mathematics

MDPI

Article

Deep Learning Cascaded Feature Selection Framework for Breast Cancer Classification: Hybrid CNN with Univariate-Based Approach

Nagwan Abdel Samee [1], Ghada Atteia [1,*], Souham Meshoul [1,*], Mugahed A. Al-antari [2] and Yasser M. Kadah [3,4]

[1] Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[2] Department of Artificial Intelligence, College of Software & Convergence Technology, Daeyang AI Center, Sejong University, Seoul 05006, Korea
[3] Electrical and Computer Engineering Department, King Abdulaziz University, Jeddah 22254, Saudi Arabia
[4] Biomedical Engineering Department, Cairo University, Giza 12613, Egypt
* Correspondence: geatteiaallah@pnu.edu.sa (G.A.); sbmeshoul@pnu.edu.sa (S.M.)

Abstract: With the help of machine learning, many of the problems that have plagued mammography in the past have been solved. Effective prediction models need many normal and tumor samples. For medical applications such as breast cancer diagnosis framework, it is difficult to gather labeled training data and construct effective learning frameworks. Transfer learning is an emerging strategy that has recently been used to tackle the scarcity of medical data by transferring pre-trained convolutional network knowledge into the medical domain. Despite the well reputation of the transfer learning based on the pre-trained Convolutional Neural Networks (CNN) for medical imaging, several hurdles still exist to achieve a prominent breast cancer classification performance. In this paper, we attempt to solve the Feature Dimensionality Curse (FDC) problem of the deep features that are derived from the transfer learning pre-trained CNNs. Such a problem is raised due to the high space dimensionality of the extracted deep features with respect to the small size of the available medical data samples. Therefore, a novel deep learning cascaded feature selection framework is proposed based on the pre-trained deep convolutional networks as well as the univariate-based paradigm. Deep learning models of AlexNet, VGG, and GoogleNet are randomly selected and used to extract the shallow and deep features from the INbreast mammograms, whereas the univariate strategy helps to overcome the dimensionality curse and multicollinearity issues for the extracted features. The optimized key features via the univariate approach are statistically significant (p-value ≤ 0.05) and have good capability to efficiently train the classification models. Using such optimal features, the proposed framework could achieve a promising evaluation performance in terms of 98.50% accuracy, 98.06% sensitivity, 98.99% specificity, and 98.98% precision. Such performance seems to be beneficial to develop a practical and reliable computer-aided diagnosis (CAD) framework for breast cancer classification.

Keywords: feature dimensionality curse (FDC); deep transfer learning; feature reduction and selection; CAD system; breast cancer

MSC: 92B20

check for updates

Citation: Samee, N.A.; Atteia, G.; Meshoul, S.; Al-antari, M.A.; Kadah, Y.M. Deep Learning Cascaded Feature Selection Framework for Breast Cancer Classification: Hybrid CNN with Univariate-Based Approach. *Mathematics* **2022**, *10*, 3631. https://doi.org/10.3390/math10193631

Academic Editors: Joanna Dipnall and Lan Du

Received: 5 September 2022
Accepted: 28 September 2022
Published: 4 October 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/).

1. Introduction

According to the World Health Organization (WHO) [1], there were 2.3 million new cases of breast cancer diagnosed in women around the world in 2020, leading to 685,000 deaths. It is projected that by the end of the year 2020, there will be 7.8 million living women who have been diagnosed with breast cancer within the previous five years,

making it the most prevalent form of cancer in the world [2]. Radiotherapy, including digital mammography, ultrasound [3], and magnetic resonance imaging, plays an essential role in the treatment of breast cancer. With breast cancer in its early stages, radiation can prevent a woman from requiring a mastectomy. Radiologists may misdiagnose 30% of breast cancers based on breast density [4]. Even experienced radiologists struggle to interpret many screening mammograms. Computer-aided diagnosis, CAD, is a major medical imaging diagnosing tool. CAD is an Artificial Intelligence-based diagnosing tool and was originally developed to help radiologists in examining medical images and highlighting possible areas of concern [5,6].

