


Article

# Application of Deep Learning to Construct Breast Cancer Diagnosis Model

Rong-Ho Lin <sup>1</sup>, Benjamin Kofi Kujabi <sup>1,\*</sup> , Chun-Ling Chuang <sup>2</sup>, Ching-Shun Lin <sup>1</sup> and Chun-Jen Chiu <sup>1</sup>

<sup>1</sup> Department of Industrial Engineering and Management, National Taipei University of Technology, 1, Sec. 3, Zhongxiao E. Rd., Taipei 10608, Taiwan; rhlin@mail.ntut.edu.tw (R.-H.L.); sun8d@msn.com (C.-S.L.); t102379004@ntut.edu.tw (C.-J.C.)

<sup>2</sup> Department of Information Management, Kainan University, Taoyuan City 33857, Taiwan; clchuang@mail.knu.edu.tw

\* Correspondence: kofi\_ben3@hotmail.com; Tel.: +886-975841720

**Abstract:** (1) Background: According to Taiwan's ministry of health statistics, the rate of breast cancer in women is increasing annually. Each year, more than 10,000 women suffer from breast cancer, and over 2000 die of the disease. The mortality rate is annually increasing, but if breast cancer tumors are detected earlier, and appropriate treatment is provided immediately, the survival rate of patients will increase enormously. (2) Methods: This research aimed to develop a stepwise breast cancer model architecture to improve diagnostic accuracy and reduce the misdiagnosis rate of breast cancer. In the first stage, a breast cancer risk factor dataset was utilized. After pre-processing, Artificial Neural Network (ANN) and the support vector machine (SVM) were applied to the dataset to classify breast cancer tumors and compare their performances. The ANN achieved 76.6% classification accuracy, and the SVM using radial functions achieved the best classification accuracy of 91.6%. Therefore, SVM was utilized in the determination of results concerning the relevant breast cancer risk factors. In the second stage, we trained AlexNet, ResNet101, and InceptionV3 networks using transfer learning. The networks were studied using Adaptive Moment Estimation (ADAM) and Stochastic Gradient Descent with Momentum (SGDM) based optimization algorithm to diagnose benign and malignant tumors, and the results were evaluated; (3) Results: According to the results, AlexNet obtained 81.16%, ResNet101 85.51%, and InceptionV3 achieved a remarkable accuracy of 91.3%. The results of the three models were utilized in establishing a voting combination, and the soft-voting method was applied to average the prediction result for which a test accuracy of 94.20% was obtained; (4) Conclusions: Despite the small number of images in this study, the accuracy is higher compared to other literature. The proposed method has demonstrated the need for an additional productive tool in clinical settings when radiologists are evaluating mammography images of patients.



**Citation:** Lin, R.-H.; Kujabi, B.K.; Chuang, C.-L.; Lin, C.-S.; Chiu, C.-J. Application of Deep Learning to Construct Breast Cancer Diagnosis Model. *Appl. Sci.* **2022**, *12*, 1957. <https://doi.org/10.3390/app12041957>

Academic Editor: Giancarlo Mauri

Received: 4 January 2022

Accepted: 9 February 2022

Published: 13 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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/>).

**Keywords:** breast cancer; neural network; support vector machine; deep learning; convolutional neural network

## 1. Introduction

According to the 2020 statistics globally sourced by the World Health Organization, countries have reported that breast cancer is considered a dominant disease ranked the highest cause of death in women [1]. Taking Taiwan as an example, according to the Ministry of health's 2019 death statistics of women, the leading cause of death among all diseases such as diabetes, chronic respiratory disease, hypertension, chronic liver diseases, etc., breast cancer mortality rate was ranked fourth. Its mortality rate significantly increased from 12.8% (1439) in 2006 to 22.2% (2633) in 2019 [2]. Compared to the United States of America (USA), the USA projected 41,760 deaths in 2019 at a rate of 19.9 per 100,000 women per year, and more than 3.8 million have a history of breast cancer [3,4].

Despite the presence of an upsurge in the annual mortality rate, early detection of breast cancer increases the survival rate of patients if the appropriate treatment is provided

to avoid the need for a surgical procedure. R. Kate and R. Nadig (2016) [5] mentioned that physicians and healthcare workers may make more informed decisions regarding a patient's condition if breast cancer survivability can be accurately predicted. In the last decade, numerous data mining tools have been used to determine the factors affecting the survival of patients with breast cancer [6–12]. Due to the advancement of technology, many machine learning tools were used to predict and diagnose patients with breast cancer [13]. These data mining tools have assisted doctors in making accurate diagnoses. However, one of the most used machine learning methods for detecting or diagnosing diseases are classifiers [14]. It was noted by [15–17] that ANN and SVM are among the most commonly used supervised learning methods in the medical field for breast cancer diagnosis. Despite the fact that multiple classifiers were used to deal with medical-related classification problems. Ren (2012) [18] suggested that when the training samples are imbalanced, balanced learning with an optimized decision is employed to improve the performance of both ANN and SVM. Although, some previous works have highlighted significant gains in the area of deep learning to identify lesion in breast cancer [19]. Despite the small number of images in this study, we developed a model architecture by exploiting ANN, SVM, AlexNet [20], ResNet 101 [21], and InceptionV3 [22] networks to improve the diagnostic accuracy and reduce the misdiagnoses rate of patients with breast cancer.

