



Article

Breast Cancer Classification Using Concatenated Triple Convolutional Neural Networks Model

Mohammad H. Alshayegi * and Jassim Al-Buloushi

Computer Engineering Department, College of Engineering and Petroleum, Kuwait University, Safat, P.O. Box 5969, Kuwait City 13060, Kuwait

* Correspondence: m.alshayegi@ku.edu.kw

Abstract: Improved disease prediction accuracy and reliability are the main concerns in the development of models for the medical field. This study examined methods for increasing classification accuracy and proposed a precise and reliable framework for categorizing breast cancers using mammography scans. Concatenated Convolutional Neural Networks (CNN) were developed based on three models: Two by transfer learning and one entirely from scratch. Misclassification of lesions from mammography images can also be reduced using this approach. Bayesian optimization performs hyperparameter tuning of the layers, and data augmentation will refine the model by using more training samples. Analysis of the model's accuracy revealed that it can accurately predict disease with 97.26% accuracy in binary cases and 99.13% accuracy in multi-classification cases. These findings are in contrast with recent studies on the same issue using the same dataset and demonstrated a 16% increase in multi-classification accuracy. In addition, an accuracy improvement of 6.4% was achieved after hyperparameter modification and augmentation. Thus, the model tested in this study was deemed superior to those presented in the extant literature. Hence, the concatenation of three different CNNs from scratch and transfer learning allows the extraction of distinct and significant features without leaving them out, enabling the model to make exact diagnoses.

Keywords: artificial intelligence; machine learning; breast tumors; convolutional neural networks; Bayesian optimization



Citation: Alshayegi, M.H.; Al-Buloushi, J. Breast Cancer Classification Using Concatenated Triple Convolutional Neural Networks Model. *Big Data Cogn. Comput.* **2023**, *7*, 142. <https://doi.org/10.3390/bdcc7030142>

Academic Editor: Salvador García López

Received: 12 July 2023

Revised: 6 August 2023

Accepted: 14 August 2023

Published: 16 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tumors are unnatural and aberrant cell growth in the body. They are often caused by genetic issues. Cancerous or malignant tumors are the most harmful form of tumor cells because they can develop and spread to neighboring tissue and organs. Benign tumors fall into the noncancerous category and do not invade and spread to other cells [1]. There were 45.62 million cancer cases worldwide and 5.75 million cancer-related deaths in 1990. In 2017, there were 9.56 million cancer-related fatalities and 100.48 million cancer cases. Worryingly, the rates of cancer incidence and death have been rising yearly. One leading cause of cancer is breast tumors. Of the 18 million new cases of cancer documented in 2018, 11.6% were characterized as breast cancer. Of the 9.5 million deaths due to cancer documented in the same year, 6.6% were due to breast cancer. Projections published in 2019 anticipated that there would be 10 million cancer deaths, 6.5% of which would be attributable to breast cancer, and 19 million new cancer cases, 11.5% of which would be breast cancers [2]. In 2020, breast cancer claimed the lives of 685,000 persons worldwide and it became the most common cancer among women since 7.8 million women had been diagnosed with it [3]. By 2040, there will likely be 3.2 million more new cases of breast cancer worldwide than the number documented in 2020 [4].

Breast cancer develops when breast tissue grows to form a tumor because of breast cell mutations that cause the cells to divide. Once formed, they have the potential to develop and spread into adjacent cells and tissues. It is unclear what variables prompt the onset

of the development of harmful cells. Age, gender, genetics, and family history all have links to breast cancer risk. Additionally, it is more likely to develop in individuals who use hormone replacement therapy (HRT), smoke, drink alcohol, are obese, and/or have been exposed to radiation [5].

Treatment for breast cancer varies according to a range of variables, including the tumor's size, location, and whether it has spread to other organs. Surgery, chemotherapeutic agents, radiation, hormone therapy, immunotherapy, targeted medication therapy, etc., are all possible forms of treatment. The study in [6] found that the average survival percentage for women who were given a Stage I breast cancer diagnosis was 96.8%. Periodic breast examinations by medical professionals can help to facilitate the early detection of a tumor. Using a variety of techniques, including mammograms, positron emission tomography (PET) scanning, magnetic resonance imaging (MRI), etc., healthcare professionals check the breasts to look for breast malignancies [7].

In this study, breast tumor classifications were performed on mammograms labeled as negative or positive in binary classifications. In multi-classifications, positive cases are divided into benign masses, malignant masses, benign calcifications, and malignant calcifications. Each type of breast tumor has different characteristics that can be inspected using mammograms to classify them correctly [8]. The masses were large white lumps appearing in the breast tissues. They can be found in different shapes, which helps identify whether they are benign or malignant. Benign masses are smoother, more regular, round, or almost round, whereas malignant masses are irregular and often have spiky edges [9,10]. Moreover, calcifications are small white dots or round bodies in groups appearing in breast tissues. Differences in the size and distribution of these calcifications can help identify whether they are benign or malignant. Benign calcifications are larger than malignant calcifications and are scattered over a region. In contrast, malignant calcifications are small and distributed linearly in segments [11].

The highest disease prediction accuracy with reliability is the key factor when applying deep learning (DL) techniques to medical fields. Although Convolutional Neural Networks (CNN) models that are already developed could offer better prediction performance, model reliability needs to be ensured. Here, we address this issue by concatenating three networks where each would be able to extract distinct and significant features without leaving any out.

This research presents a novel machine learning (ML) model for breast cancer classification based on mammography scans using concatenated triple CNN models to ensure that every significant feature responsible for classification decisions is considered. To achieve this, a CNN model is developed from scratch in addition to the two models by transfer learning (InceptionResNetV2 and Xception) techniques, and the performances are compared individually in addition to the overall model performance. To analyze the techniques that can improve prediction accuracy in classification applications, data augmentation, and Bayesian optimization experiments will be conducted. The focus is to develop a multi-classification framework that can classify the cancer state in addition to binary classification as positive or negative cancer cases, and the proposed model is capable of offering superior performance compared with alternative state-of-the-art models developed using the same dataset.

The rest of this paper is arranged as follows. Section 2 presents a literature review of related papers. Section 3 is about the materials and techniques used, and Section 4 presents the methodology that underpins the proposed approach. Section 5 discusses the experiments, results, and performance evaluations. The conclusions are presented in Section 6 along with an overview of the few limitations of the study and recommendations for future studies.

2. Related Studies

The approaches that have been described in the extant literature are thoroughly examined in this section. Surveys on deep neural network (DNN) methods for analyzing breast

cancer mammography images were conducted by the authors in [12,13]. The issue and the most current developments in the field were first discussed. The authors subsequently analyzed and compared the algorithms employed at each stage of image analysis using a variety of criteria. They examined the use of DNN technology to analyze mammography images for breast cancer and highlighted the difficulties of these models and potential future study areas. While considering the studies on breast cancer classification models using DL techniques, the majority of papers that offered superior performance were based on transfer learning.

The CNN classification models used in medical applications were investigated by the authors of [14] using the transfer learning technique. They performed tests to determine the efficacy of the transfer learning approach depending on a variety of variables, such as data size and the distance between the source data and train data. They drew the following conclusions from their study: ImageNet is the best database to apply transfer learning to image applications. Less inductive bias in the model and a smaller dataset size will produce better results. Finally, they stated that although their study was focused on medical applications, they anticipated that the outcomes would be the same for other applications.

The approach presented in [15] was based on transfer learning of a few well-known CNN architectures (Inception, ResNet, and VGG) to find masses in mammograms. The authors examined various architectures using a digitized dataset, i.e., the images in the dataset were manually scanned from mammography scans. The superiority of the three architectures was then tested using a second digital dataset of images that were directly obtained from the scanning system. The authors achieved a true positive rate of 0.98 ± 0.02 . They did not classify the tumors, which was beyond the scope of their research. However, another study [16] examined the prediction accuracy of four well-known CNN architectures (AlexNet, VGG, GoogleNet, and ResNet) based on two scenarios with two different datasets for each scenario. The initial scenario involved teaching each architecture from scratch without using predetermined weights. In the second scenario, pre-trained weights were used to fine-tune these structures. According to the findings, utilizing a fine-tuned architecture produced superior results to training the architectures from scratch. The best accuracy among all other architectures was achieved by ResNet50 and ResNet101. To verify the differences between different CNN architectures, the authors recommended supplementing the model with larger datasets.

The focus of the authors in [17] was on a hybrid transfer learning strategy for mammography-based breast cancer diagnosis. The VGG 16 network was used, and the final layers were adjusted. According to the findings, this model predicted breast tumors more accurately than the models it was compared against. The authors also noted that the model was only used for binary classifications in this research. However, they planned to use a similar strategy for a categorical classification and to incorporate other variables, such as tissue density, into the model. Another similar study [18] implemented transfer learning with ResNet50 and Nasnet-Mobile networks.