Convolutional neural networks, CNNs, are a type of machine learning approach and necessitate the training of numerous parameters. Typically, an effective training of CNNs for computer vision applications requires millions of training samples [7,8]. Representation learning enables a CNN to learn how to extract features from an image using convolutional layers from the shallow to the deep. Lines and edges are more generic in the shallow layers of a CNN but become more relevant to the target application as the layers become deeper. It is possible to extract features using convolutional filter weights. Transfer learning is commonly used to train CNNs for extracting the deep features in medical imaging due to the lack of medical datasets [9]. Transfer learning is oriented on the premise that previously acquired knowledge can be uniquely applied for solving new problems more efficiently and effectively. Consequently, transfer learning involves reusing previously learned relevant knowledge [10,11]. In 2016, numerous pre-trained CNN learning models, including AlexNet [12], GoogLeNet [13], ResNet [14], VGG [15], and Inception V3 [16], emerged to solve classification problems in natural images using ImageNet. Subsequently, transfer learning was recently applied to breast cancer imaging [17]. The utilization of pre-trained CNNs as a feature extraction approach for the classification of lesions in breast tissues has overcome the overfitting problem [18]. Because of the relatively small amount of data compared to the extensive number of network parameters, overfitting is a significant problem that arises when deep learning models are applied to medical data [19]. However, there still exist some challenges when attempting to apply the transfer learning approach in the classification of medical images.

The dimension space of the extracted features through the use of transfer learning has the potential to be quite large due to the fact that the number of features extracted is dependent on the architecture and the number of layers of the pre-trained CNN that is being used. For instance, the number of features retrieved from AlexNet/VGG16 and GoogleNet are, respectively, 4096 and 1024. The high dimensionality of the space of the extracted deep features and the limited number of the available medical samples creates the 'Curse of Dimensionality' problem [9]. In addition, the extracted features from the pre-trained CNN may suffer from the multicollinearity problem. When the extracted features of a dataset are highly correlated with one another, a phenomenon known as multicollinearity occurs. The effectiveness of regression and classification models is impacted by this problem [20]. Subsequently, a subset of the significant features has be chosen either before or at the same time as the classifier is being formulated. In this work, we extend our prior work in using deep learning for breast cancer classification in digital mammograms by proposing a cascaded feature selection framework using pre-trained CNNs and a univariate-based paradigm. The extracted features from training the pre-trained CNNs to classify breast mammograms are submitted to a next stage of feature filtration using our previously published univariate FS-based method for finding biomarkers in gene microarrays [21].

1.1. Research Questions

In order to overcome the inaccuracy of breast tumor diagnosis in mammography, the following are some research questions that need to be answered.

- How to construct a CAD system based on pre-trained CNNs for more precisely classifying breast lesions from mammograms as benign or malignant?

- How to deal with the feature dimensionality curse and multicollinearity in pre-trained CNN extracted features?

1.2. Research Contributions

The following contributions have been provided within the scope of this study:

- A novel breast classification framework is proposed based on the hybrid CNN with a univariate-based approach.
- Transfer learning using three state-of-the-art deep learning models is used to derive the high-level deep features.
- Resolving the overfitting problem that could arise when the pre-trained deep learning models are applied to the medical breast mammograms.
- Resolving the feature dimensionality curse (FDC) problem to avoid feature redundancy and select only the optimal significant key features for breast cancer classification.
- INbreast, a publicly available benchmark dataset, was used to perform a comprehensive evaluation of the proposed CAD system.
- A comprehensive evaluation and comparison study is conducted to show the capability and the feasibility of the proposed CAD system against the latest deep learning models.