## 2. Materials and Methods

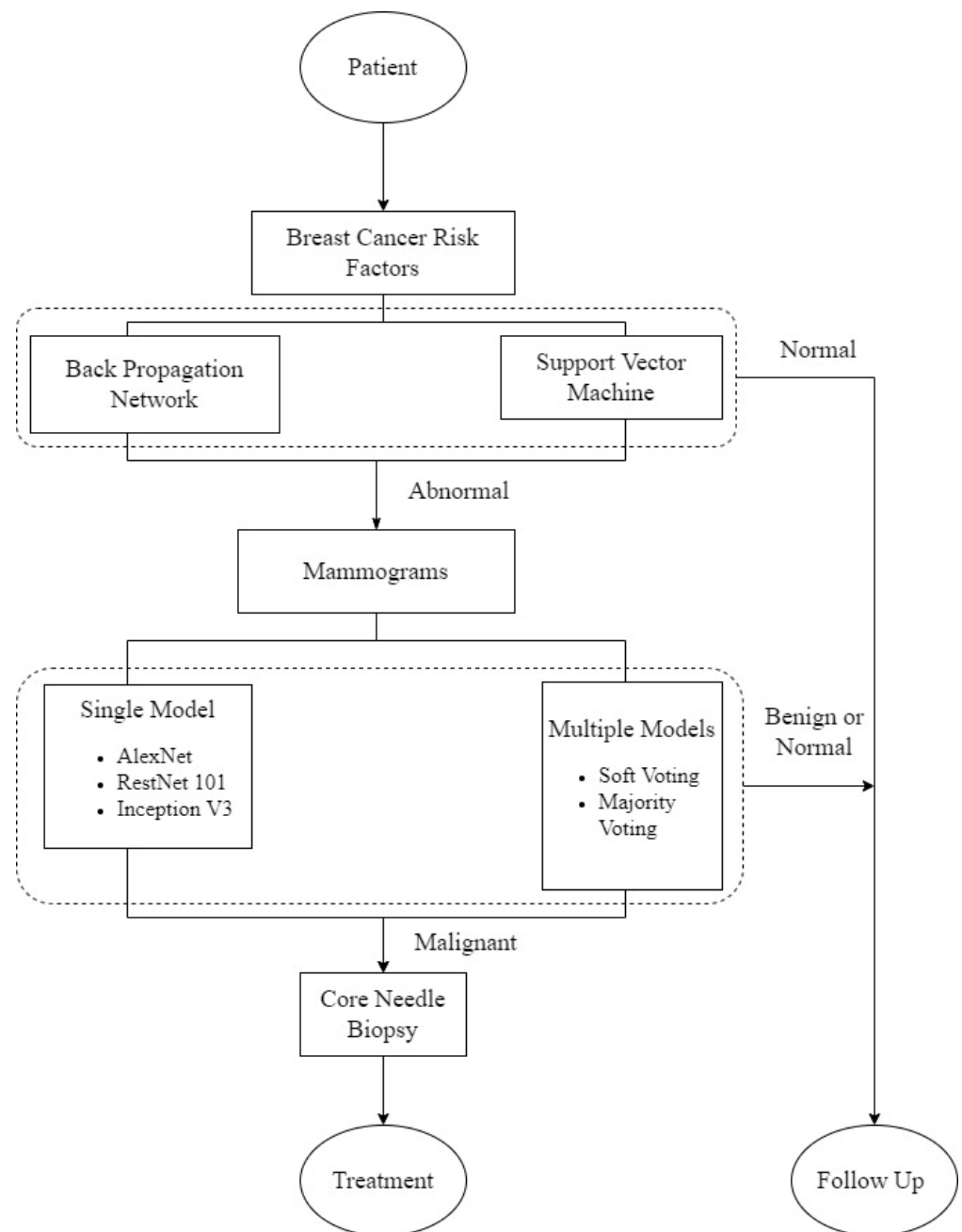
This study applied the use of machine learning methods in artificial intelligence to train and construct machine learning models for breast cancer diagnosis using data related to patients' breast cancer risk factors and mammograms. In the first stage, ANN and SVM were applied to the breast cancer dataset, and their performances were compared to determine the relevant breast cancer risk factors. In the second stage, an image recognition model was setup using pretrained AlexNet, ResNet101, Inception\_V3 fed with preprocessed mammograms. The networks were studied using Adaptive Moment Estimation (ADAM) and Stochastic Gradient Descent with Momentum (SGDM) based optimization algorithms to diagnose benign or malignant tumors, after which their accuracies were evaluated. Since the accuracy of breast cancer survivability is essential, the stepwise approach has significantly boosted the classification accuracy level. Furthermore, soft-voting was applied to average prediction results obtained from the pretrained networks.

### 2.1. Research Framework

This study focuses on using machine learning in artificial intelligence to build a system that can assist doctors in making diagnostic decisions. First, the patient's breast cancer risk factor data is inputted into the breast cancer diagnosis model, and the model is used to classify the patient's current status as normal or breast cancer, and when the result is normal, the patient is followed up regularly; when the result shows a high risk of breast cancer, the patient is further scheduled to undergo mammography for in-depth examination, and the images obtained from mammography are processed and inputted into the breast cancer confirmation model for suspected tumors. To determine whether the mammogram is a benign or malignant tumor, the mammogram model is used to provide an additional basis for the physician to make a judgment and allow the patient to receive the required treatment. The overall structure is shown in Figure 1.

#### 2.1.1. Classification

In this study, two classification methods, Backpropagation Neural Network (BPN) and Support Vector Machine (SVM) were used to classify the breast cancer risk factors of breast cancer patients.



**Figure 1.** Breast cancer risk factor model. (The dotted lines on the top are the first phase of the model where the BPN and SVM are applied to the dataset. The second dotted lines at the bottom are the second phase of the recognition model where the deep learning networks and the single and majority voting are applied).

### 2.1.2. Back Propagation Network (BPN)

#### Step1. Determining the network architecture

This includes variables that determine the input and output layers, select the number of hidden layers, neurons in the hidden layer, and the activation function.

1. Input layer: Generally, data that are about to enter the input layer need to be pre-processed to reduce possible prediction errors caused by different data units. After the data were preprocessed by normalization, the weight adjustment rate of the data should be similar to avoid weight dispersion. Additionally, the number of neural processing units depends on the problem, and the number of neurons in the input layer of this study is the number of attributes of the cancer data used.

2. Output layer: The output layer represents the output variables of the network, and the number of neural processing units depends on the problem. According to this research, the output was set to benign and malignant.
3. Hidden layer: The hidden layer represents the interaction between the neural processing units of the input layer. In this study, the hidden layer represents the interaction between the input and output layers. The computation of the hidden layers, the number of neurons, and the activation function were determined here.
  - (a) Number of hidden layers: A neural network with one hidden layer can approximate the most complex functions with the required accuracy [23]. Therefore, this study sets the hidden layers to one.
  - (b) The number of neurons in the hidden layer: if the number of neurons in the hidden layer is too small, it will not be enough to construct a nonlinear relationship between the output and the input, leading to an error, but overfitting may occur if it exceeds a certain number. Studies have pointed out that the number of neurons in the hidden layer can be set using either Equation (1) or (2).

$$\text{number of neurons in the hidden layer} = \frac{\text{number of neurons in the input layer} + \text{number of neurons in the output layer}}{2} \quad (1)$$

$$\text{number of neurons in the hidden layer} = \sqrt{\text{number of neurons in the input layer} \times \text{number of neurons in the output layer}} \quad (2)$$

- (c) Activation Function: the primary function of the activation function is to convert the output value of the function into the output of the processing unit. In this study, a sigmoid function, as shown in Equation (3) was used. This function converts the output value to a value between zero and one.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

#### Step 2. Finding the best parameter combination

1. Learning Rate and momentum: the learning rate is mainly used to control the magnitude of the weight each time it changes. If it is too large or too small, it may negatively impact the network in ways such as (1), causing the model to converge too quickly to a suboptimal solution or (2) causing the process to jam. Therefore, in this study, we set a large initial value (e.g., 0.01) for the learning rate, and then gradually reduce it during the training process, to strike a balance between speeding up the convergence and avoiding the oscillation. This concept is also known as decaying learning rate as shown in Equation (4).

$$\alpha_{current} = \frac{\alpha_{initial}}{1 + (r_{decay} \times n_{epoch})} \quad (4)$$

where  $\alpha_{current}$  is the learning rate at the current stage,  $\alpha_{initial}$  is the initial learning rate,  $r_{decay}$  is the decay rate,  $n_{epoch}$  is the current iteration number, In this study,  $\alpha_{initial}$  was set to 0.01,  $r_{decay}$  is set to 0.9 to prevent the learning rate from being close to 0 due to too long iterations, and the minimum learning rate was set to  $1 \times 10^{-8}$ .