An automated DL-based BC diagnostic method [19] utilized the pre-trained ResNet34 for feature extraction and chimp optimization algorithm to optimize its parameters. They employed a wavelet neural network for classification. To enhance BC categorization, the transferable texture CNN was introduced in [20]. Eight DCNN models that were combined yielded deep features and robust features selected to differentiate breast tumors. In [21], the YOLOX model separated breast tissue to pinpoint regions of interest (ROI) that might contain lesions. The data are then run through the EfficientNet or ConvNeXt model to determine whether any ROIs are benign or malignant.

To explore DL features using supervised classification algorithms, the authors in [22] employed CNN technique coupled with support vector machine (SVM). This model involved image preprocessing and contrast enhancement utilizing the contrast-limited adaptive histogram equalization (CLAHE) technique. Later, images were manually segmented using ROI and automatically segmented using a region-based/threshold approach. During the feature extraction/selection and classification phases, the authors used an AlexNet-

based CNN architecture that was refined with an SVM classification algorithm connected to the last layer and pretrained the architecture with the ImageNet dataset. Additionally, they created a confusion matrix based on the forecasts made on two distinct datasets for evaluation before computing accuracy, precision, and F1-Score.

Another recent paper [23] employed three different DL CNN models as feature extractors: Inception-V3, ResNet50, and AlexNet. This approach makes use of the term variance (TV) feature selection technique to extract meaningful features, features are combined, and then another selection is applied and fed into a multi-class SVM classifier. The proposed method was tested using the image database of the Mammographic Image Analysis Society (MIAS). However, along with CNN features, handcrafted features, such as histogram of oriented gradients (HOG)-based, local binary pattern (LBP)-based, and shape features were fused and given to ML classifiers in another study [24].

In [25], a new CNN architecture that created a new O-net architecture by fusing two U-net architectures was offered. Finding features of various sizes and contrast levels required the suggested architecture to go through several convolutional and deconvolutional layers. Additionally, the authors demonstrated that the proposed design achieved an accuracy that exceeded that of comparable studies. They recommend that future research focus on improving its architecture.

The authors presented their dual CNN architecture solution in [26] for the simultaneous segmentation and classification of masses simultaneously. The foundation of their strategy was the use of two routes, the first of which is known as the locality-preserving learner and handles the large-scale regions of interest. The second route, dubbed conditional graph learner, oversees addressing ROIs of a small size. The authors concluded that their proposed system performed better than cutting-edge methods for the related application they utilized their model for.

A model with two phases was suggested in [27] as a method of diagnosing breast tumors. In the first phase, the model examined the entire image using a low-capacity memory efficient network to identify any potential benign or malignant features. The first phase's findings were employed in the second phase's greater capacity network, where the data from the two phases were combined to determine the categorization. The suggested solution performed better in terms of accuracy, speed, and memory utilization.

The abovementioned strategies were integrated by the authors [28], who conducted four experiments. The first employed a fully trained CNN model. The second method used CNN for feature extraction and SVM with various kernels for classification. The third involved the implementation of feature fusion to assess whether it enhanced or degraded the SVM's classification accuracy. Fourth, principal component analysis (PCA) was performed to condense a vast feature vector, which minimizes the number of variables that are targeted while barely altering the overall importance of the information seen. According to their research, among the four studies, the third approach, which used merged characteristics, had the highest accuracy.

In the evaluated studies, the transfer learning principle was used to create the majority of the high-performance models. While other efforts concentrated on creating CNN models from scratch, they failed to attain higher accuracy. Some models were made specifically for the detection of lumps, while others classified breast cancer as benign or malignant. No precise models were discovered that could conduct multiple classifications. The strategies that increase classification prediction accuracy were not examined in any of the publications we reviewed here. This study examines a model that is formed of both built-from-scratch and pre-trained architectures. A Bayesian optimization approach is used to determine the optimal hyperparameters in each structure, including those that are generated from scratch and additional layers that are added on top of pre-trained layers. These hyperparameters' effects are examined in depth. Additionally, three models are trained, and their outputs are combined to predict the outcome. This analysis of predictions is performed for each model evaluated. Finally, we conduct data augmentation [29] on the training dataset to balance out all the classes such that they have roughly the same percentage of the dataset.

After that, we compare how the model performs when trained on the dataset before and after augmentation. Since three CNNs together make the prediction decisions, the final model performance could be enhanced along with sufficient reliability, which are the key requirements when applying ML techniques in the medical field.

3. Materials and Methods

In this section, an overview of existing techniques that are used and the way they operate will be discussed and elaborated. A workflow outline of the proposed model is given in Figure 1.

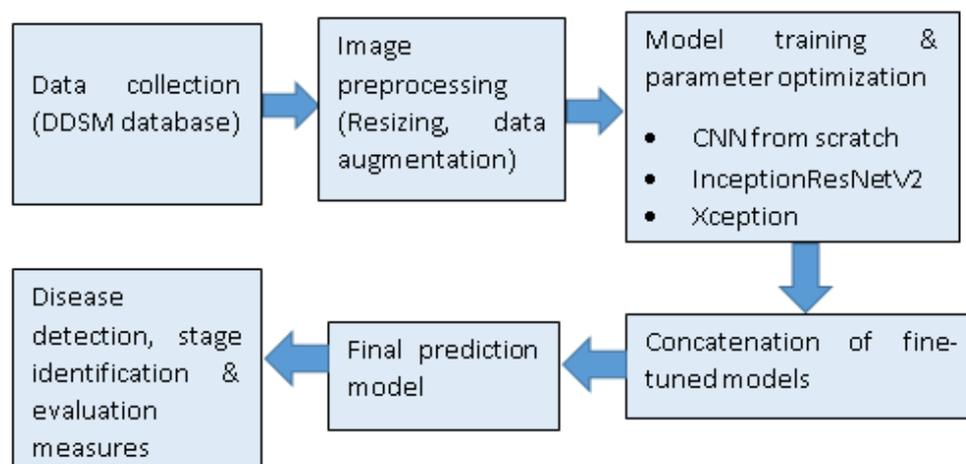


Figure 1. Workflow outline.

3.1. CNN Models

A CNN model [30] is an artificial neural network (ANN) algorithm that consists of input and output layers, with a variable number of hidden layers in between, depending on how the model is built. The hidden layers of a CNN model are blocks of numerous types of default and custom-made layers, some of which are used for feature extraction and others of which are used to map the features extracted into final outputs, which are mostly either raw features or classification predictions. These blocks of layers build a hierarchical model that has neurons in each level after feature extraction to simulate the way the human brain's neural network works. Each connection between 2 neurons in the model gets assigned a weight and bias values where the weight is a presentation of the importance of the feature that this neuron represents, and it is multiplied by the value that enters the neuron through its connection, and the bias is a value that is added to it to shift the activation function to the left or to the right.

The CNN model in this study is built using a convolutional layer to extract the essential features, pooling layers to summarize the features, a fully connected layer to perform classification, dropout layers to drop some of the features generated by previous layers randomly based on a preset probability and activation function layers to add nonlinearity to the process. The architecture built by arranging these layers will vary the performance offered by the CNN model.

3.2. Transfer Learning Technique

The transfer learning technique [31] uses a preexisting model or a part of it as a segment of another model. It can be retrained from scratch while preserving the layers, and the hyperparameters that are set on them, or part of the layers of it can be set as nontrainable to keep the features that the model trained on. Since transfer learning uses a pre-trained model, it will require less data so that the model can be trained on another dataset, which makes this technique efficient for use with small datasets. Additionally, it is

simpler and easier than building new models from scratch, so it enhances the process of building CNN models.

3.3. Bayesian Optimization

Bayesian optimization algorithm [32] is used to mathematically try to find the best set of parameters that will obtain a black box function to achieve the best output results. This algorithm is used when the function is a black box with no closed form and when it is expensive to evaluate the function. The algorithm starts with several sample points and evaluates them to determine their results. Then, using the previously acquired points, it computes a function called the surrogate function to build a surrogate model, which approximates the true objective function. Next, it iterates a loop where it adds an additional point using an acquisition function, which is the selection function that selects the parameter where it is maximized, evaluates the newly added points, and then re-evaluates the surrogate function. This process is repeated until the preset number of iterations is reached, or there are no more max points that were not evaluated after evaluating a surrogate function at an iteration.

3.4. Experimental Setup

The experiment was conducted using an Nvidia GeForce RTX 2070 Super GPU, AMD Ryzen 9 3900X 12-core CPU, 32 GB RAM, and 8 GB VRAM with a Windows 10 Pro operating system. In addition, the programming part was developed using Python 3.8, the Jupyter notebook, Keras, TensorFlow, and GPy libraries.