2. Related Work

Breast cancer is the second major cause of death in women, affecting 12.5 percent of women in various societies worldwide [22]. According to current literature, early detection of breast cancer is vital since it can result in up to a 40% decrease in mortality [23]. Medical imaging examination is the most effective method of diagnosing breast cancer. Some of the imaging techniques used for the diagnosing of breast cancer include digital mammography, ultrasound, and magnetic resonance imaging (MRI); however, digital mammography imaging can be considered the most important method for early detection [24]. Many of the issues that are associated with mammography in terms of the classification and detection of breast cancer have been resolved through the use of machine learning. These issues include false positive rates, subjective judgments, and a limitation in indicating changes caused by cancer [25]. Building effective prediction models requires a vast number of normal/tumor samples. However, it is challenging to obtain the necessary training data and develop effective learning models when used in medical applications such as breast mammography [26]. As a result, it is recommended to reduce, as much as possible, the amount of time and effort needed to acquire the training data [27]. In situations like these, it would be beneficial to transfer the knowledge gained from one activity to the target task. Transfer learning allows for a model that was trained on one domain to be used as the focus of learning in another one [10,28,29]. Transfer learning has been used extensively in mammography classification to improve CNN architectures [30–35]. Improvements in classification accuracies, precision, and speed of training are the major advantages of transfer learning [35]. Fusion of extracted deep features has yielded better performance of transfer learning in the classification of mammograms [36,37].

As the number of extracted deep features depends on the architecture and the number of layers of the pre-trained CNN being utilized, the dimension space of the extracted deep features through the application of transfer learning has the potential to be rather huge. To the best of our knowledge, only a few previous studies have addressed the Curse of Dimensionality induced by the extracted deep features using transfer learning. In [38] the space of the extracted deep features has been reduced using the principal component analysis (PCA) and then submitted to conventional classifier, Support Vector Machine (SVM). The application of PCA for reducing the feature space for retrieved features from transfer learning was originally proposed in [39] and has resulted in greater performance in the classification precision than yielded one from the usage of only the extracted deep features. Another contribution for resolving the problem of the Curse of Dimensionality of the recovered deep features has been introduced by Samee et al., in [18], by empowering the performance of PCA through applying the Logistic Regression to identify the significant

principal components produced from the PCA analysis. Inspired by such mentioned ideas, we propose a new framework for selecting the most prominent key deep features from the extracted ones via the pre-trained CNNs after being trained on breast mammograms of benign or cancerous breast tissues. The remainder of the paper is structured as follows: Section 2 introduces the proposed methodology; Section 3 summarizes the experimental results and discusses them. The conclusion and future trends are presented in Section 5.

3. Methodology

We propose a cascaded Feature Selection framework for selecting non-redundant and relevant features for the classification of breast lesions in digital mammograms in this study. This work is an extension of our previous work utilizing transfer learning to classify breast cancer. Two cascaded FS stages are used to pick the features. Pre-trained CNNs such as AlexNet, VGG, and GoogleNet are used in the initial stage to extract shallow and deep features in the region of interest in breast mammograms. This will aid in avoiding the overfitting issue discussed in [40]. The selection of non-redundant significant characteristics in the returned set of features from the first stage is strengthened in the second stage by the use of the univariate-based FS paradigm, which can help in overcoming the high dimensionality and multicollinearity issues due to the extracted deep features. As depicted in Figure 1, the proposed framework is comprised of four modules. These modules include preprocessing of digital mammograms and region of interest (ROI) extractions, feature selection, feature classification utilizing conventional machine learning techniques, and classifier evaluation utilizing the performance metrics that are associated with CAD systems.

