169, are employed. Densenet169, Resnet50, and Resnet101 obtained the highest accuracy without any data pre-processing procedures, achieving 62 %, 68 %, and 85 % accuracy, respectively. In addition, the paper illustrates the impact of various data preparation procedures on the performance of the most efficient models. According to the experimental findings, data augmentation and segmentation improve the accuracy of the top models by 20 %, 17 %, and 6 %, respectively. To further improve the accuracy of the models, an ensemble learning technique was used to aggregate the most accurate models. The results indicated that the highest accuracy achieved was 92.5 %. The obtained accuracy can be further enhanced.

With the goal of predicting relapse and metastasis in HER2-positive breast cancer patients, authors in [9] developed a novel multimodal deep learning approach in 2021 that combines whole slide Hematoxylin and Eosin (H&E) images with clinical data. In this study, researchers gathered full H&E staining photographs from breast cancer patients' surgical tissues and downsized them to 512×512 pixels. These pictures are fed into a deep Convolutional Neural Network (CNN) that pulls out visual attributes that are then mixed with clinical data according to the characteristics. After that, a cutting-edge multimodal model is developed to forecast individual patients' outcomes. The model had an Area Under the Curve (AUC) of 0.76, as calculated by two-fold Cross-Validation (CV). Researchers downloaded data from The Cancer Genome Atlas (TCGA) for 123 Human Epidermal growth factor Receptor 2 (HER2)-positive breast cancer patients with available H&E images, known recurrence, and metastatic status in order to further examine the model's performance. The proposed model achieved an AUC of 0.72 for the TCGA data despite substantial variation in race and experimental approaches. In conclusion, HER2-positive breast cancer patients' recurrence and metastatic risk could be estimated using H&E images, clinical data, and cutting-edge deep learning algorithms. Preliminary evidence from the TCGA datasets suggests that deep learning can predict breast cancer recurrence and metastasis using H&E-stained histological images and clinical data. This study also points the way toward a new routine clinical use of deep learning. Investigation on this topic is still in its early stages. The model will be widely used as a diagnostic and therapeutic tool after its clinical use will be established with greater rigor.

Using Ultrasound Images (USIs), authors in 2022 [10] described an Ensemble Deep-Learning-Enabled Clinical Decision Support System for Breast Cancer Diagnosis and Categorization (EDLCDS-BCDC) approach. Using USIs, the suggested EDLCDS-BCDC approach aimed to detect the presence of breast cancer. In this method, USIs are initially preprocessed in two steps, namely wiener filtering and contrast enhancement. In addition, the Chaotic Krill Herd Algorithm (CKHA) is utilized in conjunction with Kapur's Entropy (KE) in the image segmentation procedure. In addition, for feature extraction, an ensemble of three deep learning models, VGG-16, VGG-19, and Squeeze Net, is employed. Cat Swarm Optimization (CSO) with the Multilayer Perceptron (MLP) model is used to identify the photos according to the presence or absence of breast cancer. Extensive simulations are performed on benchmark databases, and the results demonstrate that the proposed

EDLCDS-BCDC strategy outperforms current approaches. The ImageNet database is utilized in their project. The obtained accuracy can be further enhanced.

The study presented in [11] provided a multi-fold deep learning method for detecting breast cancer using microscope images of biopsies. Diverse forms of convolutional nets with a high degree of depth are utilized. The impact of different data preparation procedures on the performance of deep learning models was then explored. The research culminated with the introduction of an ensemble method for improving performance by combining the best models. The study introduced a deep learning approach to biopsy diagnosis to save time for both patients and pathologists while also lowering the risk of misdiagnosis, which could lead to other complications and decline in health. For the analysis, BACH dataset was used, but the obtained accuracy could be further enhanced.

Authors in 2021 [12] proposed a deep learning-based method for classifying breast masses in ultrasound images. In medical image analysis, transfer learning using CNNs is frequently utilized to construct object-recognition models. The most prevalent strategies for fine-tuning try to change the weights of pre-trained networks to address specific medical concerns. However, fine-tuning might be challenging when the number of trainable network parameters is enormous and there are few relevant medical data. To address this problem, the author proposes a new transfer learning approach based on Deep Representation Scaling (DRS) layers, which are inserted between the CNN's pre-trained blocks to boost the network's information flow. During network training, the parameters of the DRS layers are simply modified to adapt the pre-trained CNN to breast mass ultrasound picture processing. The researcher demonstrated that the DRS-based strategy drastically decreases the number of trainable parameters and delivers performance that is either superior or equivalent to that of conventional transfer learning techniques. With an accuracy of 91.5 %, in conjunction with conventional fine-tuning procedures, the suggested DRS layer method displayed effective breast mass classification performance. Other alternative forms for the DRS layers must be researched, as well as the applicability of the approach in various network designs.