3.5. Evaluation Criteria

The proposed solution is evaluated based on test data prediction results. The terms used are: True Positive (TP), where the model predicts that there are cancerous cells and the patient also has cancerous cells, True Negative (TN), where the model predicts that there are no cancerous cells and the patient does not have cancerous cells, False Positive (FP), where the model predicts that there are cancerous cells but the patient does not have cancerous cells, and False Negative (FN), where the model predicts that the patient does not have cancerous cells and the patient actually has cancerous cells. These data were collected based on the predictions. Subsequently, the evaluation parameters were calculated and compared with similar studies. The evaluation parameters were as follows:

Accuracy measures the percentage of accurate predictions out of the total predictions (Equation (1)).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision measures the percentage of correct positive predictions out of the total positive predictions (Equation (2)).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Specificity was defined as the percentage of correct negative predictions that were correct out of the total negative predictions (Equation (3)).

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

Recall is defined as the percentage of correct positive predictions out of the total positive samples in the dataset (Equation (4)).

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

The F1-score measures the balance between precision and recall in a model (Equation (5)).

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

Different models may misclassify the inputs by focusing on incorrect features. While concatenating these models, if any two of the three models predict the correct class, integrating their predictions into a single model may result in more accurate predictions.

4. Proposed Methodology

Since CNN models have been shown to perform better than more traditional methods in this field [12,15], they make up most of the models proposed in this paper. Although numerous DL frameworks are available, the field is broad, and room for improvement remains. Additionally, due to the widespread appearance of lesions and their varied intensity distributions, which can cause benign diseases to be mistaken for malignant abnormalities, the automatic detection and identification of breast cancer using DL models from mammography scans is a difficult issue. As a result, we construct the complete framework in two phases to increase the prediction accuracy.

One CNN model is created from scratch in the first phase, while two other models (InceptionResNetV2 and Xception) are created using transfer learning principles [33]. The feature extraction capability, classification performance, etc., of each CNN depend on several parameters in the CNN architecture, including the number of layers, filter size, number of filters, etc. Hence, each CNN model will be making classification decisions based on the features that are extracted by them that are unique. They are adjusted using hyperparameter optimization, and the dataset is balanced using augmentation. The models are then concatenated to create a model that offers enhanced binary and multi-classification accuracy performance. By following this strategy, we rely on one of three different networks, where each network can extract unique features responsible for making the diagnosis decisions, and the concatenated model offers prediction model reliability in addition to improved diagnosis accuracy. Additionally, the proposed study performs multi-classification of the cancer state with values ranging from 0 to 4. The overall framework, divided into different steps, is detailed below.

4.1. Preprocessing

The accuracy of the model's classification of the images and the length of time needed to train the model are both significantly improved by preprocessing the images. The input images are first preprocessed by being resized to 100×100 and rendered in grayscale. The negative cases are left untouched, while the images for the positive cases have ROIs extracted from them using masks while still leaving a border around the ROIs. The original dataset is used to create the augmented dataset, which is then processed with a 40-degree rotation, 20% width and height shifts, 20% sheer and zoom, 20% random vertical-horizontal flipping, etc. Finally, during these procedures, if pixel filling is necessary, it is set to use the closest available pixel.

The training dataset for the binary classification consists of 87% of images, which are 38,877 samples labeled with class 0, and 13% of images, which are 5831 samples labeled with class 1. After performing data augmentation on images, each class consists of 38,877 samples and 77,754 samples in total. The dataset for multi-classification consists of 87% of images, which are 38,877 samples labeled with class 0, 3.8% of images, which are 1682 samples labeled with class 1, 3.4% of images, which are 1529 samples labeled with class 2, 2.6% of images, which are 1170 samples labeled with class 3, and 3.2% of images, which are 1450 samples labeled with class 4. After performing data augmentation on images of classes 1, 2, 3, and 4, the dataset became 20% of each class, where each class consisted of 38,877 samples, and the total number of samples for the training dataset was 194,385.

4.2. Single Built from Scratch CNN Model

By using the preprocessed images, the first CNN model is developed from scratch. For that, we designed the CNN architecture where SeperableConv2d is utilized in this model's convolutional layers because it performs a depthwise spatial convolution on distinct channels before performing a pointwise convolution to combine the output findings, which makes it faster than the default convolutional layer. The kernel sizes utilized alternate between 3×3 and 2×2 . The number of filters, pool sizes, neurons, and dropout percentage all need to be optimized. Figure 2 depicts the design of this model.

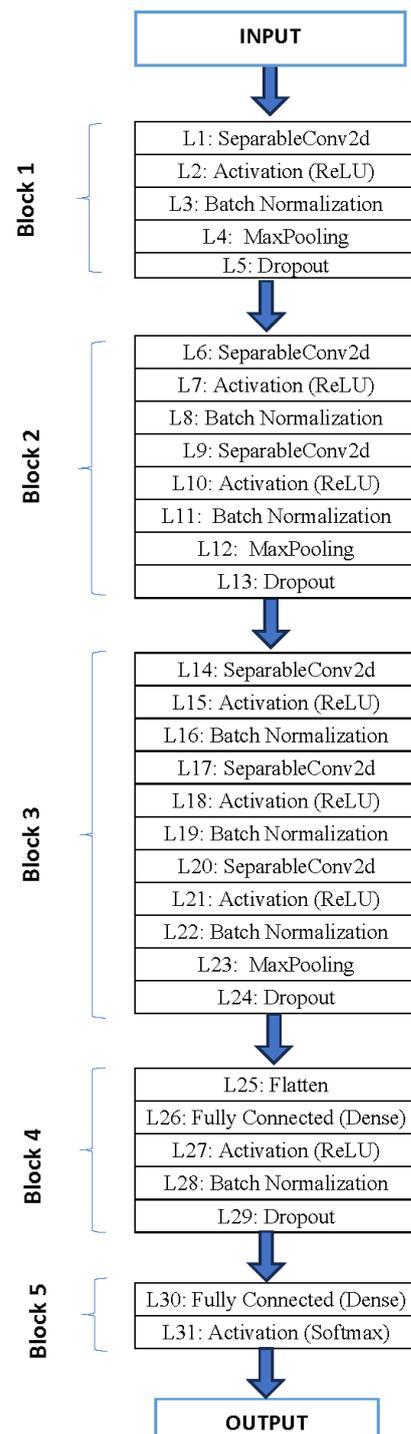


Figure 2. Layer architecture of the custom-built CNN model.

The feature extraction procedure is carried out using the first three blocks, and the classification process is carried out using the next two blocks. Starting with a flattened layer in the fourth step, the 2×2 array of features is converted into a vector. The completely linked layer of the fifth block has 2 neurons for binary classification of the inputs into solely positive or negative situations and 5 neurons for multi-classification.

4.3. Transfer Learning Model Based on the InceptionResNetV2 Model

The second CNN model is built on InceptionResNetV2 [34,35] and uses a copy of this architecture as its foundation before more layers are added. Hyperparameter optimization is used to optimize three dropout rates and two numbers of neurons in fully linked layers. Figure 3 depicts the model's overall architecture design.

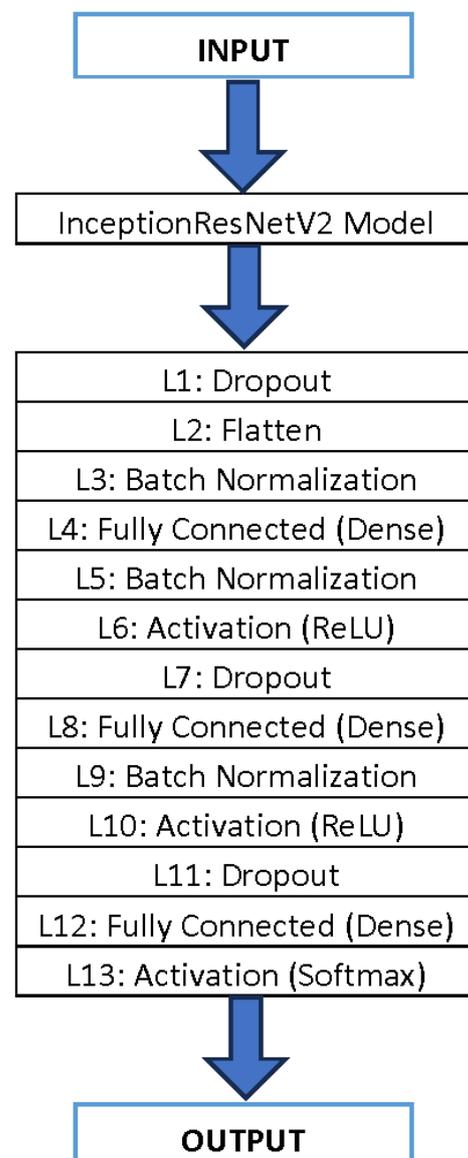


Figure 3. High-level architecture of the transfer learning model based on the InceptionResNetV2 model.

4.4. Transfer Learning Model Based on the Xception Model

The third model takes the form of a CNN framework that is an adaptation of the Xception model [36,37]. Five values in this model are optimized using the hyperparameter method: Three dropout percentages and two numbers of neurons in fully linked layers. The high-level architecture design for the model is shown in Figure 4.