2. Convergence conditions: to find a stable and predictive network architecture, certain evaluation indicators should set criteria for network architecture selection. Since this study is a classification problem, the classification accuracy rate and Mean Square Error (MSE) are used as an index to evaluate the performance of the network prediction ability. The accuracy rate and MSE are shown in Equations (5) and (6), respectively.

$$\text{Accuracy Rate} = \frac{\text{number of sample classified correctly}}{N} \times 100\% \quad (5)$$

$$\text{MSE} = \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N} \quad (6)$$

where  $N$  is the total number of samples,  $\hat{y}_i$  is the predicted value of the  $i$ th sample, and  $y_i$  is the actual value of the  $i$ th sample. In this study, the convergence criteria were based on the highest Accuracy Rate for selecting the best network architecture and parameter combination was the first priority and the test sample with the smallest MSE value was the second priority.

3. Support Vector Machine (SVM): is a machine learning method published by Cortes and Vapnik in 1995 [24]. SVM has been widely used in recent years to solve various classification problems [25], [26]. By calculating the training data, SVM can find an optimal hyperplane and classification decision function to effectively separate the data points belonging to two categories, and when a new case is predicted for classification, the hyperplane can be used to determine the category to which the case belongs [27].

If you want to use SVM to deal with nonlinear problems, you can use the kernel function to map the data group, mapping the observation point to a higher-dimensional feature space, make it a linear hyperplane, and then find its solution; in other words, the kernel function converts nonlinear data into linear data, and the classifier performs the classification work. The definition of the core function is shown in Equation (7).

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \quad (7)$$

where  $\varphi$  is the mapping function,  $x$  can be mapped to a higher-dimensional feature space through  $\varphi$ .

The four commonly used core functions are Linear, Polynomial, Radial Basis Function (RBF) and Sigmoid, as shown in Equation (8) to (11), respectively.

$$Lin\_K(x_i, x_j) = x_i^T x_j \quad (8)$$

$$Poly\_K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0 \quad (9)$$

$$RBF\_K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (10)$$

$$Sig\_K(x_i, x_j) = \tanh(x_i^T x_j + r) \quad (11)$$

In the above equation,  $\gamma$ ,  $r$  and  $d$  are the core parameters. Different core functions are matched with different core parameters to have different classification effects. Therefore, the selection of the core functions and parameters are important. According to Hsu et al. (2010) [28], the selection of core functions should give priority to Radial Basis Function (RBF) because of its advantages:

1. Radial Basis Function can classify non-linear and high-dimensional data.
2. By simply adjusting the Upper Bound parameter  $C$  and the core parameter  $\gamma$ , the operation is less complex and achieves better prediction capability.
3. The input data is limited between 0 and 1 to reduce the complexity of the calculation time.
4. Due to the high dimensional nature of the data used in this study, the RBF was used to find the optimal  $C$  and  $\gamma$ , and input the test data to evaluate the prediction accuracy rate (Accuracy Rate) of the model.

## 2.2. Image Recognition

In this study, a Convolutional Neural Network (CNN) was used for image recognition of mammograms to determine the stage of breast cancer in patients. The functions of the Convolutional Layer, Pooling Layer, and Fully Connected Layer of the CNN and the parameters required to construct a CNN are described below.

### 2.2.1. Convolutional Neural Network (ConvNet)

CNNs are types of biologically-inspired feed-forward networks characterized by a sparse connectivity and weight sharing among their neurons [29]. It accepts two-dimensional input data in contrast with other DL algorithms [30]. A CNN can also be referred to as a sequence of convolution and subsampling layers where the entire network will take an image input of size  $(h,w,c)$ , where  $H$  is height,  $W$  weight, and  $C$  is the number of channels in the image. Those channels are mostly referred to different (RGB) colors [31] and output the conditional probability distribution over the categories  $p(y|x)$ . This is carried out by a sequence of the nonlinear level image [32]. For each pixel in an image, the kernel multiplies the pixel and its adjacent pixels that the kernel covers by the opposite kernel pixels. The products are then totalled and their results are set as the pixel value in the convolved image at the preliminary pixel location [33].

The CNN architecture consists of three layers. Two consecutive conventional pooling layers and a final fully-connected layer [34]

The convolutional layer: is the main layer that forms a ConvNet and is also the more computationally intensive layer. Its main function is to extract features from the input image pixels. The first few layers of ConvNet can extract features at lower levels. As the network progresses to a deeper level, the features that can be extracted from the convolutional layers gradually increases. The calculation is based on the Element-wise multiplication of Input and Filter and then summed up. The so-called Filter (also known as Kernel) can be regarded as a window composed of several weights, by which the window is padded on the image, and each pixel value in the area covered by the window during the padding is multiplied with the weight of the window at its corresponding position, resulting in a convolution, hence, the name convolutional layer.

### 2.2.2. Pooling Layer or Downsampling

In most ConvNet, after the computation of the convolutional layer, the Input usually enters the pooling layer. The main purpose of the pooling layer is to reduce the spatial dimension of the Feature Map (Resolution) [35]. Downsampling is conducted along the width and height of the image to reduce the computational requirements progressively through the network and minimize overfitting.

The role of pooling is the process of reducing the image size by padding the image through the Filter window after the convolution is completed. The image size was reduced by extracting a specific pixel value (maximum or average) from the Filter window at each transition. In addition, when the image size is reduced, it also means that the number of parameters to be calculated reduces. As the complexity of the parameters decreases, the computation time reduces.

Fully connected layers: In a typical ConvNet architecture, besides the convolutional layer and the pooling layer, the last layer of ConvNet is the Fully Connected layer. When Input starts from the beginning of ConvNet, it passes through several layers of convolutional layer and pooling layer, and then passes through the last pooling layer (provided that there is a pooling layer), when it is about to enter the fully connected layer, The neurons in the previous pooling layer will connect to the feature values, which have already been activated and the structure becomes a common neural network.

Transfer Learning: Pretrain network architectures on an extensive large dataset and uses the trained model on a dataset with minimal size on a new classification task [36] The application of transfer learning to ConvNet is to extract the features of pictures using the weights in the pretrained model and using the extracted features for classification.

### 2.2.3. Stochastic Gradient Descent

The common algorithms used to calculate the gradient in training networks are Stochastic Gradient Descent with Momentum (SGDM) and Adaptive Moment Estimation (ADAM) [37,38]. In this study, we used different optimizers, including SGDM and Adaptive Moment Estimation (ADAM), to train the network and later compared the performance of

the two. The SGDM algorithm is shown in Equation (12), while the ADAM algorithm uses Equations (13) and (14) and updates the network parameters with Equation (15).

$$P_{l+1} = P_l - \alpha \nabla E(P_l) + \gamma(P_l - P_{l-1}) \quad (12)$$

$$m_l = \beta_1 m_{l-1} + (1 - \beta_1) \nabla E(P_l) \quad (13)$$

$$v_l = \beta_2 v_{l-1} + (1 - \beta_2) [\nabla E(P_l)]^2 \quad (14)$$

$$P_{l+1} = P_l - \frac{\alpha m_l}{\sqrt{v_l} + \epsilon} \quad (15)$$

When there is enough data for training a model, a new model can be built and trained from scratch, but when there is less data, the problem of overfitting can easily occur, and then transfer learning can be used to overcome overfitting.