It is worth to mention that many other researchers tried to diagnose breast cancer by employing techniques other than deep learning, for further details please refer to [13], [14], [15], [16], [17], [18] and [19].

3. Research Method

The methods employed in the proposed system for detecting breast cancer and assessing if a case is normal, benign, or malignant are described in this section. The system is divided into three steps. The data is pre-processed in the first step to increase its size. CNN is used to extract features from the data in the second step. The third step involves detecting and categorizing the patient's condition using CNN. Figure 1 shows the general structure of the breast cancer categorization system.

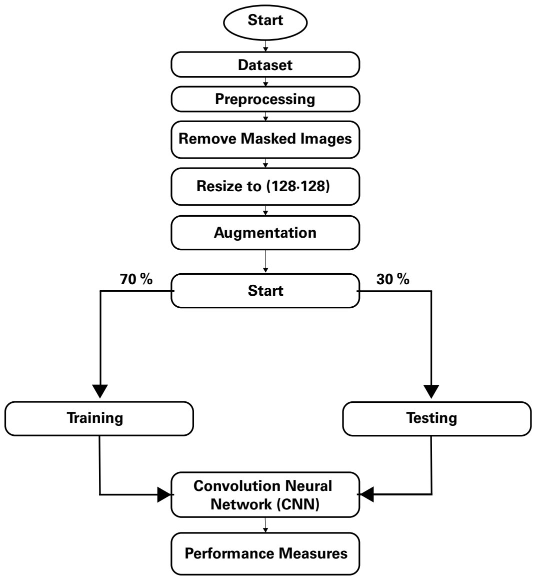


Fig. 1: The general structure of the breast cancer categorization system.

3.1. Dataset

The data examines medical images of breast cancer that are obtained from an ultrasound scan and can be downloaded from the following website [20].The data is collected at the outset from women aged from 25 to 75 in 2018. There is a total of 600 female patients. The collection contains 780 PNG images with an average resolution of 500×500 pixels. The images are separated into three categories: normal (133 images), benign (437 images), and malignant (210 images).

3.2. Image Pre-processing

Image pre-processing is the first stage in displaying various useful image features for subsequent usage. The pre-processing stage consists of minor enhancements to the breast cancer images. The steps utilized in this work are described in the following subsections.

1) Remove Masked Images

The first stage of processing is to remove the mask images from the whole dataset.

2) Image Resize

One of the main limitations of CNN is the requirement for image resizing in the dataset to a consistent dimension. In this phase, images are transformed into an array of pixels and then scaled before being sent into CNN. The goal of shrinking photos is to minimize computational load, speed up the training technique, and generate an accurate test model. By applying the resizing function for the image's width and height, the proposed system reduces the image size from 500×500 to 128×128. This process aided in the speeding up of the training process while also preserving the data's integrity without affecting the original image data.

3.3. Augmentation

Augmentation, also known as data transformations, is a common pre-processing technique that involves expanding the dataset and distorting the images in various ways so that a broader range of data may be displayed to the network. This method successfully increases the amount of the dataset, when lowering the danger of overfitting. Scaling, rotation, and other related adjustments are prevalent in image augmentation. Augmentation is used to extend datasets and supply neural networks with a range of visual variants, where this technique increases the possibility that the model will detect objects of any shape. In this paper, data augmentation is employed for the breast cancer dataset. It is crucial to note that after the data augmentation, the number of images in the datasets is increased to 15,474. The augmentation techniques used in this work are as follows:

- Rotation range: It is a form of augmentation that aids the network in identifying the item in any direction in the picture. All images in this work have been rotated using a random rotation angle ($0°$–$45°$).
- Width shift range: Its role is to shift the image to the right or left (i.e., in horizontal direction), implying that the displacement is done on the horizontal axis. For all images in this work, the width shift range = 0.2 has been implemented.
- Height shift range: Its function is to shift the image up or down (vertical shifts), suggesting that the displacement occurs on the vertical axis. The height shift range = 0.2 has been applied to all the images in this work.
- Shear range: Image shear refers to the distortion of an image along an axis, usually to produce or correct perception angles, in which one axis is fixed while a certain angle (shear angle) is stretched. The shear range is 0.2, and it is utilized for all images in the dataset.
- Zoom range: Zooming allows us to zoom in or out. It is vital to understand that if the zoom range is larger than one, the image is enlarged, and if the zoom range is less than one, the image is reduced. This strategy is useful because it enlarges the disease's afflicted areas, allowing the model to learn just the properties relevant to the damaged areas. All the images in this work have been zoomed using the range = 0.2.
- Horizontal flip: This method is employed to render the trained model to be insensitive to reflect objects as it turns the images horizontally.
- Fill mode: This is a mechanism for compensating for missing pixels following the modifications described above. The primary goal of pixel compensation is to preserve image quality. There are several methods for offsetting these pixels, including Reflect, Constant, and Nearest filling. It assigns the values using the fill mode option. For all the images in this work, the fill mode = nearest has been used. To conserve image quality, such a decision is based on selecting the nearest pixel value for each empty value.

3.4. Dataset Splitting

It is vital to analyze a classifier's performance to anticipate its future estimate accuracy. The dataset is split after the pre-processing stage into two sets: one for training and the other for testing. The training set (also known as the estimation set) is used to estimate the model parameters. The collection of tests (validation) is used to evaluate the model's performance. The proposed work used the holdout approach as a common manner of dividing the data using a 70/30 ratio, which means that 70 % of the data is used to train the model, with the remaining 30 % being utilized to evaluate the model. The splitting ratio is the best obtained one after testing many ratios.

3.5. CNN Architecture

The employed CNN architecture is made up of multiple layers, including an input layer, three hidden layers, and an output layer. Convolution layers, max-pooling layers, and flatten layers are common hidden layers in CNNs. CNN is a popular deep learning network because it has numerous advantages over other networks, such as using fewer parameters with fewer neurons, which results in shorter training time. A customized CNN architecture is utilized in the proposed work. Figure 2 shows the overall architecture utilized for the CNN design, with certain parameters modified for improved results. The following subsections give further information about the model's layer.

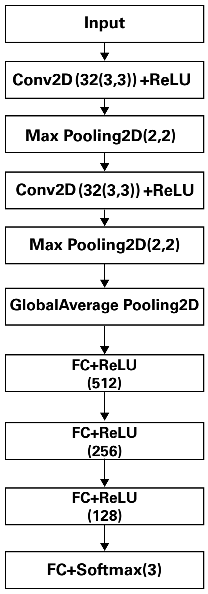


Fig. 2: Design of the proposed CNN structure.

1) Input Layer

The input layer contains the input image as well as its pixel values. The input layer in the proposed work comprises images of normal, benign, and malignant breast cancer inserted as a matrix of pixels, and its size is previously decreased to 128×128×3 during the pre-processing stage.

2) Convolution Layer

This layer is the most significant aspect of the CNN architecture since it is in charge of extracting features and creating feature maps. This layer has weight arrays in the convolution layer filters, and each element in the array represents the weight being taught. To put it another way, the behavior of this layer is controlled by an array of numbers that is referred to as the kernel or filter matrix. To extract features, these filter arrays are applied to the entire image. This layer's outputs are feature maps, which means that each feature map represents the outcome of the convolution process. The convolution layer is where computations and difficult tasks are conducted. To establish any convolution layer, the kernel must be configured, which consists of four hyperparameters: the number of filters, the kernel initializer, the padding amount, and the input shape. Except for filter settings, the three parameters of all convolution layers in the proposed system are as follows: The kernel size is set to three because a smaller kernel size makes the network deeper.

The proposed method incorporates two stages of convolution. The first layer has 32 filters of 3×3 sizes, appropriate padding, and input form 128×128. The second convolution layer has 32 filters of 3×3 size as well as appropriate padding.

Following each convolution layer, the non-linear function, also known as the activation function, is applied, which operates the non-linear transformation of the convolution layer output. Rectified Linear Unit (ReLU) has been used in the suggested system, which is the best and the most renowned kind chosen after many tests in the CNN area. The negatives in the previous stage's ReLU values are converted to 0, while the positives remain unaltered. ReLU has been found to expedite CNN training, and it is simple to deploy.