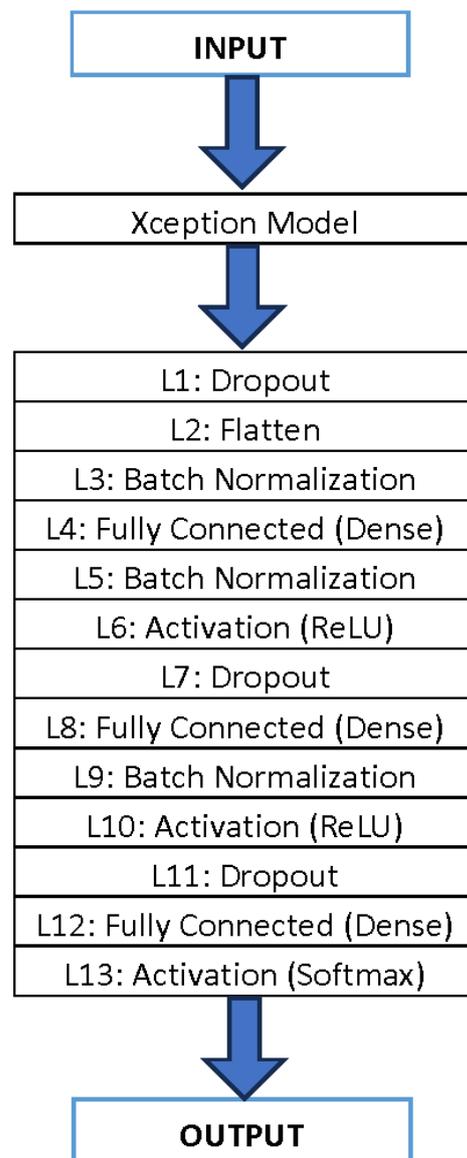


Figure 4. High-level architecture of the transfer learning model based on the Xception model.

4.5. The Triple CNN Concatenated Model

This model essentially combines the three concepts that were previously discussed into a single framework. The custom-constructed architectures, InceptionResNetV2, and Xception transfer learning models are combined, and the results are combined, as illustrated in Figure 5, by linking each model's final layer to the final fully connected layer in the primary model. During this process, a single completely linked layer is introduced without any hyperparameter adjustment. To discover the ideal hyperparameters for the layers, the Bayesian optimization [38] technique is applied to each network. In the convolutional layers, the values 16, 32, 64, and 128 are assigned to the number of filters. Additionally, the maximum pooling settings for filter sizes are 2×2 , 3×3 , or 4×4 . Furthermore, fully connected layers have 16, 32, 64, 128, 256, 512, or 1024 neurons. The dropout rates are configured to be a float value between 0 and 0.5 or a percentage between 0% and 50%. The motivation for developing this model is that different models extract information in varying ways and quantities. Different models may, therefore, misclassify some inputs because of focusing on the incorrect features. Therefore, combining the predictions of three models into a single model may improve predictions when only two of the three models predict the proper classification.

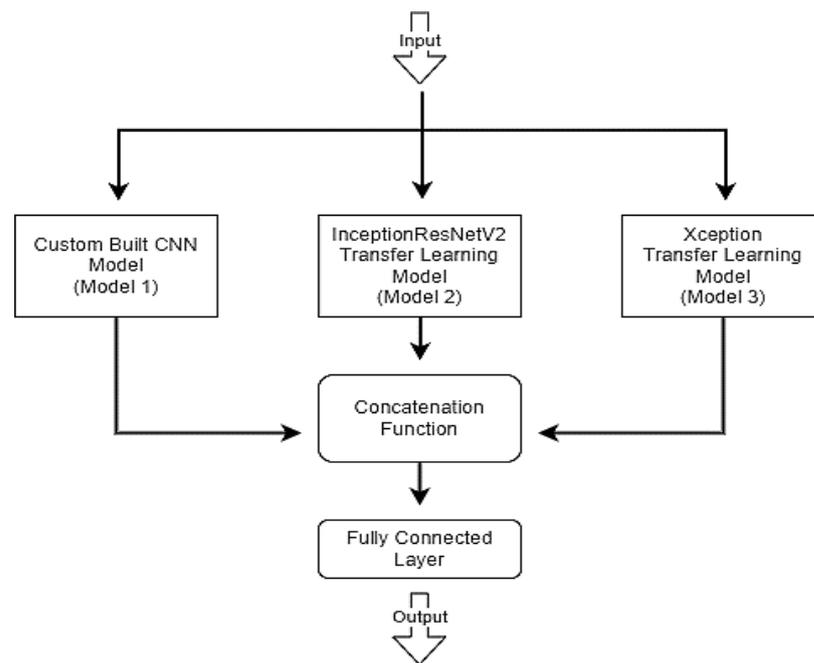


Figure 5. High-level architecture of the triple CNN concatenated model.

5. Results and Discussion

The experimental findings for each phase of the proposed model are presented below, along with tables, graphs, and statements.

5.1. Dataset

Using the Digital Database for Screening Mammography (DDSM) dataset, the architecture is trained and tested [39]. There are 55,890 images in total, 13% of which are positive instances, while 87% are negative cases. The DDSM [40] and Curated Breast Imaging Subset of DDSM (CBIS-DDSM) [41] datasets are used to build this dataset. To make dataset reuse with classification models simpler, the images in this dataset are increased to 3 RGB channels from 299×299 and grayscale. Each image in this dataset has two different types of labels, one of which is either 0 or 1, with 0 being a negative case and 1 a positive case. The second is an integer between 0 and 4, with 0 denoting a negative case, 1 denoting a benign calcification, 2 denoting a benign mass, 3 denoting a malignant calcification, and 4 denoting a malignant mass. Figure 6 displays a sample of the dataset with the labels placed on top of each image. The dataset was split into two parts: 20% for model testing and 80% for training, of which 30% was used for validation after each epoch.

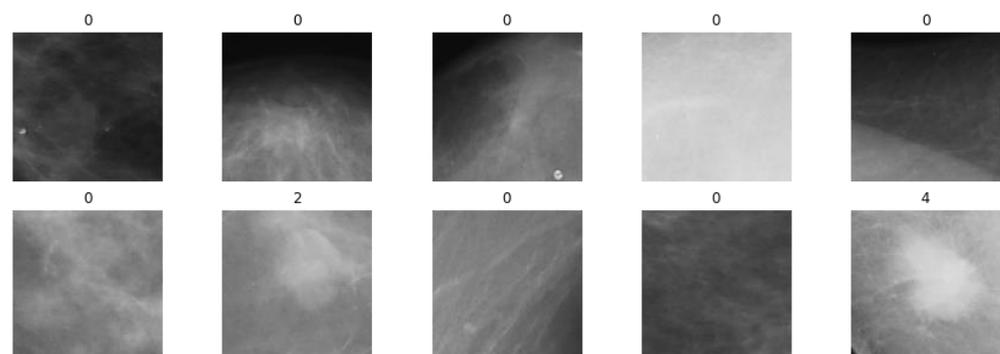


Figure 6. Samples of the dataset.

During the process of training DL models, the training parameters are set as 50 epochs, early stopping callback, verbose 1, and a split ratio of 30%.

5.2. CNN Model Development and Training

Three models, one built from scratch and the other two (InceptionResNetV2, Xception) by transfer learning techniques, are developed and refined.

5.2.1. Results and Analysis of Single Built-from-Scratch CNN Model

Before data augmentation, this model runs a binary classification test where it predicts the class of test datasets with an accuracy of 93.03% in the first iteration and the best accuracy result of 95% in the twenty-fourth iteration. Following data augmentation, it predicts with an accuracy of 94.53%. In the fifth iteration, as shown in Figure 7, it achieves the highest accuracy of 95.86%. Over the course of 25 iterations, the optimizer discovers hyperparameters with prediction accuracy improvements of 2% over the first iteration and 8.4% over iteration #6, which exhibits the lowest accuracy among the iterations performed prior to data augmentation. Additionally, it has a prediction accuracy that is 1.33 percent greater than the initial iteration and 1.86 percent better than iteration #23's accuracy, which is the lowest among the iterations following data augmentation.

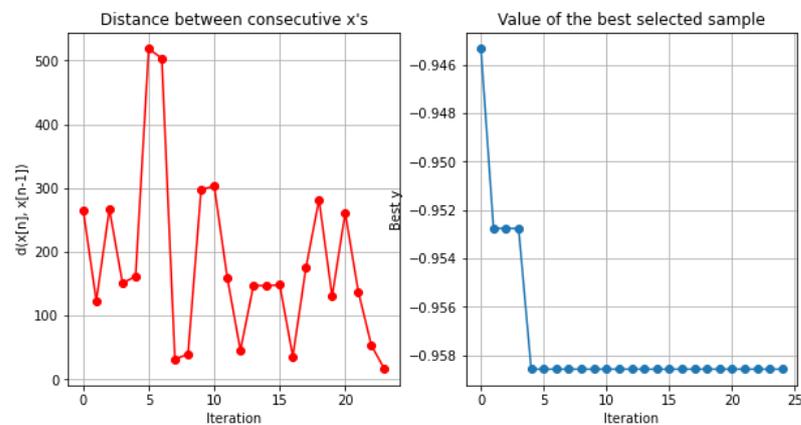


Figure 7. Hyperparameter optimization chart of a single built-from-scratch CNN model on binary classification after data augmentation.