#### 2.2.4. Soft Voting

After averaging the output probabilities of breast cancer obtained from each model with Equation (16) [39], the Equation (17) was used to obtain the final classification probability result. Benign or normal cases of patients with breast cancer are represented as B or N (benign means a lump is present in the breast and normal means no breast lump is present), malignant is represented as M, and the classification probability is represented as [B or N probability, M probability].

$$Output_i = \frac{1}{n} \sum_{j=1}^n net_j(i) \quad (16)$$

where  $Output_i$  is the output of the voting combination model,  $n$  is the number of CovNet used for the voting model and  $net_j(i)$  is the output  $i$  of the  $j$ th ConvNet.

$$y_{predict} = \operatorname{argmax}[p(i_0 | x), p(i_1 | x)] \quad (17)$$

In Equation (6), where  $y_{predict}$  is the category predicted using the combined voting model;  $i_0$  represents category B or N;  $i_1$  represents category M, that is, if the classification probability is greater than category B or N, the output is category B or N; if the classification probability is greater than category M, then the output is category M.

#### 2.2.5. Confusion Matrix

The model performance was evaluated through a standard data classification system based on accuracy, sensitivity, and specificity. True Positive (TP) and true negative (TN) results represent correctly classified cases. A test's Accuracy is computed by estimating the fraction of true positive and negative instances in all cases as computed in Equation (18). Sensitivity, are correctly generated positive cases with either cancer or cancer free (also known as TP rate) as in Equation (19). Specificity, correctly generated negative cases of those without cancer or cancer-free (also known as the TN rate) as in Equation (20) [40].

$$\text{Accuracy} : \frac{TP + TN}{TP + FN + FP + TN} \times 100 \quad (18)$$

$$\text{Sensitivity} : \frac{TP}{TP + FN} \quad (19)$$

$$\text{Specificity} : \frac{TN}{TN + FP} \quad (20)$$

### 2.2.6. Data Description

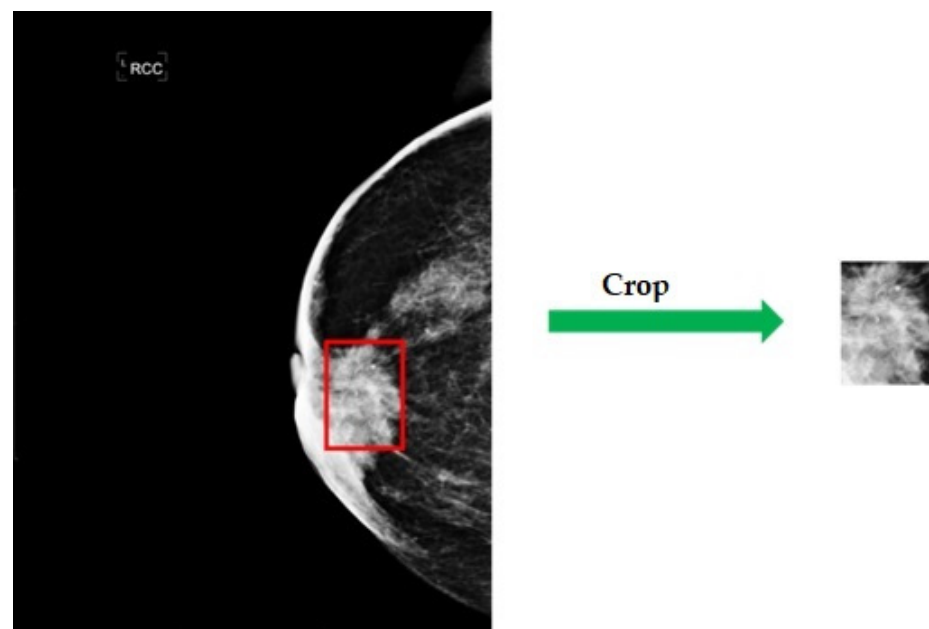
This study uses data obtained from the Breast Cancer Surveillance Consortium (BCSC), Data Resource [41,42]. The breast cancer risk factor assessment dataset is used in constructing the first stage of the breast cancer diagnosis model.

The breast cancer risk factor assessment dataset contains 2,392,998 cases with 12 attributes namely: menopause, age group, breast density, ethnicity, Hispanic origin, BMI value, age, number of relatives with breast cancer, previous breast-related surgery, last mammogram result, menopause mode, hormonal treatment or not, and response variable Class. The response variable (Class) was to evaluate whether the patient had invasive breast cancer or noninvasive breast cancer (Ductal Carcinoma in Situ). After removing the cases with missing values in the data, it was found that the number of cases in the category was imbalanced, so the Synthetic Minority Oversampling Technique (SMOTE) [43–45] Duplicate method which is considered the “de facto” standard able to learn from imbalance data was used to increase the categories with fewer cases, and after processing the data a total of 88,763 cases were used to train and classify the model.

### 2.2.7. Data Pre-Processing

Before inputting the mammogram image into ConvNet for training, it is necessary to preprocess the mammogram images, and the following steps were applied.

- (1) Cropping—The original image was cropped to retain only the main Region of Interest (ROI), i.e., the tumor area was cropped out. An example is depicted in Figure 2.
- (2) Rotation—The image was rotated at a random angle within a specific range.
- (3) Random vertical and horizontal image flip.
- (4) Vertical and horizontal image rotation (Shift).
- (5) Randomly zoom in or out in a specific area.
- (6) Vertical and horizontal image mirroring (Reflection).



**Figure 2.** Schematic diagram of cropped ROI from a mammogram.

### 2.2.8. Input Layer

The input size of all images was then scaled to the required input size for each Convolutional Neural Network,  $227 \times 227 \times 3$  for AlexNet,  $224 \times 224 \times 3$  for ResNet101, and  $299 \times 299 \times 3$  for Inception v3, where three is the number of color channels. This means that the input images of these ConvNets are all RGB images (color images). The



last fully linked layer of AlexNet, ResNet101, and Inception v3 was removed and replaced with a new spreading layer and a Softmax layer, and the number of output neurons were changed from 1000 to 2 (benign or normal (B or N) and malignant (M)).

### 2.2.9. Model Development

In this study, MATLAB R2018a software [46] was used in building the model. Since the weights of a specific number of layers need to be fixed when conducting transfer learning, the learned weights were used to extract the features of the image to reduce the probability of overfitting, but at the same time allowing deeper convolutional layers to conduct higher-level feature extraction of the image may also improve the classification accuracy of the model. Therefore, we let the deeper convolutional layers of ResNet101 and Inception v3 with deeper network depth learn the images (i.e., the weights of the deeper layers were not fixed) to compare and find the best fixed-weight layers between the two models. Since the network depth (number of layers) of AlexNet was not as deep as that of ResNet101 and Inception v3, the number of layers with fixed weights was not explored here.