3) Pooling Layer

The pooling layer is put after the non-linear layer. This layer has no parameters that need to be modified during the backpropagation phase. However, the absence of parameters does not mean that pooling performs no function in backpropagation. The pooling layer is in charge of sending values to the next and previous levels during forward and backward propagation. Because of its efficiency and capacity to produce the best results, maximum pooling is used in the proposed work after multiple tests. There are options for the window size and stride when using maximum pooling. In the proposed system, the window size is set to two and the stride is also set to two, which are common settings of maximum pooling. These configurations boost the network's convergence rate.

4) Fully Connected Layer

Feature vectors are sent to the Fully Connected (FC) layer, which is used to analyze the best features that the model learns. These characteristics are then used to categorize the inputs into distinct groups. Four completely linked layers are employed in the suggested model. The first one is completely linked layer employing 512 units (weights) that uses Rectified Linear Unit (ReLU) as an activation function. The second completely linked layer employs 256 units (weights) using ReLU as an activation function. The third fully connected layer employs 128 units (weights) using ReLU as an activation function. To extract the outcome from a model, the last fully connected layer employs three units (nodes) reflecting the number of categories, which are appended with the SoftMax activation function. SoftMax's primary role is to calculate probabilities, based on which the cancer is categorized into groups depending on the data utilized. Each input in the FC layer is linked to each output unit with the corresponding probability. This implies that each neuron in this layer may gather information from inputs, which aids in predicting the value of the proper class in the SoftMax layer. The FC layer output is a value for each category, and a SoftMax is created for it. Consequently, the category will be determined in each phase depending on the attributes derived from images from previous layers.

It is advised to use the SoftMax function to terminate the CNN in multi-classification problems. The values of the findings will be between zero and one, which is ideal because it avoids binary categorization and matches with many of the data in the model. SoftMax's role is to train the network by transforming labels into probabilities and then using the loss function to estimate the model's loss. The higher the probability value, the more likely the category is correct.

3.6. CNN Training

After determining the configuration of each layer, the network is trained using the training set. The goal of training the CNN is to fine-tune its internal parameters to improve its performance on input data. Training a network is the process of obtaining kernels from convolution layers and weights from FC layers to reduce discrepancies between truth labels on a training dataset and output predictions. This is one of CNN's most important phases, as the network is designed to learn by gathering user-data-derived properties. Each image is categorized because of network learning. Numerous choices must be recognized and altered to carry out the training process. These options include initialize weights, update parameters, Adam calculation, and backpropagation. This helps estimate the loss function needed to evaluate the training process. In addition, the kind of loss function and accuracy metric are utilized to evaluate the model. The following subsections explains the CNN training process.

1) Weights Initialization and Parameter Updating

When it comes to the training phase, it is critical for a CNN to alter the internal parameters to enhance classification. Two main processes of training are needed to carry out learning: The initial stage is to give training data to the network so that it can finish a whole epoch at a time. The loss is then determined and based on the value of the loss, the network modifies the parameters in the backpropagation process in each epoch. All trials in the proposed system employed 200 epochs since they were sufficient for learning. Later, 400 epochs are utilized to enhance the system using the tweaking algorithm and the results are excellent as will be shown in the next section.

Because parameters and hyperparameters have such a large influence on network performance, they must be fine-tuned after numerous tests. The number of units, the kind of optimization function, and other class parameters are examples. Additional parameters, such as learning rate and batch size are employed to train the suggested basic model.

A learning rate is the number of steps taken to update the weights during training. The learning rate hyperparameter controls how fast or slowly the model learns. When the learning rate is too low, training is sluggish and error reduction is very slow, requiring a lengthy period for convergence. A learning rate that is too high, on the other hand, may result in too large weight changes, and the model's performance (such as its loss on the training dataset) would swing throughout training epochs, resulting in no convergence at higher learning rates. The learning rate to be used in the suggested system is 0.001 because it was effective in terms of learning speed and reducing error. This value is found to be the best value out of many tested values during the work.