Before data augmentation, this model executes a multi-classification run where it predicts the class of testing datasets with an accuracy of 89.8% in the first iteration and an accuracy of 90.74% in the fourteenth iteration. After data augmentation, it forecasts with the best accuracy being 93.33% in the twenty-first iteration, as shown in Figure 8. It has an accuracy of 92.17% in the first iteration. The optimizer discovers hyperparameters after 25 iterations that improve prediction accuracy by 0.94% compared to the first iteration and 1.6% compared to iteration #22, which has the lowest accuracy from the iterations prior to data augmentation. Additionally, it outperforms earlier iterations in terms of prediction accuracy, outperforming iteration #1 by 1.16% and iteration #18 by 3.45%, which exhibits the worst accuracy among iterations following data augmentation.

Before data augmentation, the model with the best hyperparameters for binary classification contains 26,736 total parameters, 25,904 of these are trainable, and the remaining 832 cannot be trained. Additionally, it has 102,461 total parameters after data augmentation. Of those, 101,373 can be trained, while the remaining 1088 cannot. In contrast, multi-classification has 44,928 parameters before data augmentation, of which 44,096 are trainable and the remaining 832 are not. In contrast, it has 42,144 parameters after data augmentation, of which 41,440 are trainable and 704 are not. Table 1 provides information on the optimum hyperparameters of the CNN model.

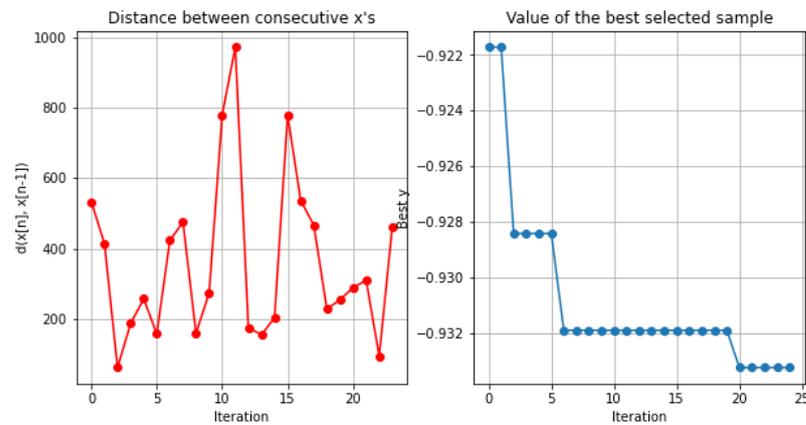


Figure 8. Hyperparameter optimization chart of a single built-from-scratch CNN model on multi-classification after data augmentation.

Table 1. Optimized hyperparameters for the CNN model built from scratch.

Layer #	Hyperparameter	Binary Classification		Multi-Classification	
		Value before Data Augmentation	Value after Data Augmentation	Value before Data Augmentation	Value after Data Augmentation
1	Number of filters	128	128	64	64
4	Pool size	2 × 2	2 × 2	2 × 2	3 × 3
5	Dropout rate	8%	18%	16%	42%
6	Number of filters	64	128	128	32
9	Number of filters	32	32	64	32
12	Pool size	3 × 3	3 × 3	3 × 3	3 × 3
13	Dropout rate	19%	19%	43%	36%
14	Number of filters	32	64	64	128
17	Number of filters	128	64	32	32
20	Number of filters	16	64	32	32
23	Pool size	4 × 4	4 × 4	4 × 4	2 × 2
24	Dropout rate	22%	26%	7%	12%
26	Number of Neurons	16	64	32	32
29	Dropout rate	39%	7%	49%	39%

5.2.2. Results and Analysis of the Transfer Learning Model Based on the InceptionResNetV2 Model

Before data augmentation, when this model is used for binary classification, it commences by correctly predicting the class of test datasets in the first iteration with 95.65% accuracy and ultimately produces the best accuracy result in the twenty-second iteration with an accuracy of 96.1%. As shown in Figure 9, the best model is trained in the third iteration with an accuracy of 96.4% compared to the initial accuracy of 94.64% before data augmentation. The optimizer discovers hyperparameters after 25 iterations that have a prediction accuracy of 0.45% higher than the first iteration and 0.54% higher than iteration #6, which has the lowest accuracy among the iterations conducted prior to the addition of additional data. Additionally, it exhibits a 1.76% improvement in accuracy over the initial iteration, which has the worst accuracy among the subsequent rounds of data augmentation.

Before data augmentation, the multi-classification run of this model starts by accurately predicting the class of testing datasets in the first iteration with 92.2% accuracy. It produces the highest accuracy results in the ninth iteration, with 92.42% accuracy. After augmentation, it begins forecasting with 95.31% accuracy, and in the twentieth iteration, as shown in Figure 10, it achieves the highest accuracy of 98.68%. The optimizer discovers hyperparameters after 25 iterations that improve prediction accuracy by 0.22% compared

to the first iteration and 1.6% compared to iteration #22, which has the lowest accuracy from the iterations prior to data augmentation. However, following data augmentation, it outperforms the first iteration, which has the lowest prediction accuracy across all iterations, with a prediction accuracy of 3.37%.

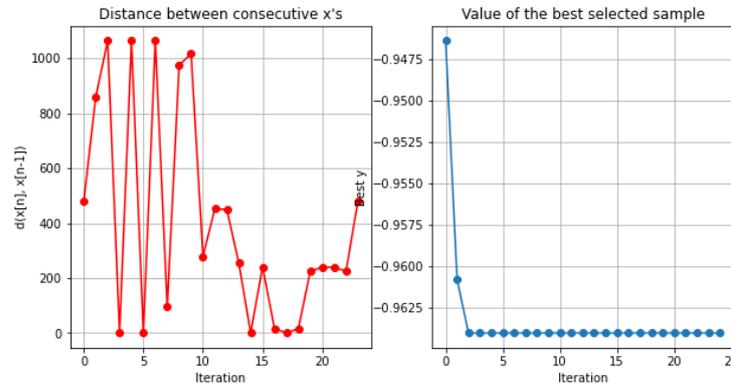


Figure 9. Hyperparameter optimization chart of the transfer learning model based on the Inception-ResNetV2 model for binary classification after data augmentation.

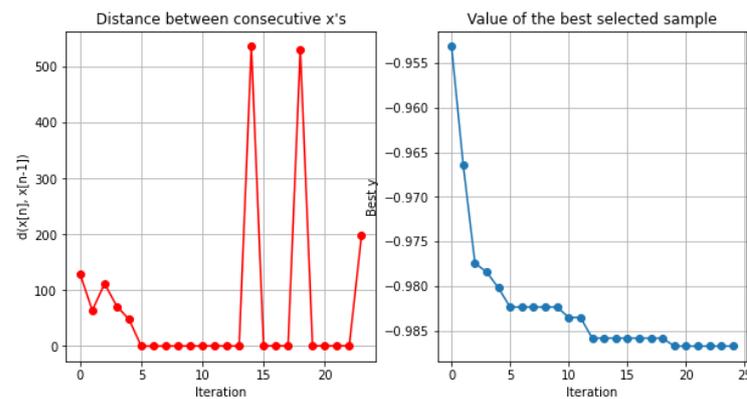


Figure 10. Hyperparameter optimization chart of the transfer learning model based on the Inception-ResNetV2 model on multi-classification after data augmentation.

Before data augmentation, the model with the best hyperparameters for binary classification has 54,368,274 total parameters, 54,304,562 of which are trainable, while the remaining 63,712 are not. After data augmentation, it has a total of 54,678,370 parameters, of which 54,612,450 are trainable and 65,920 are not. In contrast, multi-classification has a total of 56,185,573 parameters before data augmentation, of which 56,119,397 are trainable and the remaining 66,176 are not. Following data augmentation, it has 54,442,693 parameters, of which 54,378,917 are trainable and 63,776 are not. Table 2 provides information on the CNN model’s optimized hyperparameters based on the InceptionResNetV2 network.

Table 2. Optimized hyperparameters for the transfer learning model based on the InceptionResNetV2 model.