### 2.2.10. Back Propagation Network (BPN)

The number of neurons in the input layer of BPN is the number of attributes in the BCSC breast cancer risk factor data set (there were 12 neurons in the input layer), and the number of neurons in the output layer is used to determine the presence of invasive breast cancer or noninvasive breast cancer ( $N = 1$ ). Based on this, to find the best network architecture for BPN, we set the hidden layer of the network to one and determined the number of neurons in the hidden layer using the Equations (21) and (22).

$$\text{Number of neurons in the hidden layer} = \frac{12 + 1}{2} = 6.5 \quad (21)$$

$$\text{Number of neurons in the hidden layer} = \sqrt{12 \times 1} = 3.46 \quad (22)$$

According to the above equations, the number of neurons in the hidden layer was set to two, three, four, five, six, seven, and tested. The sigmoid function was used for the activation function, and the results were averaged over three times for each number of neurons (Table 1).

**Table 1.** Selection of the number of neurons in the hidden layer of 1BPN.

Number of Neurons in the Hidden Layer	First Result	Second Result	Third Result	Average
2	73.1%	74.0%	73.6%	73.6%
3	74.4%	73.4%	74.1%	73.9%
4	74.6%	74.0%	75.2%	74.6%
5	74.3%	74.8%	74.8%	74.6%
6	76.6%	76.4%	76.7%	<b>76.6%</b>
7	76.1%	74.9%	76.5%	75.8%

Bold text indicates the highest accuracy rate.

The BCSC breast cancer risk factor dataset was entered into the BPN, and the results were obtained by training the architecture with six hidden layer neurons, 31,901 cases with cancer were correctly classified, while 10,212 cases were classified as not having cancer. Among those without cancer, 12,291 cases were correctly classified, and 34,339 cases were classified as having cancer.

The experiment performed on our breast cancer risk factors model accumulated the results of the various evaluation metrics. Tables 2 and 3 depicts the accuracy, sensitivity, and specificity for the BPN and SVM. The accuracies of the BPN and SVM are 74.63% and 91.6%, respectively, showing a march between the predicted and the actual instances.

**Table 2.** BPN confusion matrix.

Prediction Category	Genuine Category		Total	Accuracy	Sensitivity	Specificity
	Cancer	Cancer-Free				
Cancer	31,901	10,212	42,113	74.63%	24.24%	75.75%
Cancer-free	12,291	34,339	46,630	25.35%	73.64%	26.36%

**Table 3.** SVM confusion matrix.

Prediction Category	Genuine Category		Total	Accuracy	Sensitivity	Specificity
	Cancer	Cancer-Free				
Cancer	42,092	21	42,113	91.6%	0.05%	99.95%
Cancer-free	7437	39213	46,650	8.4%	84.06%	15.94%

The sensitivity and specificity are inversely proportional, meaning as the sensitivity increases, the specificity decreases and vice versa [47]. For instance, the BPN sensitivity is 24.24% while the specificity is 75.75%. Under cancer-free the sensitivity is 73.64% while the specificity is 26.36%.

#### 2.2.11. Support Vector Machine (SVM)

The core function used here is the Radial Basis Function (RBF) SVM with outputs zero (normal) and one (breast cancer), and the algorithm used to train this SVM is the Sequential Minimal Optimization. The SVM was then cross-validated 10 times to obtain a loss of 0.0842, i.e., a classification accuracy of 91.6% and an AUC of 0.96. According to the classification results, 42,092 cases were correctly classified, and 21 cases were not classified as having cancer. Among those without cancer, 39,213 cases were correctly classified, and 7437 cases were classified as having cancer (Table 3).

From the above experimental results, we found that both SVM and BPN were excellent in correctly classifying cancer patients. The performance of the support vector machine was better than that of the backpropagation neural network in terms of classification accuracy (91.6%) and AUC (0.96).

#### 2.2.12. Breast Cancer Validation Model

Due to personal information attached to the mammograms; data collection was difficult. As a result, the amount of data obtained was relatively small, and when the data is small, training the network from scratch can easily cause overfitting and affect the generalization of the model. Therefore, transfer learning will be ideal for training the image recognition model.

Since the number of mammograms in this study was very minimal, among which the benign tumor images are the least, the diagnostic results of benign and normal were combined into one category (B or N), and malignant (M) into another category, this reduces the occurrence of overfitting.

### 3. Results

#### 1. AlexNet–Optimizer: Adaptive Moment Estimation (ADAM)

The classification accuracies obtained from the training and testing on three occasions were 79.71%, 81.16%, and 81.16%, respectively, with an average classification accuracy of 80.68% (Table 4).

**Table 4.** AlexNet classification accuracy (ADAM).

First Accuracy Rate	Second Accuracy Rate	Third Accuracy Rate	Average
79.71%	81.16%	81.16%	<b>80.68%</b>

Bold text indicates highest accuracy rate.

## 2. AlexNet–Optimizer: Stochastic Gradient Descent with Momentum (SGDM)

The classification accuracies obtained from the training and testing on three occasions were 81.16%, 85.51%, and 84.06%, and the average classification accuracy was 83.58% (Table 5).

**Table 5.** AlexNet classification accuracy (SGDM).

First Accuracy Rate	Second Accuracy Rate	Third Accuracy Rate	Average
81.16%	85.51%	84.06%	<b>83.58%</b>

Bold text indicates highest accuracy rate.

Comparing the average accuracy in Table 4 with that of Table 5, the classification accuracy obtained by AlexNet using SGDM as the optimizer during training was better than that obtained using ADAM. The results were 24 of the benign or normal cases were correctly classified and six were misclassified as malignant; 32 of the malignant cases were correctly classified and seven were misclassified as benign or normal (Table 6).

**Table 6.** AlexNet classification results.

Prediction Category	Genuine Category	
	B or N	M
B or N	24	7
M	6	32

## 3. ResNet101–Optimize: Adaptive Moment Estimation (ADAM)

From the experimental results, it can be found that when using the ADAM optimizer, the average classification accuracy of 81.16% was obtained after fixing the weights of each layer in front of module 5c, 83.09% was obtained after fixing the weights of each layer in front of module 5b, and 81.16% was obtained after fixing the weights of each layer in front of module 5a (Table 7).

**Table 7.** The classification accuracy of different fixed layers.

Fixed-Weight Layer	Classification Accuracy 1	Classification Accuracy 2	Classification Accuracy 3	Average
Before fixing module 5c (1-333)	81.16%	79.71%	82.61%	81.16%
Before fixing module 5b (1-323)	82.61%	84.06%	82.61%	<b>83.09%</b>
Before fixing module 5a (1-311)	79.71%	82.61%	81.16%	81.16%

Bold text indicates highest accuracy rate.

It can be seen from Table 7 that not allowing all the deeper convolutional layers of the model to perform higher-level feature extraction on the image will improve the classification accuracy. From the perspective of ResNet101, when using the ADAM optimizer, fixing the weights of the first to the 323rd layer can make the model develop a higher classification accuracy.

4. ResNet101–Optimizer: Stochastic Gradient Descent with Momentum (SGDM)

From the experimental results, it was observed that the average classification accuracy of 79.71% was obtained after fixing the weights of each layer in front of module 5c, 82.61% was obtained after fixing the weights of each layer in front of module 5b, and 79.71% was obtained after fixing the weights of each layer in front of module 5a (Table 8).