Batch size is the technique of separating data into batches before it is transmitted over a network. It refers to the amount of training instances in a single forward/backward pass in another sense. The more memory space used, the larger the batch size. The used batch size in the proposed work is 192 since the network trains quicker with smaller batches and consumes less memory. Furthermore, if all samples are used at the same time during propagation, the network parameter will only be updated once.

2) Adaptive Moment Estimation

In this study, an optimization strategy based on the Adaptive Moment Estimation (Adam) algorithm is used. It is a technique for updating CNN parameters (weights) based on the loss function value recorded in each epoch since the major goal of training is to lower the loss using backpropagation and the Adam function to update the weights. However, finding a solution to make the loss equal to zero is challenging because the error function is not convex; it contains multiple local optima in addition to the global optimum. Therefore, the iteration approach is proposed to identify the best solution, in which the process of updating the weights is repeated in each iteration until the best limit is reached. The Adam method is used to update CNN weights as well as compute rates of learning of gradient parameters belonging to the first (i.e. mean) and second moments (i.e. variance). Furthermore, Adam estimates the decay average of past gradients and squares it. Adam's optimizer is described in the following pseudo code:

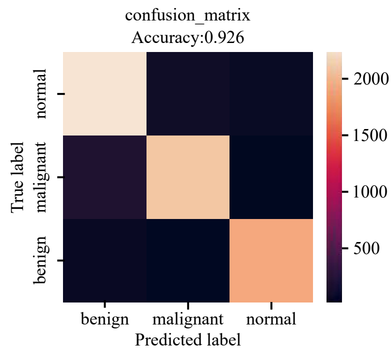
1. Set $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\eta = 10^{-8}$.
2. $m_0 = 0$ // Initial vector of first moment.
3. $v_0 = 0$ // Initial vector of second moment.
4. $t = 0$ // Initial step of time.
5. Repeat 6 to 12 Until $\varnothing_t$ is converged.
6. $t = t + 1$.
7. $g_t = \nabla_{\varnothing} f_t(\varnothing_t - 1)$ // Gradients computation with respect to stochastic objective at time step τ.
8. $m_t = \beta_1 \cdot m_t - 1 + (1 - \beta_1) \cdot g_t$ // Update 1^{st} moment estimate.
9. $v_t = \beta_2 \cdot v_t - 1 + (1 - \beta_1) \cdot g_t^2$ // Update 2^{nd} moment estimate.
10. $\hat{m}_t = \frac{m_t}{1-B_2^t}$ // Create unbiased estimate $\hat{m}_t$.
11. $m_t = \frac{v_t}{1-B_1}^{t}$ // Create unbiased estimate $\hat{v}_t$.
12. $\varnothing_t = (\varnothing_{t-1} - \alpha) \cdot \frac{\hat{m}_t}{\sqrt{v_t + \epsilon}}$ // Parameters updating.
13. Return $\varnothing_t$ // Parameters results (weights).

3.7. Proposed Model Testing

After the construction and training of the suggested model, the model testing step begins, which is an assessment procedure for the model. The evaluation is carried out using the previously established dataset, which represents 30 % of the whole dataset. Furthermore, the review procedure is based on several indicators. The accuracy measures are used to evaluate the proposed system, in addition to other metrics like precision, recall, and F-score. Accuracy is determined via a confusion matrix, which indicates how many categories are properly classified and wrongly classified. The loss during evaluation is calculated in addition to the accuracy using the loss scale. After the stage of careful training and evaluation of the proposed model, the model can classify and predict correctly.

4. Results and Discussion

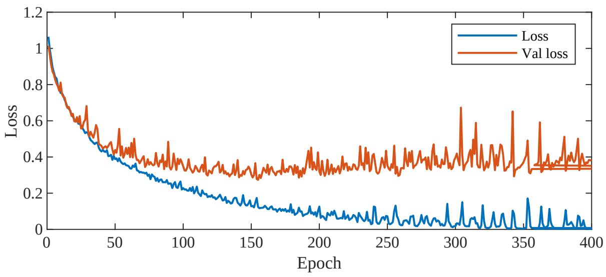
The proposed CNN model for breast cancer disease diagnosis is assessed after each epoch. The attained performance accuracy of the proposed CNN model is about 93 %. Also, the proposed system achieved Precision of 93 %, Recall of 93 %, and F1-Score of 93 %. Figure 3 and Fig. 4 depict the proposed CNN model's training accuracy, validation accuracy, and loss retraction when 400 epochs are used. The variations in the training loss function demonstrate typical Adam algorithm behavior.