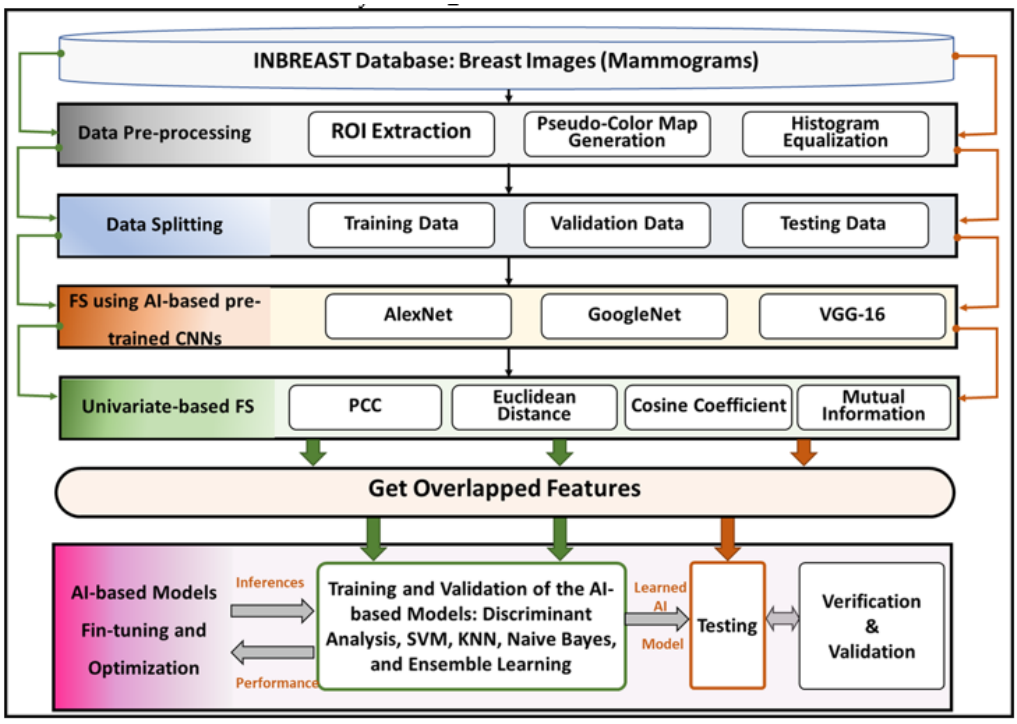


Figure 1. Deep learning cascaded feature selection framework for breast cancer classification.

3.1. Images Pre-Processing and ROI Extraction

The data that were used in this investigation came from a variety of publicly available sources, such as the INbreast for data based on full-field, high-resolution digital mammography. All of the mammograms included in the INbreast database were acquired in a Breast Centre located within a University Hospital [41]. The INbreast contains a total of 410 mammograms, which correspond to 115 unique patient cases. The INbreast dataset contains the correct diagnosis for all mammograms as well as the locations of the potential

abnormalities within each image, specified by the center and radius of the surrounding circle for each lesion. To perform our experiments, 32 × 32 square regions of interest (ROIs) inside of the lesion were picked up. Keeping the ROI as narrow as possible and also providing a statistically representative sample of the lesion facilitated improved localization performance of the developed method [42]. To eliminate any possibility of bias, the aberrant locations were hand-picked from among the available lesions. The samples also reflected the various abnormality subclasses, with lesion or cluster sizes that were adequate to encompass the selected ROI size.

We have recently introduced an effective method, pseudo-color mapping [18], for the preprocessing of these datasets and it has yielded good impact on the precision of the classification of lesions in the mammograms. When using a CNN, the pseudo-color mapping can take advantage of the three input channels available (Red, Green, and Blue). Therefore, we have employed the same datasets, pseudo-colored images in this study. The pseudo-color mapping can be summarized as follows. An area of size of 32 by 32 pixels was extracted from the lesion, ROIs. We chose such a size for the ROI in order to compare our proposed framework to previous work that handled the same classification problem using other techniques [43–45]. The produced files contain 34, and 73 ROIs extracted from the benign/malignant samples in the INbreast dataset. Data augmentation, principally based on the flipping (up/down and left/right) and rotation approach, was employed to increase the size of the datasets used. Pseudo-color mapping incorporates the original image into the red channel. The green and blue channels, on the other hand, receive different processed images as a result of the application of contrast-limited adaptive histogram equalization (CLAHE) [46] and intensity adjustment [43]. The pseudo-color mapping makes it possible for several levels of global information to be embedded into each pixel.