**Table 8.** Accuracy rate of each classification with fixed number of layers (SGDM).

Fixed-Weight Layer	Classification Accuracy 1	Classification Accuracy 2	Classification Accuracy 3	Average
Before fixing module 5c (1-333)	79.71%	81.16%	78.26%	79.71%
Before fixing module 5b (1-323)	82.61%	84.06%	81.16%	<b>82.61%</b>
Before fixing module 5a (1-311)	81.16%	78.26%	79.71%	79.71%

Bold text indicates highest accuracy rate.

From Table 8, it can be observed that when ResNet101 uses SGDM as the optimizer, fixing the weights of layer one to layer 323 still results in a better average classification accuracy rate of 82.61%. Although, the performance is not as good as when using ADAM (83.09%). Generally, the classification accuracy rate when using ADAM optimizer is still slightly better than when using SGDM. After applying the data to ResNet101 with ADAM as the optimizer, the classification accuracy of 85.51% was obtained, and the results were; 25 of the benign or normal cases were correctly classified and five were misclassified as malignant; 34 of the malignant cases were correctly classified and five were misclassified as benign or normal (Table 9).

**Table 9.** ResNet101 classification results.

Prediction Category	Genuine Category	
	B or N	M
B or N	25	5
M	5	34

5. Inception v3–Optimizer: Adaptive Moment Estimation (ADAM)

From the experimental results, it can be observed that when using ADAM as the optimizer, the average classification accuracy of 85.51% was obtained after fixing the weight of each layer before the merge point mixed10. The average classification accuracy of 87.44% was obtained after setting the weights of each layer before the merge point mixed9; the average classification accuracy of 90.71% was obtained after setting the weights of each layer before the merge point mixed8 (Table 10).

**Table 10.** Fixed classification accuracy of different layers (ADAM).

Fixed-Weight Layer	Classification Accuracy 1	Classification Accuracy 2	Classification Accuracy 3	Average Accuracy
Before fixing mixed10 (1-281)	85.51%	86.96%	84.06%	85.51%
Before fixing mixed9 (1-250)	88.41%	86.96%	86.96%	87.44%
Before fixing mixed8 (1-230)	89.86%	91.30%	89.86%	<b>90.71%</b>
Before fixing mixed7 (1-198)	82.61%	84.06%	81.16%	82.61%

Bold text indicates highest accuracy rate.

From Table 10, it can be observed that the classification accuracy of Inception v3 increases as the deeper layers are allowed to conduct higher-level feature extraction on the images. For this reason, the number of layers with fixed weights were reduced to test whether reducing the number of layers with fixed weights could increase the classification accuracy again (i.e., fixed weights from layer one to layer 198).

From Table 10 the results showed that the classification accuracies obtained through three intervals of training and testing were 82.61%, 84.06%, and 81.16%, with an average accuracy of 82.61%; in other words, higher classification accuracy was obtained when the weights of Inception v3 were fixed from layer one to layer 230.

#### 6. Inception v3–Optimizer: Stochastic Gradient Descent with Momentum (SGDM)

From the experimental results, it can be observed that when using SGDM as an optimizer, the average classification accuracy of 83.09% was obtained after fixing the weight of each layer before the merge point mixed10. The average classification accuracy of 81.64% was obtained after fixing the weights of each layer before mixed9; 83.58% was obtained after fixing the weights of each layer before mixed8 (Table 11).

**Table 11.** Fixed classification accuracy of different layers (SGDM).

Fixed-Weight Layer	Classification Accuracy 1	Classification Accuracy 2	Classification Accuracy 3	Average Accuracy
Before fixing mixed10 (1-281)	84.06%	82.61%	82.61%	83.09%
Before fixing mixed9 (1-250)	81.16%	79.71%	84.06%	81.64%
Before fixing mixed8 (1-230)	84.06%	81.16%	85.51%	<b>83.58%</b>

Bold text indicates highest accuracy rate.

As shown in Table 11, the highest classification accuracy is still obtained when using the SGDM optimizer with fixed weights from the first layer to the 230th layer, but in contrast with when using the ADAM optimizer. The second-highest classification accuracy was obtained using SGDM optimizer with fixed weights from the first layer to the 281st layer instead of the fixed weights from the first layer to the 250th layer. It was also observed that the overall performance of Inception v3 using ADAM optimizer was better than the performance using SGDM.

Since the best classification accuracy was obtained using Inception v3 with ADAM optimizer and transfer learning with fixed weights from the first layer to the 230th layer, the test data set was integrated into this completed model and 91.3% classification accuracy was obtained. Twenty-eight (28) of the cases (benign or normal cases) were correctly classified, and two cases were incorrectly classified as malignant, 35 of the malignant cases were correctly classified, and four were incorrectly classified as benign or normal (Table 12).

**Table 12.** Inception v3 classification results.

Predicted Category	Genuine Category	
	B or N	M
B or N	28	4
M	2	35

Generally, from the classification results of AlexNet, ResNet101, and Inception v3, Inception v3 had the best performance, while the classification accuracy of ResNet101 was higher than that of the AlexNet. It can also be observed that as the network deepens, the classification accuracy increases. The results obtained from the training and testing of every single model are summarized in Tables 13–15.

**Table 13.** Accuracy rates of each single model.

	AlexNet	ResNet101	Inception v3
Training	83.58%	83.09%	90.71%
Testing	81.16%	85.51%	91.30%

#### 7. Soft-voting model

**Table 14.** Soft-voting model classification results.

Predicted Category	Genuine Category	
	B or N	M
B or N	29	3
M	1	36

#### 8. Majority voting

**Table 15.** Majority voting model classification results.

Predicted Category	Genuine Category	
	B or N	M
B or N	27	4
M	3	35

The accuracy of the soft-voting model is 94.20%, which is about 2.9% higher than the accuracy of the single model Inception v3. Table 14 shows the classification results of the soft-voting model, and Table 15 shows the classification results of the majority voting. It can be observed that applying soft-voting can successfully reduce the cases of misclassification and improve the classification accuracy, and this led to a significant classification accuracy rate of 89.85%.

## 4. Discussion

The proposed model demonstrates that it is capable to Improve diagnostic accuracy and reduce the misdiagnosis rate of breast cancer. The results showed that when the three networks were compared using ADAM and SGDM, the InceptionV3 achieved the highest accuracy 91.30% when compared to [48]. This was due to the deep network of the InceptionV3 after being fine-tuned. Although, the AlexNet is capable of achieving excellent results on highly challenging datasets using purely supervised learning but if a single convolution layer is removed the network's performance degrades [49]. In relation to this, it was observed when we fix module 5a layers in Table 7 and it degraded the performance of the network, resulting to an accuracy of 81.16%.

In comparing ADAM and SGDM, ADAM outperformed SGDM since its adaptation of learning rate scale for different layers instead of hand-picking manually in SGDM [50].

Regarding SVM and ANN, the SVM outperformed the ANN and this is attributed to the ability of SVM handling large feature space, avoiding overfitting and condensing of information for a given dataset [51]. In this regard, the SVM results have demonstrated a highly classification accuracy of 91.60%.