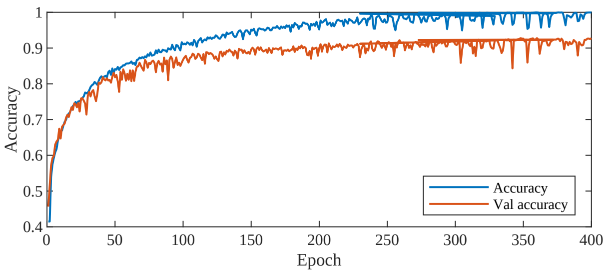
	precision	recall	f1-score	support
benign	0.90	0.92	0.91	2407
malignant	0.94	0.91	0.92	2247
normal	0.95	0.95	0.95	1980
accuracy			0.93	6634
macro avg	0.93	0.93	0.93	6634
weighted avg	0.93	0.93	0.93	6634

Fig. 3: Classified accuracy for precision, recall, and F1-Score support for each class.

The confusion matrix is used as the basis for the calculations that determine the overall accuracy, precision, recall, and F1-score. The proposed CNN model for breast cancer disease diagnosis by using the validation dataset (test data) is evaluated after each epoch. The achieved performance accuracy of the proposed CNN model was nearly 100 % when using the training set, while the performance accuracy, precision, recall, and F1-score were about 93 % when using the test set. The accuracy is growing gradually as the number of epochs increases. This clearly proves that the proposed system is not a fluke of optimization. The exponential trend illustrates the suggested CNN model's quick learning and demonstrated the efficient behavior of the applied classifier. Also, the loss function's behavior demonstrates inverse tracking;



(a)



(b)

Fig. 4: Loss function for training and validation.

the loss begins with a high value when little learning has happened and subsequently decreases to the lowest value at the most utilized epoch. The validation test's average behavior was typically identifying the training ones, which ensures that the classifier achieves a very good performance.

The proposed system is executed for 10 times in a row with identical settings, and the results were so close, with the accuracy growing gradually as the number of epochs increases. This clearly proved that the proposed system is not a fluke of optimization. The exponential trend illustrates the suggested CNN model's quick learning and demonstrated the efficient behavior of the applied classifier. Also, the loss function's behavior demonstrates inverse tracking; the loss begins with a high value when little learning has happened and subsequently decreases to the lowest value at the most utilized epoch. The validation test's average behavior was typically identifying the training ones, which ensures that the classifier achieves a very good performance.

5. Conclusion

Detecting subtle changes in the breast requires expertise in this field. However, the human eye may not always detect such subtle alterations. Medical aid enabled by computer vision and deep learning could save a lot of lives. With this impulse, this paper researched this disease by employing CNN. Breast cancer is one of the most dangerous diseases that must be detected quickly and correctly. It is, therefore, necessary to suggest an effective system for detecting cases of breast cancer. A set of steps is used in this project to diagnose breast cancer. First, a dataset is collected, then pre-processing removes masked data, then images are resized, then augmentation is used to increase data, and finally, CNN is used to train networks. The proposed system is evaluated on several samples obtained from the Kaggle website, and the results showed that it succeeded in the early detection of breast cancer with a rate of 93 %. The following conclusions are drawn from implementing the proposed model:

1. Deep learning is a modern research method that can effectively solve diagnosis problems to help physicians.
2. The suggested CNN model could accurately classify three classes of benign, malignant, and normal breast cancer.
3. CNNs are extremely successful in feature extraction.
4. Data augmentation increased the quantity of training data available throughout the training phase, which helped the proposed model to overcome the overfitting problem.
5. To improve the performance of the CNN model, parameters such as learning rate or number of epochs are adjusted. The increasing number of epochs improves performance. However, numerous trials must be carried out before making a judgment on the epochs and learning rate.
6. It is determined that the model with more data performed better.
7. The lower the number of parameters that may be trained, the better the performance.

The suggested model for identifying and diagnosing breast cancer disorders is a flexible system, therefore there are some suggestions for further development, including the following:

- using transfer learning as a try to increase the diagnosing accuracy,
- using other types of deep learning algorithms such as Resnet50 and U-Net as a try to increase the diagnosing accuracy.