Layer #	Hyperparameter	Binary Classification		Multi-Classification	
		Value before Data Augmentation	Value after Data Augmentation	Value before Data Augmentation	Value after Data Augmentation
1	Dropout rate	25%	26%	38%	28%
4	Number of Neurons	16	128	1024	64
7	Dropout rate	43%	1%	11%	41%
8	Number of Neurons	32	1024	256	16
11	Dropout rate	31%	17%	42%	25%

5.2.3. Results and Analysis of the Transfer Learning Model Based on the Xception Model

Prior to the addition of the augmentation data, this framework’s binary classification run accurately identifies the class of test datasets in the first iteration with 96.53% accuracy, and it achieves its highest accuracy in the seventeenth iteration with 96.64%. The best model, as shown in Figure 11, starts with 96.19% accuracy after data augmentation and finishes with 96.68% accuracy in the second iteration. After 25 iterations, the optimizer finds hyperparameters with predictions that were 0.11 percent more accurate than the first iteration and 0.51 percent more accurate than iteration 24, which has the lowest accuracy of the iterations prior to data augmentation. The model also finds hyperparameters with a 0.68% greater accuracy than iteration #4, which has the lowest accuracy of the iterations and is 0.49% higher than the first iteration following data augmentation.

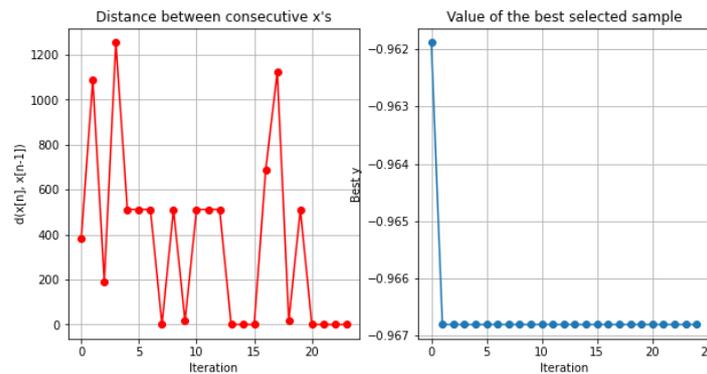


Figure 11. Hyperparameter optimization chart of the transfer learning model based on the Xception model on binary classification after data augmentation.

This model starts predicting the class of testing datasets prior to data augmentation in the multi-classification run with an accuracy of 93.09%, which is the highest across all iterations. Furthermore, following data augmentation, it begins to predict with an accuracy of 97.82% in the first iteration and achieves its best outcomes in the twentieth iteration with an accuracy of 99.08%, as shown in Figure 12. After 25 iterations, the optimizer identifies the hyperparameters that have the highest prediction accuracy in the first iteration, as well as an accuracy that is 1.27% greater than that of iteration #14—the iteration with the lowest accuracy prior to data augmentation. After data augmentation, the prediction accuracy is 1.27% greater than iteration #9, which has the lowest accuracy of all the iterations and is 1.26% higher than the first iteration.

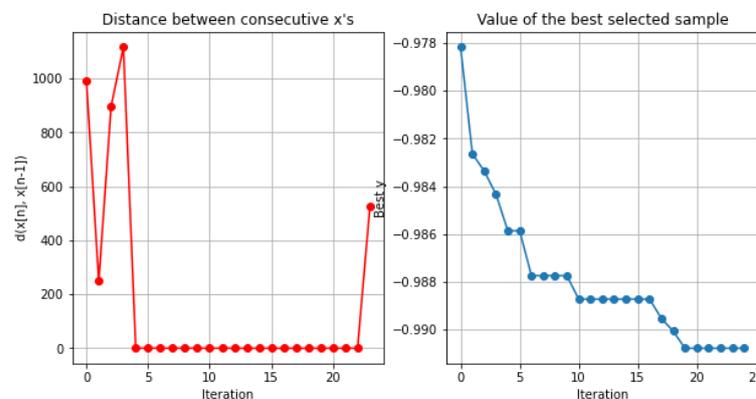


Figure 12. Hyperparameter optimization chart of the transfer learning model based on the Xception model on multi-classification after data augmentation.

Before data augmentation, the binary classification model with the best hyperparameters has 21,236,090 parameters in total. There are 91,936 untrainable parameters and

21,144,154 trainable parameters. Additionally, after data augmentation, it has a total of 21,254,162 parameters, of which 21,452,682 are trainable and 92,480 are not. Furthermore, it has a total of 21,243,517 parameters in the multi-classification before data augmentation, of which 21,151,069 are trainable and 92,448 are not. After data augmentation, it also has 21,243,517 parameters, of which 21,151,069 are trainable and 92,448 are not. Table 3 provides information on the CNN model’s optimal hyperparameters based on the Xception network.

Table 3. Optimized hyperparameters for the transfer learning model based on the Xception model.

Layer #	Hyperparameter	Binary Classification		Multi-Classification	
		Value before Data Augmentation	Value after Data Augmentation	Value before Data Augmentation	Value after Data Augmentation
1	Dropout rate	28%	5%	34%	29%
4	Number of Neurons	16	32	16	16
7	Dropout rate	28%	25%	3%	45%
8	Number of Neurons	256	512	512	512
11	Dropout rate	23%	35%	40%	28%

5.3. Triple CNN Concatenated Model

Before data augmentation, the model accurately predicted the outcomes for binary classification (96.74%) and multi-classification (93.9%). After data augmentation, an accuracy of 97.26% in binary classification and 99.13% in multi-classification is achieved. According to [42], Tables 4 and 5 contain the performance requirements for each model for binary and multicategorization both before and after augmentation. The results demonstrate that, both before and after data augmentation, the concatenated model outperforms the three models that were utilized to individually develop this triple CNN model. Before data augmentation, the binary classification model takes 2 h and 27 min to train and evaluate, and the multi-classification model takes 2 h and 34 min. Additionally, following data augmentation, the binary classification takes 4 h and 50 min to train and evaluate, while the multi-classification model needs 2 h and 8 min.

Table 4. Comparison between the binary classification evaluation performances of the studied models.

Model	Augmentation	Class	Precision	Specificity	Recall	F1-Score	Accuracy (%)
Single Built from Scratch CNN model	Without	0	96.1	73.7	98.2	97.1	95
		1	86	98.2	73.7	79.4	
		0	97	80	98.2	97.6	
Transfer learning model based on InceptionResNetV2 model	With	1	87	98.2	80	83	95.86
		0	96.8	78	98.7	97.8	
		1	90	98.7	78	84	
Transfer learning model based on Xception model	Without	0	97.3	82	98.6	97.9	96.4
		1	90	98.6	82	86	
		0	97.38	82	98.8	98.08	
Triple CNN concatenated model	With	1	91	98.8	82	86	96.64
		0	98	86	98.2	98.09	
		1	88	98.2	86	87	
Triple CNN concatenated model	Without	0	97.2	81	99.08	98.15	96.74
		1	93	99.08	81	87	
		0	97.88	85.73	99	98.43	
	With	1	92.73	99	85.73	89.1	97.26

Table 5. Comparison between the multi-classification evaluation performances of the studied models.

Model	Augmentation	Class	Precision	Recall	F1-Score	Accuracy (%)
Single Built from Scratch CNN model	Without	0	95	99	97	90.74
		1	50	37	42	
		2	52	33	40	
		3	42	29	37	
		4	52	48	50	
	With	0	97	98	98	93.33
		1	62	54	57	
		2	70	65	67	
		3	54	55	54	
		4	70	73	71	
Transfer learning model based on InceptionResNetV2 model	Without	0	96	99	97	92.42
		1	66	33	43	
		2	68	60	64	
		3	41	39	40	
		4	70	60	65	
	With	0	99	100	99	98.68
		1	95	87	91	
		2	98	95	96	
		3	97	92	95	
		4	97	98	98	
Transfer learning model based on Xception	Without	0	96	99	98	93.09
		1	64	43	51	
		2	72	58	64	
		3	45	43	44	
		4	74	62	67	
	With	0	99	100	99	99.08
		1	93	93	93	
		2	97	97	97	
		3	99	95	97	
		4	99	99	99	
Triple CNN concatenated model	Without	0	96	100	98	93.09
		1	68	38	49	
		2	72	62	67	
		3	47	39	43	
		4	73	60	66	
	With	0	99	100	100	99.13
		1	96	91	94	
		2	98	97	98	
		3	99	95	97	
		4	99	99	99	

5.4. Final Results and Comparisons

With an accuracy of 97.26%, the study's final model correctly predicts the testing dataset for binary classification based on a dataset of 77,754 samples. Additionally, it accurately predicts the testing dataset in the multi-classification, which has a dataset size of 194,385 samples, with 99.13% accuracy. The binary dataset and the multi-class dataset are identical. The samples labeled as 1 in the binary dataset are split into classes 1 to 4 in the multi-class dataset, the only variation being the labels. By treating predictions from classes 1–4 as positive cases and predictions from class 0 as negative cases, the framework trained on the multi-class model may be used for binary classification with at least the same accuracy. Table 6 shows that the suggested model outperforms all other models that have been researched and trained using the same dataset.