3.2. Feature Selection Using Pre-Trained CNNs and the Univariate-Based Approach

Pre-trained CNNs have convolutional, pooling, and fully connected (FC) layers [47]. A fully connected layer classifies the features extracted by convolution layers [48]. In this investigation, we used the convolutional and pooling layers of pre-trained CNNs to extract key characteristics from benign/malignant breast cancer images. AlexNet, VGG, and Googlenet are popular pre-trained CNN image classifiers. Alex Krizhevsky and his team introduced AlexNet [12] in 2012's ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Their CNN has more levels than earlier models. They've achieved top-5 error rates in the ILSVRC challenge. As shown in Figure 2, AlexNet has an input image layer of size 227 × 227 × 3, five convolutional layers, three max pooling layers, two fully connected (FC) layers, and a SoftMax layer as an output layer. The 1st convolutional layer has 96 filters [49], kernels, of sizes of 11 × 11. The output of this layer is normalized using the response normalization [12] and followed by an overlapping max pooling layer. The 2nd convolutional layer has 256 kernels of sizes of 5 × 5, and the output is then normalized and pooled by the same approaches mentioned above. The structure of the 3rd, 4th, and 5th convolutional layers is similar and each one has 384 filters of sizes of 3 × 3. Then, the output of the 5th convolutional layer is similarly normalized and pooled and applied to two FC layers, each having 4096 neurons. Finally, a SoftMax layer of two neurons are used as an output layer.

Karen Simonyan and Andrew Zisserman introduced VGG [15] in 2014. Using VGG, network depth affects the CNN's accuracy. After convolutional layers and before pooling, ReLU is activated. The architecture of the VGG 16 is very straightforward, as illustrated in Figure 3. It contains an input layer that takes images with a size of 224 × 224 × 3. The input layer is followed by a stack of convolutional layers. Each convolutional layer comprises a set of small size filters, 3 × 3. The small size of the convolution kernels permits adding the large number of convolutional layers in the VGG networks and has helped in yielding an enhanced of their performance [15]. The stack of convolutional layers is followed by three fully connected layers. The first two FC layers contain 4096 neurons each, but the third one

has only two neurons for the binary classification of the mammogram's tissues (normal, and abnormal). Finally, a soft-max layer of size $1 \times 1 \times 2$ is employed.

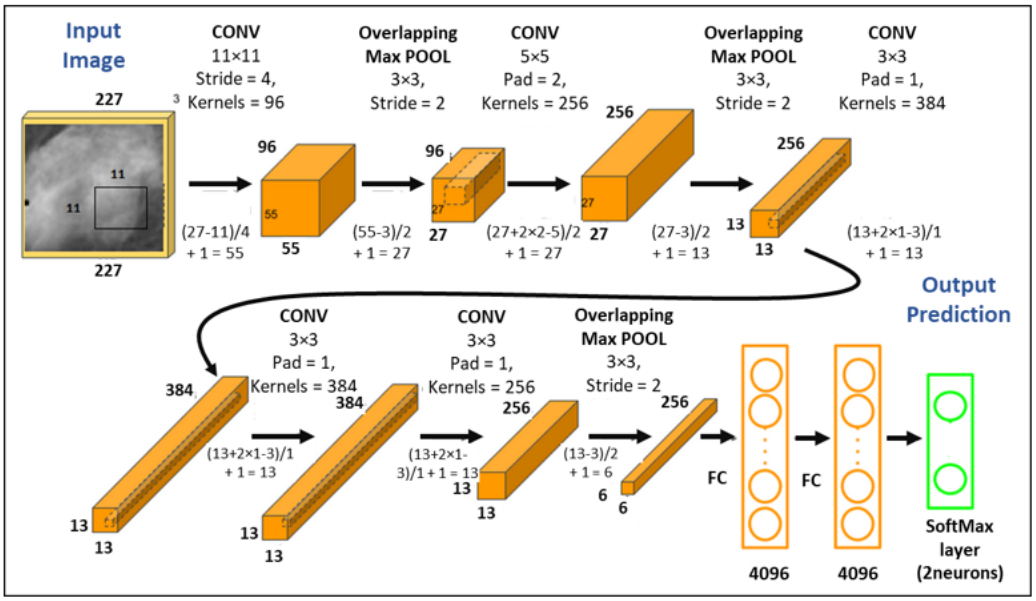


Figure 2. AlexNet flowchart architecture.