Author Contributions

N.K.K. and H.W.A. performed the analytic calculations and performed the numerical simulations.

B.A.-K. developed the theoretical formalism. All authors contributed to the final version of the manuscript. B.A.-K. supervised the project.

References

[1] JEMAL, A., R. SIEGEL, J. XU and E. WARD. Cancer Statistics. *CA: A Cancer Journal for Clinicians*. 2010, vol. 60, iss. 5, pp. 277–300. ISSN 1542-4863. DOI: 10.3322/caac.20073.

[2] LECUN, Y., Y. BENGIO and G. HINTON. Deep Learning. *Nature*. 2015, vol. 521, iss. 7553, pp. 436–444. ISSN 1476-4687. DOI: 10.1038/nature14539.

[3] DENG, L. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing*. 2014, vol. 3, iss. 1, pp. 1–29. ISSN 2048-7703. DOI: 10.1017/atsip.2013.9.

[4] KELLEHER, J. D. *Deep Learning*. 1st ed. Cambridge: MIT press, 2019. ISBN 978-0-262-53755-1.

[5] LITJENS, G., T. KOOI, B. E. BEJNORDI, A. A. A. SETIO, F. CIOMPI, M. GHAFOORIAN, J. A. W. M. VAN DER LAAK, B. VAN GINNEKEN and C. I. SANCHEZ. A survey on deep learning in medical image analysis. *Medical Image Analysis*. 2017, vol. 42, iss. 1, pp. 60–88. ISSN 1361-8415. DOI: 10.1016/j.media.2017.07.005.

[6] BEKKER, A. J., H. GREENSPAN and J. GOLDBERGER. A multi-view deep learning architecture for classification of breast microcalcifications. In: *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*. Prague: IEEE, 2016, pp. 726–730. ISBN 978-1-4799-2349-6. DOI: 10.1109/ISBI.2016.7493369.

[7] RASHED, E. and M. S. A. EL SEOUD. Deep learning approach for breast cancer diagnosis. In: *Proceedings of the 2019 8th International Conference on Software and Information Engineering*. Cairo: ACM, 2019, pp. 243–247. ISBN 978-1-4503-6105-7. DOI: 10.1145/3328833.3328867.

[8] ELDIN, S. N., J. K. HAMDY, G. T. ADNAN, M. HOSSAM, N. ELMASRY and A. MOHAMMED. Deep Learning Approach for Breast Cancer Diagnosis from Microscopy Biopsy Images. In: *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*. Cairo: IEEE, 2021, pp. 216–222. ISBN 978-1-66541-243-8. DOI: 10.1109/MIUCC52538.2021.9447653.

[9] YANG, J., J. JU, L. GUO, B. JI, S. SHI, Z. YANG, S. GAO, X. YUAN, G. TIAN, Y. LIANG and P. YUAN. Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. *Computational and Structural Biotechnology Journal*. 2022, vol. 20, iss. 1, pp. 333–342. ISSN 2001-0370. DOI: 10.1016/j.csbj.2021.12.028.

[10] RAGAB, M., A. ALBUKHARI, J. ALYAMI and R. F. MANSOUR. Ensemble Deep-Learning-Enabled Clinical Decision Support System for Breast Cancer Diagnosis and Classification on Ultrasound Images. *Biology*. 2022, vol. 11, iss. 3, pp. 1–17. ISSN 2079-7737. DOI: 10.3390/biology11030439.

[11] MOHAMED, A., E. AMER, SARA NOOR ELDIN, JANA KHALED, M. HOSSAM, N. ELMASRY and G. T. ADNAN. The Impact of Data processing and Ensemble on Breast Cancer Detection Using Deep Learning. *Journal of Computing and Communication*. 2022, vol. 1, iss. 1, pp. 27–37. ISSN 2636-3577. DOI: 10.21608/jocc.2022.218453.

[12] BYRA, M. Breast mass classification with transfer learning based on scaling of deep representations. *Biomedical Signal Processing and Control*. 2021, vol. 69, iss. 1, pp. 1–8. ISSN 1746-8094. DOI: 10.1016/j.bspc.2021.102828.

[13] LIU, N., E.-S. QI, M. XU, B. GAO and G.-Q. LIU. A novel intelligent classification model for breast cancer diagnosis. *Information Processing & Management*. 2019, vol. 56, iss. 3, pp. 609–623. ISSN 0306-4573. DOI: 10.1016/j.ipm.2018.10.014.