Table 6. Comparison between the accuracy of the proposed model with studied models.

Model	Accuracy
CNN model for detection of architectural distortion [43]	87.50%
lightweight end-to-end improved CNN [44]	97.20%
VGG16 [44]	95.04%
SVM [44]	92.23%
PCA + SVM [44]	90.59%
Deep Vision supervised learning [45]	97%
CNN Model with Transfer learning in binary classification [46]	92%
CNN Model with Transfer learning in multi-class classification [46]	85%
Haze-reduced local-global + EfficientNet-b0 [47]	95.4%
Stacked ensemble of residual neural networks [48]	85.39%
CNN with less learnable parameters [49]	90.68%
Case-based reasoning system [50]	86.71%
DL feature fusion + satin bowerbird optimization-controlled Newton Raphson feature selection [51]	94.5%
Proposed model in binary classification	97.26%
Proposed model in multi-classification	99.13%

6. Discussion

Automatic breast cancer diagnosis from mammography scans becomes challenging, especially because of the difficulty in identifying malignancy from the lesion appearance, which results in confusing intensity distributions and leads to the misidentification of malignant abnormalities as benign. Hence, a reliable model with improved performance for disease-detection stage identification is employed here. Binary classification performs disease detection, and multi-classification provides stage identification. Because accuracy is the most important metric when applying an ML technique for medical image classification applications, techniques for improving prediction accuracy were also investigated.

Because Xception-based CNN models are more efficient in BC classification, as shown in Table 5, we designed a CNN based on its architecture. To improve the prediction accuracy, we concatenated these three models (Xception from transfer learning, InceptionResNetV2 from transfer learning with a similar architecture, and CNN built based on Xception architecture). Concatenating the three models in this manner is the main innovation of the network structure. In addition, we showed in the manuscript how this network structure with Bayesian optimization and augmentation improves performance.

One of the main issues regarding ML model development is the limited availability of labeled data. Hence, data balancing was performed through augmentation. Usually, transfer learning networks perform better even with less data, but for a built-from-scratch CNN, we need sufficient data to perform better. With transfer learning, InceptionResNetV2 and Xception achieved an accuracy of 96.4% and 96.68%. However, with data augmentation, our CNN from scratch also obtained a comparable accuracy of 95.86%, and its parameters were fine-tuned by Bayesian optimization. The concatenated model performed better, with 97.26% accuracy for disease prediction.

The data augmentation effect was more effectively visible with the transfer learning networks when considering multi-classification because of the large variation in the number of data belonging to each class. Here, InceptionResNetV2 and Xception showed accuracy improvements of 6.26% and 5.99%, respectively, with augmentation. The CNN from scratch could provide only 93.33% accuracy. However, concatenating it with transfer learning models enhanced the accuracy of disease stage identification by 99.13%.

From these results, it is clear that different models extract information in different ways and to different extents. Hence, the model performance of each network varied. Because a CNN model may concentrate on the wrong features, it may misclassify some inputs. Therefore, when only two of the three models predicted the correct categorization, integrating the predictions of the three models into one enhanced the predictions. Therefore, the final model can provide diagnostic prediction reliability, in addition to improved accuracy, which makes the model appropriate in the medical field.

Limitations and Future Work:

This study had some limitations, including data paucity and hardware constraints. Massive amounts of memory and computing power are required for ML algorithms. ML is a probabilistic mathematical model that involves numerous computations. Millions of parameters may need to be calculated and updated during the runtime for the model. The task requiring the most computer power is to train the ML model. A higher processing power translates to more rapid framework training and evaluation. The more data there are, the more accurate these frameworks are. Creating synthetic data samples from those that are already accessible is one approach to circumvent this problem. To further enhance the results, future research could use the methodology described in this paper on larger datasets with more real samples and machines with more memory and processing power. Moreover, the possibility of combining single models made using recent techniques with higher accuracy levels than those used here will be investigated.

7. Conclusions

The purpose of this study was to develop a precise and reliable model for breast cancer classification as well as to examine the possibility of improving CNN classification accuracy by combining multiple models while optimizing the hyperparameters in each model and balancing and enriching the datasets that the models are trained with. The findings reveal that the combined models perform better than any one model working in isolation. Additionally, for each model, we can observe how varying the hyperparameters leads to varying degrees of accuracy and how an optimization approach such as Bayesian optimization can be beneficial. The proposed triple-concatenated framework is able to improve the accuracy from 96.74% to 97.26% in binary classification and from 93.09% to 99.13% in multi-classification. In addition to the improvement in prediction accuracy, the misclassifications can be reduced, and reliability can be ensured since we are relying on multiple model networks for the final decision. These characteristics make the suggested work appropriate for medical field applications.

These results can be further enhanced by combining single models that have higher accuracy levels than those employed here and by boosting the dataset's size with additional actual or augmented samples. To create better models for various types of datasets, these strategies can also be applied to other applications. The proposed technique would be applicable to any image classification problem, especially in medical image classification for disease diagnosis and stage identification applications where classification accuracy, reduced misclassification, and decision reliability matter.

Author Contributions: Conceptualization, M.H.A.; formal analysis, M.H.A.; investigation, M.H.A. and J.A.-B.; methodology, M.H.A. and J.A.-B.; software, J.A.-B.; supervision, M.H.A.; validation, M.H.A.; writing—original draft, M.H.A. and J.A.-B.; writing—review and editing, M.H.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this paper is publicly available [39].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. NCI. What Is Cancer? Available online: <https://www.cancer.gov/about-cancer/understanding/what-is-cancer> (accessed on 4 November 2022).
2. Global Cancer Observatory. Available online: <https://gco.iarc.fr/> (accessed on 4 November 2022).
3. Breast Cancer. Available online: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (accessed on 1 December 2022).
4. Cariolou, M.; Abar, L.; Aune, D.; Balducci, K.; Becerra-Tomás, N.; Greenwood, D.C.; Markozannes, G.; Nanu, N.; Vieira, R.; Giovannucci, E.L.; et al. Postdiagnosis recreational physical activity and breast cancer prognosis: Global Cancer Update Programme (CUP Global) systematic literature review and meta-analysis. *Int. J. Cancer* **2022**, *152*, 600–615. [CrossRef]