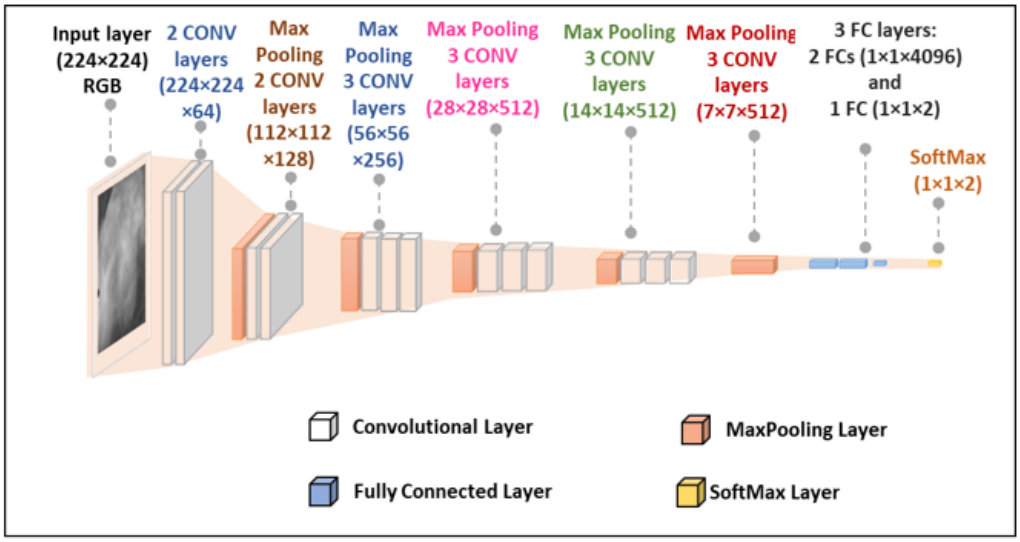


Figure 3. Pretrained CNN, VGG16, flowchart architecture.

Szegedy C. et al., introduced GoogleNet [13] in 2014's ILSVRC challenge. GoogleNet CNN contains 22 layers. The architecture of GoogleNet is wider and deeper than AlexNet. The processing of such deeper network can be very expensive unless other approaches are used to help in decreasing such computations. So, in the GoogleNet, there are three approaches from the NIN (Network in Network) that have been employed, including: inception modules, global average pooling, and the 1×1 convolution. The major concept of the inception module is employing multiple convolutional and pooling operations with several kernel sizes in parallel, which can help in extracting more features from the input patches. As shown in Figure 4a, an inception module has three convolutional layers, colored by gray color, of sizes 1×1, 3×3, and 5×5 that works in parallel along with a max pooling layer of size 3×3. The inception module receives its input from its earlier layer and applies its parallel operations, then the output feature maps are concatenated and passed to the following module. This approach creates a wider network. The processing of the inception

module has been alleviated by using a 1 × 1 convolution, shown in Figure 4b, on its internal layers. The 1 × 1 convolution has been created in the NIN deep network. This approach can help in reducing the number of required operations in the processing of the deep convolutional layers. It has been used in the GoogleNet as a step of nonlinear dimensional reduction, which in turn reduces the necessary computational cost. By decreasing the computation overhead of the deep convolution layers, more inception modules can be added. Finally, the global average pooling is employed before the FC layer by averaging the extracted feature maps. The overall architecture of GoogleNet consists of 22 layers, including the input layer, convolutional layers, max pooling, inception modules, FC, and softmax layers. GoogleNet contains 12 less parameters than AlexNet, making it easier to train.