[14] MOHAMMED, M. A., B. AL-KHATEEB, A. N. RASHID, D. A. IBRAHIM, M. K. ABD GHANI and S. A. MOSTAFA. Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Computers & Electrical Engineering*. 2018, vol. 70, iss. 1, pp. 871–882. ISSN 0045-7906. DOI: 10.1016/j.compeleceng.2018.01.033.

[15] SADOUGHI, F., Z. KAZEMY, F. HAMEDAN, L. OWJI, M. RAHMANIKATIGARI and T. T. AZADBONI. Artificial intelligence methods for the diagnosis of breast cancer by image processing: a review. *Breast Cancer: Targets and Therapy*. 2018, vol. 10, iss. 1, pp. 219–230. ISSN 1179-1314. DOI: 10.2147/BCTT.S175311.

[16] GAYATHRI, B. M., C. P. SUMATHI and T. SANTHANAM. Breast Cancer Diagnosis Using Machine Learning Algorithms - A Survey. *International Journal of Distributed and Parallel systems*.

2013, vol. 4, iss. 3, pp. 105–112. ISSN 2229-3957. DOI: 10.5121/ijdps.2013.4309.

[17] KHURIWAL, N. and N. MISHRA. Breast cancer diagnosis using adaptive voting ensemble machine learning algorithm. In: *2018 IEEMA Engineer Infinite Conference (eTechNxT)*. New Delhi: IEEE, 2018, pp. 1–5. ISBN 978-1-5386-1138-8. DOI: 10.1109/ETECHNXT.2018.8385355.

[18] KHAMPARIA, A., S. BHARATI, P. PODDER, D. GUPTA, A. KHANNA, T. K. PHUNG and D. N. H. THANH. Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. *Multidimensional Systems and Signal Processing*. 2021, vol. 32, iss. 2, pp. 747–765. ISSN 1573-0824. DOI: 10.1007/s11045-020-00756-7.

[19] RABIEI, R. Prediction of Breast Cancer using Machine Learning Approaches. *Journal of Biomedical Physics and Engineering*. 2022, vol. 12, iss. 3, pp. 297–308. ISSN 2251-7200. DOI: 10.31661/jbpe.v0i0.2109-1403.

[20] AL-DHABYANI, W., M. GOMAA, H. KHALED and A. FAHMY. Breast Ultrasound Images Dataset. In: *Kaggle* [online]. 2020. Available at: `https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset`.

About Authors

Noor Kareem KADHIM received her B.Sc. degree Computer Science from the Computer Science Department, College of Computer Science & Mathematics, the University of Thi - Qar, Iraq, in 2016, She is currently a high diploma student, Iraqi Commission for Computer and Informatics, Informatics Institute for Postgraduate Studies, Baghdad, Iraq. Her current research interests include applying deep learning techniques in the health care sector.

Belal AL-KHATEEB received the B.Sc. (honors) (first class) degree in computer science from Al-Nahrain University, Baghdad, Iraq, in 2000, and the M.Sc. degree in computer science from Al-Nahrain University, Baghdad, Iraq, in 2003, and the Ph.D. degree from the School of Computer Science, University of Nottingham, Nottingham, United Kingdom, in 2011. He is currently a professor at the College of Computer Science and Information Technology, University of Anbar. He has published over 77 refereed journal and conference papers. His current research interests include evolutionary and adaptive learning particularly in computer games, expert systems, and heuristics and meta/hyper-heuristics. He has a particular interest in computer games programming. Prof. Al-Khateeb is a reviewer of twenty-three international journals (including three IEEE Transactions) and thirty conferences.

Huda Wadah AHMED received the B.Sc. degree in computer science from Al-Rafidien University, Baghdad, Iraq, in 2002, and the M.Sc. degree in computer science from Iraqi Commission for Computers & Informatics/ Informatics Institute for Postgraduate Studies, Baghdad, Iraq, in 2012, and the Ph.D. degree from the Iraqi Commission for Computers & Informatics/ Informatics Institute for Postgraduate Studies, in 2021. She is currently a lecturer at the Informatics Institute for Postgraduate Studies. She has published over 6 refereed journal and conference papers. Her current research interests include applying deep learning techniques in the health care sector.