Although, our soft voting model was able to correct misclassified data using a single model. We are cognizant of the fact that there was a small proportion of the data used with limited computational resources which has hindered our efforts to perfectly fine-tune the networks. Thus, in future research, we would consider employing a large dataset and carryout more exhaustive tests to optimize the performance of the deep learning networks and test other algorithms such as AdaBelief [52] optimizer which converges fast and has high accuracy on image classification and language modeling.

## 5. Conclusions

This research was aimed at developing a stepwise breast cancer model architecture to improve diagnostic accuracy and reduce the misdiagnosis rate of breast cancer. In the first stage, a breast cancer risk factor dataset was used. In the second stage, an image recognition model was set up using pretrained AlexNet, ResNet101, Inception\_V3 fed with preprocessed mammograms. The networks were studied using Adaptive Moment Estimation (ADAM), and SGDM based optimization algorithms to diagnose benign or malignant tumors, and their accuracies were evaluated. Since the accuracy of breast cancer survivability is essential, the stepwise approach has significantly boosted the classification accuracy level. It was observed that using a single model may misclassify a patient with benign or normal tumor as malignant; or misclassify a patient with malignant tumor as benign or normal, resulting in a missed opportunity to receive appropriate treatment. However, using multiple ConvNets voting models, soft voting can classify several cases that were originally misclassified using a single model to the correct category. This allows patients to have more time to receive proper treatment.

**Author Contributions:** Conceptualization, R.-H.L. and C.-J.C.; methodology, C.-J.C., B.K.K. and C.-S.L.; software, C.-L.C.; validation, R.-H.L.; writing—original draft preparation, B.K.K.; writing—review and editing, B.K.K. and R.-H.L.; visualization, B.K.K., C.-J.C. and C.-L.C.; supervision, R.-H.L., C.-S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Technology Development Project Foundation of Taiwan and National Taipei University of Technology and Chang Gung Memorial Hospital Joint Research Program.

**Institutional Review Board Statement:** Ethical review and approval were waived for this study due to the fact that, the data was obtained from the SEER website.

**Informed Consent Statement:** Ethical review and approval were waived for this study due to the fact that, the data was obtained from the SEER website.

**Data Availability Statement:** We thank the BCSC participants, investigators, mammography facilities, and radiologists for the data they have provided for this study. You can learn more about the BCSC at: <http://www.bsc-research.org/> (accessed on 3 January 2022).

**Acknowledgments:** We wish to acknowledge the support from the Ministry of Science and Technology Development Project Foundation of Taiwan (MOST 107-2221-E-027-072-MY2) and the National Taipei University of Technology and Chang Gung Memorial Hospital Joint Research Program (NTUT-CGMH-107-02).

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

## References

1. Breast Cancer. Available online: <https://www.who.int/news-room/fact-sheets/detail/breast-cancer> (accessed on 15 August 2021).
2. Cause of Death Statistics. 2017. Available online: <https://www.mohw.gov.tw/np-128-2.html> (accessed on 15 August 2021).
3. Cancer Facts & Figures 2021 | American Cancer Society. Available online: <https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2021.html> (accessed on 8 July 2021).
4. Cancer of the Breast (Female)—Cancer Stat Facts. Available online: <https://seer.cancer.gov/statfacts/html/breast.html> (accessed on 3 January 2022).
5. Kate, R.; Nadig, R. Stage-Specific Predictive Models for Breast Cancer Survivability. *Int. J. Med. Inf.* **2016**, *97*, 304–311. [[CrossRef](#)]
6. Foster, K.R.; Koprowski, R.; Skufca, J.D. Machine learning, medical diagnosis, and biomedical engineering research—commentary. *Biomed. Eng. Online* **2014**, *13*, 94. [[CrossRef](#)] [[PubMed](#)]
7. Kourou, K.; Exarchos, T.P.; Exarchos, K.P.; Karamouzis, M.V.; Fotiadis, D.I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 8–17. [[CrossRef](#)]
8. Al-Azzam, N.; Shatnawi, I. Comparing supervised and semi-supervised Machine Learning Models on Diagnosing Breast Cancer. *Ann. Med. Surg.* **2021**, *62*, 53–64. [[CrossRef](#)] [[PubMed](#)]

9. Roberto Cesar, M.-O.; German, L.-B.; Paola Patricia, A.-C.; Eugenia, A.-R.; Elisa Clementina, O.-M.; Jose, C.-O.; Marlon Alberto, P.-M.; Fabio Enrique, M.-P.; Margarita, R.-V. Method Based on Data Mining Techniques for Breast Cancer Recurrence Analysis. In *Advances in Swarm Intelligence*; Tan, Y., Shi, Y., Tuba, M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 584–596. [CrossRef]
10. Mining, D. Application of Data Mining Techniques to Predict Breast Cancer. *Procedia Comput. Sci.* **2019**, *163*, 11–18. [CrossRef]
11. Singh, S.N.; Thakral, S. Using Data Mining Tools for Breast Cancer Prediction and Analysis. In Proceedings of the 2018 4th International Conference on Computing Communication and Automation (ICCCA), Greater Noida, India, 14–15 December 2018; pp. 1–4. [CrossRef]
12. Shah, C.; Jivani, A.G. Comparison of data mining classification algorithms for breast cancer prediction. In Proceedings of the 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), Tiruchengode, India, 4–6 July 2013; pp. 1–4. [CrossRef]
13. Jonsdottir, T.; Hvannberg, E.T.; Sigurdsson, H.; Sigurdsson, S. The feasibility of constructing a Predictive Outcome Model for breast cancer using the tools of data mining. *Expert Syst. Appl.* **2008**, *34*, 108–118. [CrossRef]
14. Oskouei, R.J.; Kor, N.M.; Maleki, S.A. Data mining and medical world: Breast cancers' diagnosis, treatment, prognosis and challenges. *Am. J. Cancer Res.* **2017**, *7*, 610–627. [PubMed]
15. Delen, D.; Walker, G.; Kadam, A. Predicting breast cancer survivability: A comparison of three data mining methods. *Artif. Intell. Med.* **2005**, *34*, 113–127. [CrossRef] [PubMed]
16. Park, K.; Ali, A.; Kim, D.; An, Y.; Kim, M.; Shin, H. Robust predictive model for evaluating breast cancer survivability. *Eng. Appl. Artif. Intell.* **2013**, *26*, 2194–2205. [CrossRef]
17. Xu, X.; Zhang, Y.; Zou, L.; Wang, M.; Li, A. A gene signature for breast cancer prognosis using support vector machine. In Proceedings of the 2012 5th International Conference on BioMedical Engineering and Informatics, Chongqing, China, 16–18 October 2012; pp. 928–931. [CrossRef]
18. Ren, J. ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging. *Knowl.-Based Syst.* **2012**, *26*, 144–153. [CrossRef]
19. Chen, H.; Dou, Q.; Wang, X.; Qin, J.; Heng, P.-A. Mitosis detection in breast cancer histology images via deep cascaded networks. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; AAAI Press: Phoenix, AZ, USA, 2016; pp. 1160–1166.
20. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
22. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. *arXiv* **2015**, arXiv:1512.00567.
23. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* **1989**, *2*, 303–314. [CrossRef]
24. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
25. Asri, H.; Mousannif, H.; Moatassime, H.A.; Noel, T. Using Machine Learning Algorithms for Breast Cancer Risk Prediction and Diagnosis. *Procedia Comput. Sci.* **2016**, *83*, 1064–1069. [CrossRef]
26. Gupta, M.D.; Banerjee, S. Similarity Based Retrieval in Case Based Reasoning for Analysis of Medical Images. *Int. J. Comput. Inf. Eng.* **2015**, *8*, 539–545.
27. Sharma, A. Stochastic nonparallel hyperplane support vector machine for binary classification problems and no-free-lunch theorems. *Evol. Intell.* **2020**, *1*–20. [CrossRef]
28. Hsu, C.; Chang, C.; Lin, C. *A Practical Guide to Support Vector Classification*; 2010; Available online: <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf> (accessed on 3 January 2022).
29. Gallego-Posada, J.; Montoya-Zapata, D.; Quintero-Montoya, O. Detection and Diagnosis of Breast Tumors Using Deep Convolutional Neural Networks. Available online: <https://www.semanticscholar.org/paper/Detection-and-Diagnosis-of-Breast-Tumors-using-Deep-Gallego-Posada-Montoya-Zapata/9566d1f27a0e5f926827d3eaf8546dab51e40e21> (accessed on 3 January 2022).
30. Albayrak, A.; Bilgin, G. Mitosis detection using convolutional neural network based features. In Proceedings of the 2016 IEEE 17th International Symposium on Computational Intelligence and Informatics (CINTI), Budapest, Hungary, 17–19 November 2016; pp. 000335–000340. [CrossRef]
31. Platania, R.; Shams, S.; Yang, S.; Zhang, J.; Lee, K.; Park, S.-J. Automated Breast Cancer Diagnosis Using Deep Learning and Region of Interest Detection (BC-DROID). In Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Boston, MA, USA, 20–23 August 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 536–543. [CrossRef]
32. Geras, K.J.; Wolfson, S.; Shen, Y.; Wu, N.; Kim, S.G.; Kim, E.; Heacock, L.; Parikh, U.; Moy, L.; Cho, K. High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks. *arXiv* **2018**, arXiv:1703.07047.
33. Zhou, H.; Zhou, H. Mammogram Classification Using Convolutional Neural Networks. Available online: <https://www.semanticscholar.org/paper/Mammogram-Classification-Using-Convolutional-Neural-Zhou-Zhou/1d62f0078a6348be9e3e619dc1f1702ac2a87c49> (accessed on 3 January 2022).