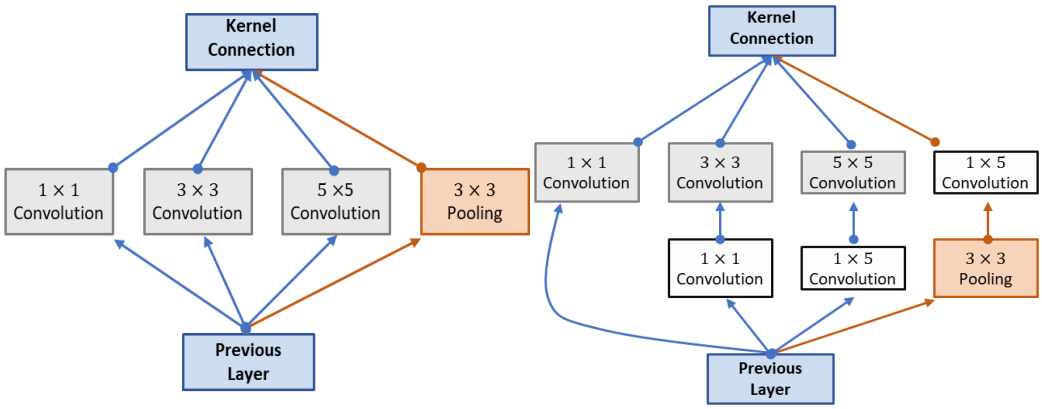


Figure 4. An inception module in the GoogleNet DCNN. (**a**) An inception module without 1 × 1 convolution layer. (**b**) An inception module with a 1 × 1 convolution layer.

Abdelhafiz et al. [50] reviewed various research articles on breast cancer classification. Because of their great accuracy and minimal complexity, AlexNet, VGG, and GoogleNet CNNs were used in this study. Given its high dimensionality and non-convex objective function, determining the ideal parameters for each pre-trained CNN to provide the highest performance is a difficult optimization problem. As a result, stochastic optimization approaches are commonly used. The stochastic gradient descent with momentum (SGDM) optimizer is utilized in this work with the learning parameters listed in Table 1. These settings were chosen after monitoring validation outcomes through testing and were applied to all networks to allow direct comparison of their results and computational costs.

Table 1. The values of learning parameters for training pre-trained CNNs.

Learning Parameters	Value
Learning rate	0.0001
Number of training epochs	400
Batch size	16
Momentum factor	0.9
L2-Regularization	0.0005

The process of selecting a feature subset (FS) involves deleting features that are either unnecessary or duplicated elsewhere in the system. The subset of features that is chosen should, in accordance with the criteria of some objective function, produce the highest level of performance. FS is an NP-hard problem, which means that it is difficult in a nondeterministic polynomial amount of time [51]. Because the amount of data that needs to be processed has become more substantial over the course of the past years, feature

selection has developed into a prerequisite for developing effective ML models. In contrast to the methods of feature extraction, the approaches of feature selection do not change the initial representation of the data [52]. Avoiding overfitting the data is one of the goals of both the feature extraction and selection. This is done in order to make further learning effective and accurate. FS algorithms are classified into three types: filter, wrapper, and embedding. The filter-based FS method retrieves features from data without requiring any learning. Wrappers employ ML approaches to determine which features are useful. Embedded approaches merge the feature selection and classifier development steps. They can be classified as either multivariate or univariate approaches [19]. Univariate approaches focus on analyzing a single feature at a time, and multivariate methods analyze the features in a way that enables them to discover links amongst themselves.

A common objective of the univariate filter-based FS approaches is feature ranking. The feature ranking can be done by several ways, including unconditional mixture modelling and information gain (IG) [53]. The unconditional mixture model makes the assumption that there are two distinct states of the feature, on and off, and determines whether or not the underlying binary state of the feature has an effect on the classification [21]. The unconditional mixture model is a univariate filter-based FS method and can do its job without taking the classifier into account. Because of this, it is extremely effective in terms of computing [53]. We have utilized it in earlier work for detecting key genes in a microarrays dataset for the classification of normal and abnormal liver tissues [19,21], and it has yielded great results. This inspired us to study the power of the unconditional mixture model univariate filter-based FS in picking the key features in the extracted set of features yielded from pre-trained CNNs for classifying breast lesions (benign/malignant). The main notion is based on identifying an on/off ideal feature whose values vary in normal and abnormal samples. As depicted in Figure 5, two ideal features have been applied in this study. The upregulated ideal feature is represented by a vector with two distinct sets of values $(-1, 1)$ for normal and abnormal instances, and a downregulated ideal feature with values of $(1, -1)$ is presented for both specimens. The Pearson correlation coefficient (PCC), the cosine coefficient (CC), the Euclidean distances (ED), and the mutual information (MI) are then used to determine the degree of similarity between the features that have been retrieved from pre-trained CNNs and the ideal features. The features are ranked according to the similarities that have been measured, and then the features that have the highest rankings are the ones that are returned as key features, which are used for the classification of breast lesions.