5. Azuero, A.; Benz, R.; McNees, P.; Meneses, K. Co-morbidity and predictors of health status in older rural breast cancer survivors. *SpringerPlus* **2014**, *3*, 102. [CrossRef]
6. Iqbal, J.; Ginsburg, O.; Rochon, P.A.; Sun, P.; Narod, S.A. Differences in breast cancer stage at diagnosis and cancer-specific survival by race and ethnicity in the United States. *JAMA* **2015**, *313*, 165–173. [CrossRef]
7. Alshayegi, M.H.; Ellethy, H.; Abed, S.; Gupta, R. Computer-aided detection of breast cancer on the Wisconsin dataset: An artificial neural networks approach. *Biomed. Signal Process. Control.* **2022**, *71*, 103141. [CrossRef]
8. Kaur, H. Dense Convolutional Neural Network Based Deep Learning Framework for the Diagnosis of Breast Cancer. *Wirel. Pers. Commun.* **2023**, *2023*, 1–16. [CrossRef]
9. Vaidehi, K.; Subashini, T.S. Automatic Characterization of Benign and Malignant Masses in Mammography. *Procedia Comput. Sci.* **2015**, *46*, 1762–1769. [CrossRef]
10. Yi, D.; Sawyer, R.L.; Cohn, D., III; Dunnmon, J.; Lam, C.; Xiao, X.; Rubin, D. Optimizing and Visualizing Deep Learning for Benign/Malignant Classification in Breast Tumors. May 2017. Available online: <https://arxiv.org/abs/1705.06362v1> (accessed on 5 August 2023).
11. Henkel, A.; Cooper, R.A.; Ward, K.A.; Bova, D.; Yao, K. Malignant-appearing microcalcifications at the lumpectomy site with the use of FloSeal hemostatic sealant. *Am. J. Roentgenol.* **2008**, *191*, 1371–1373. [CrossRef]
12. Abdelhafiz, D.; Yang, C.; Ammar, R.; Nabavi, S. Deep convolutional neural networks for mammography: Advances, challenges and applications. *BMC Bioinform.* **2019**, *20*, 281. [CrossRef]
13. Debelee, T.G.; Schwenker, F.; Ibenhal, A.; Yohannes, D. Survey of deep learning in breast cancer image analysis. *Evol. Syst.* **2019**, *11*, 143–163. [CrossRef]
14. Matsoukas, C.; Haslum, J.F.; Sorkhei, M.; Söderberg, M.; Smith, K. What Makes Transfer Learning Work for Medical Images: Feature Reuse & Other Factors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 9215–9224. [CrossRef]
15. Agarwal, R.; Diaz, O.; Lladó, X.; Yap, M.H.; Martí, R. Automatic mass detection in mammograms using deep convolutional neural networks. *J. Med. Imaging* **2019**, *6*, 031409. [CrossRef]
16. Tsochatzidis, L.; Costaridou, L.; Pratikakis, I. Deep Learning for Breast Cancer Diagnosis from Mammograms—A Comparative Study. *J. Imaging* **2019**, *5*, 37. [CrossRef]
17. Khamparia, A.; Bharati, S.; Podder, P.; Gupta, D.; Khanna, A.; Phung, T.K.; Thanh, D.N. Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. *Multidimens. Syst. Signal Process.* **2021**, *32*, 747–765. [CrossRef]
18. Alruwaili, M.; Gouda, W. Automated Breast Cancer Detection Models Based on Transfer Learning. *Sensors* **2022**, *22*, 876. [CrossRef]
19. Escorcia-Gutierrez, J.; Mansour, R.F.; Beleño, K.; Jiménez-Cabas, J.; Pérez, M.; Madera, N.; Velasquez, K. Automated Deep Learning Empowered Breast Cancer Diagnosis Using Biomedical Mammogram Images. *Comput. Mater. Contin.* **2022**, *71*, 4221–4235. [CrossRef]
20. Maqsood, S.; Damaševičius, R.; Maskeliūnas, R. TTCNN: A Breast Cancer Detection and Classification towards Computer-Aided Diagnosis Using Digital Mammography in Early Stages. *Appl. Sci.* **2022**, *12*, 3273. [CrossRef]
21. Huynh, H.N.; Tran, A.T.; Tran, T.N. Region-of-Interest Optimization for Deep-Learning-Based Breast Cancer Detection in Mammograms. *Appl. Sci.* **2023**, *13*, 6894. [CrossRef]
22. Ragab, D.A.; Sharkas, M.; Marshall, S.; Ren, J. Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ* **2019**, *7*, e6201. [CrossRef]
23. Elkorany, A.S.; Elsharkawy, Z.F. Efficient breast cancer mammograms diagnosis using three deep neural networks and term variance. *Sci. Rep.* **2023**, *13*, 2663. [CrossRef]
24. Cruz-Ramos, C.; García-Avila, O.; Almaraz-Damian, J.-A.; Ponomaryov, V.; Reyes-Reyes, R.; Sadovnychiy, S. Benign and Malignant Breast Tumor Classification in Ultrasound and Mammography Images via Fusion of Deep Learning and Handcraft Features. *Entropy* **2023**, *25*, 991. [CrossRef]
25. Rashed, E.; El Seoud, M.S.A. Deep learning approach for breast cancer diagnosis. In Proceedings of the 2019 8th International Conference on Software and Information Engineering, Cairo, Egypt, 9–12 April 2019; ACM International Conference Proceeding Series; pp. 243–247. [CrossRef]
26. Li, H.; Chen, D.; Nailon, W.H.; Davies, M.E.; Laurensen, D.I. Dual Convolutional Neural Networks for Breast Mass Segmentation and Diagnosis in Mammography. *IEEE Trans. Med. Imaging* **2020**, *41*, 3–13. [CrossRef]
27. Shen, Y.; Wu, N.; Phang, J.; Park, J.; Liu, K.; Tyagi, S.; Heacock, L.; Kim, S.G.; Moy, L.; Cho, K.; et al. An interpretable classifier for high-resolution breast cancer screening images utilizing weakly supervised localization. *Med. Image Anal.* **2021**, *68*, 101908. [CrossRef]
28. Ragab, D.A.; Attallah, O.; Sharkas, M.; Ren, J.; Marshall, S. A framework for breast cancer classification using Multi-DCNNs. *Comput. Biol. Med.* **2021**, *131*, 104245. [CrossRef]
29. Alshayegi, M.; Al-Buloushi, J.; Ashkanani, A.; Abed, S. Enhanced brain tumor classification using an optimized multi-layered convolutional neural network architecture. *Multimed. Tools Appl.* **2021**, *80*, 28897–28917. [CrossRef]
30. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]

31. Hosna, A.; Merry, E.; Gyalmo, J.; Alom, Z.; Aung, Z.; Azim, M.A. Transfer learning: A friendly introduction. *J. Big Data* **2022**, *9*, 102. [CrossRef]
32. Snoek, J.; Larochelle, H.; Adams, R.P. Practical Bayesian Optimization of Machine Learning Algorithms. *Adv. Neural Inf. Process. Syst.* **2012**, *25*.
33. Alshayegi, M.H.; Sindhu, S.C.; Abed, S. CAD systems for COVID-19 diagnosis and disease stage classification by segmentation of infected regions from CT images. *BMC Bioinform.* **2022**, *23*, 264. [CrossRef]
34. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017, Francisco, CA, USA, 4–9 February 2017; pp. 4278–4284. [CrossRef]
35. InceptionResNetV2. Available online: <https://keras.io/api/applications/inceptionresnetv2/> (accessed on 5 November 2022).
36. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807. [CrossRef]
37. Xception. Available online: <https://keras.io/api/applications/xception/> (accessed on 5 November 2022).
38. Zhang, Y.; Dai, Z.; Low, B.K.H. Bayesian Optimization with Binary Auxiliary Information. In Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence, UAI 2019, Tel Aviv, Israel, 22–25 July 2019. [CrossRef]
39. DDSM Mammography | Kaggle. Available online: <https://www.kaggle.com/datasets/skooch/ddsm-mammography> (accessed on 5 November 2022).
40. Heath, M.D.; Bowyer, K.; Kopans, D.; Moore, R.H. The Digital Database for Screening Mammography. In Proceedings of the 5th International Workshop on Digital Mammography, Toronto, ON, Canada, 11–14 June 2000; pp. 212–218.
41. Lee, R.S.; Gimenez, F.; Hoogi, A.; Miyake, K.K.; Gorovoy, M.; Rubin, D.L. A curated mammography data set for use in computer-aided detection and diagnosis research. *Sci. Data* **2017**, *4*, 170177. [CrossRef]
42. de Diego, I.M.; Redondo, A.R.; Fernández, R.R.; Navarro, J.; Moguerza, J.M. General Performance Score for classification problems. *Appl. Intell.* **2022**, *52*, 12049–12063. [CrossRef]
43. Oyelade, O.N.; Ezugwu, A.E. A deep learning model using data augmentation for detection of architectural distortion in whole and patches of images. *Biomed. Signal Process. Control.* **2021**, *65*, 102366. [CrossRef]
44. Kumar, P.; Srivastava, S.; Mishra, R.K.; Sai, Y.P. End-to-end improved convolutional neural network model for breast cancer detection using mammographic data. *J. Def. Model. Simul.* **2020**, *19*, 375–384. [CrossRef]
45. Fulton, L.; McLeod, A.; Dolezel, D.; Bastian, N.; Fulton, C.P. Deep Vision for Breast Cancer Classification and Segmentation. *Cancers* **2021**, *13*, 5384. [CrossRef]
46. Arias, R.; Narváez, F.; Franco, H. Evaluation of Learning Approaches Based on Convolutional Neural Networks for Mammogram Classification. In Proceedings of the Systems and Applications: First International Conference, SmartTech-IC 2019, Quito, Ecuador, 2–4 December 2019; Communications in Computer and Information Science. Springer: Cham, Switzerland, 2020; Volume 1154, pp. 273–287.
47. Jabeen, K.; Khan, M.A.; Balili, J.; Alhaisoni, M.; Almujally, N.A.; Alrashidi, H.; Tariq, U.; Cha, J.-H. BC2NetRF: Breast Cancer Classification from Mammogram Images Using Enhanced Deep Learning Features and Equilibrium-Jaya Controlled Regula Falsi-Based Features Selection. *Diagnostics* **2023**, *13*, 1238. [CrossRef]
48. Baccouche, A.; Garcia-Zapirain, B.; Elmaghraby, A.S. An integrated framework for breast mass classification and diagnosis using stacked ensemble of residual neural networks. *Sci. Rep.* **2022**, *12*, 12259. [CrossRef]
49. Muduli, D.; Dash, R.; Majhi, B. Automated diagnosis of breast cancer using multi-modal datasets: A deep convolution neural network based approach. *Biomed. Signal Process. Control.* **2022**, *71*, 102825. [CrossRef]
50. Bouzar-Benlabiod, L.; Harrar, K.; Yamoun, L.; Khodja, M.Y.; Akhloufi, M.A. A novel breast cancer detection architecture based on a CNN-CBR system for mammogram classification. *Comput. Biol. Med.* **2023**, *163*, 107133. [CrossRef]
51. Fatima, M.; Khan, M.A.; Shaheen, S.; Almujally, N.A.; Wang, S.H. B2C3NetF2: Breast cancer classification using an end-to-end deep learning feature fusion and satin bowerbird optimization controlled Newton Raphson feature selection. *CAAI Trans. Intell. Technol.* **2023**. online version of record before inclusion in an issue. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.