34. Wang, H.; Cruz-Roa, A.; Basavanthally, A.; Gilmore, H.; Shih, N.; Feldman, M.; Tomaszewski, J.; Gonzalez, F.; Madabhushi, A. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *J. Med. Imaging Bellingham Wash* **2014**, *1*, 034003. [CrossRef]
35. LeCun, Y.A.; Bottou, L.; Orr, G.B.; Müller, K.-R. Efficient BackProp. In *Neural Networks: Tricks of the Trade*, 2nd ed.; Montavon, G., Orr, G.B., Müller, K.-R., Eds.; Lecture Notes in Computer Science; Springer: Berlin/Heidelberg, Germany, 2012; pp. 9–48. ISBN 978-3-642-35289-8. [CrossRef]
36. Salaken, S.M.; Khosravi, A.; Nguyen, T.; Nahavandi, S. Extreme learning machine based transfer learning algorithms: A survey. *Neurocomputing* **2017**, *267*, 516–524. [CrossRef]
37. Zhou, P.; Feng, J.; Ma, C.; Xiong, C.; Hoi, S.; E, W. Towards Theoretically Understanding Why SGD Generalizes Better Than ADAM in Deep Learning. *arXiv* **2021**, arXiv:2010.05627.
38. Su, S.S.W.; Kek, S.L. An Improvement of Stochastic Gradient Descent Approach for Mean-Variance Portfolio Optimization Problem. *J. Math.* **2021**, *2021*, e8892636. [CrossRef]
39. Frazão, X.; Alexandre, L.A. Weighted Convolutional Neural Network Ensemble. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*; Bayro-Corrochano, E., Hancock, E., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 674–681. [CrossRef]
40. Wang, S.; Wang, Y.; Wang, D.; Yin, Y.; Wang, Y.; Jin, Y. An improved random forest-based rule extraction method for breast cancer diagnosis. *Appl. Soft Comput.* **2020**, *86*, 105941. [CrossRef]
41. Risk Estimation Dataset Documentation: BCSC. Available online: <https://www.bscs-research.org/data/rfdataset/dataset> (accessed on 3 January 2022).
42. Barlow, W.E.; White, E.; Ballard-Barbash, R.; Vacek, P.M.; Titus-Ernstoff, L.; Carney, P.A.; Tice, J.A.; Buist, D.S.M.; Geller, B.M.; Rosenberg, R.; et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. *J. Natl. Cancer Inst.* **2006**, *98*, 1204–1214. [CrossRef] [PubMed]
43. Mukherjee, M.; Khushi, M. SMOTE-ENC: A Novel SMOTE-Based Method to Generate Synthetic Data for Nominal and Continuous Features. *Appl. Syst. Innov.* **2021**, *4*, 18. [CrossRef]
44. Blagus, R.; Lusa, L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinform.* **2013**, *14*, 106. [CrossRef] [PubMed]
45. Fernandez, A.; Garcia, S.; Herrera, F.; Chawla, N.V. SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *J. Artif. Intell. Res.* **2018**, *61*, 863–905. [CrossRef]
46. MathWorks—Makers of MATLAB and Simulink. Available online: <https://ww2.mathworks.cn/en/> (accessed on 15 August 2021).
47. Parikh, R.; Mathai, A.; Parikh, S.; Chandra Sekhar, G.; Thomas, R. Understanding and using sensitivity, specificity and predictive values. *Indian J. Ophthalmol.* **2008**, *56*, 45–50. [CrossRef] [PubMed]
48. Guan, Q.; Wan, X.; Lu, H.; Ping, B.; Li, D.; Wang, L.; Zhu, Y.; Wang, Y.; Xiang, J. Deep convolutional neural network Inception-v3 model for differential diagnosing of lymph node in cytological images: A pilot study. *Ann. Transl. Med.* **2019**, *7*, 14. [CrossRef]
49. Alom, M.Z.; Taha, T.M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M.S.; Van Esesn, B.C.; Awwal, A.A.S.; Asari, V.K. The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches. *arXiv* **2018**, arXiv:1803.01164.
50. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
51. Aruna, S.; Rajagopalan, D.; Nandakishore, L. Knowledge based analysis of various statistical tools in detecting breast cancer. *Comput. Sci. Inf. Technol.* **2011**, *2*, 37–45. [CrossRef]
52. Zhuang, J.; Tang, T.; Ding, Y.; Tatikonda, S.; Dvornek, N.; Papademetris, X.; Duncan, J.S. AdaBelief Optimizer: Adapting Stepsizes by the Belief in Observed Gradients. *arXiv* **2020**, arXiv:2010.07468.