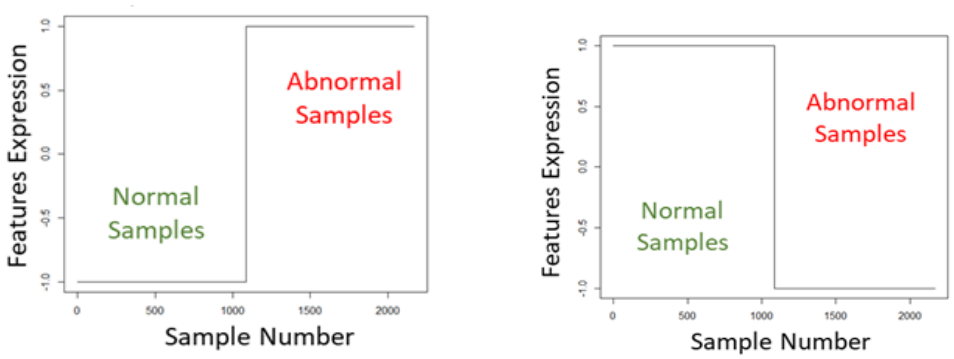


Figure 5. The on/off state of the ideal features. (**a**) Ideal feature (upregulated). (**b**) Ideal feature (downregulated).

The extracted feature matrix (EFM) from the pre-trained CNNs is then submitted to the second stage of feature filtration. The entries of the EFM have been normalized using the z-score technique. The normalized values have a standard deviation of one and are

centered around zero. The Z_{score} of a random feature F with a mean value of M and a standard deviation of σ are determined by Equation (1).

$$Z_{score} = \frac{\mathrm{F} - \mathrm{M}}{\sigma} \tag{1}$$

The Pearson's correlation coefficient [54,55], which is indicated in Equation (2), can be used to determine the degree of similarity between the ideal feature, F_{Ideal}, and a feature F. This can be done by comparing the two sets of values.

$$r = \frac{\sum_{i=1}^{S}\left(F_i - \overline{F}\right)\left(F_{Ideal_i} - \overline{F_{Ideal}}\right)}{\sqrt{\sum_{i=1}^{S}\left(F_i - \overline{F}\right)^2}\sqrt{\sum_{i=1}^{S}\left(F_{Ideal_i} - \overline{F_{Ideal}}\right)^2}} \tag{2}$$

Measuring the space-based distance between two vectors allows one to determine their similarity. Therefore, we utilized the Euclidean distance between the ideal feature, F_{Ideal}, and all features in the EFM. S denotes the size of the extracted feature matrix. F_i and F denote the value of a feature in any sample, i, and its mean value in all input samples respectively. The value of the ideal feature sample i is represented by F_{Ideal_i}, whereas the mean value of this feature across all input samples is represented by $\overline{F_{Ideal}}$.

For a given feature F_i in the EFM, the cosine coefficient can be used to determine the degree of dependence between the ideal feature and F_i, as illustrated in Equation (3). It is possible to tell if two angles are pointing in the same direction by looking at the cosine coefficient; if it is zero, it shows that the angles are independent.

$$r_{\text{cosine}} = \frac{\sum_{i=1}^{S}\left(F_i - \overline{F}\right)F_{Ideal_i}}{\sqrt{\sum_{i=1}^{S}\left(F_i - \overline{F}\right)^2}\sqrt{\sum_{i=1}^{Q}F_{Ideal}{}_i^2}} \tag{3}$$

The mutual information between the ideal feature, F_{Ideal}, and all features in the EFM can also be used to pick the key features [56–58]. In Equation (4), the formula for calculating mutual information is given, where the $H(F_{Ideal})$ is the entropy of the ideal feature and $H(F_{Ideal} \mid F)$ is the conditional entropy between F_{Ideal} and a feature, F, in the EFM.

$$I(F_{Ideal}, \mathrm{F}) = H(F_{Ideal}) - H(F_{Ideal}|\mathrm{F}) \tag{4}$$

3.3. Classification and Model Assessment

The statistical ML system that was used in this investigation consisted of the implementation of six major types of conventional classifiers. Techniques that fall within this category include decision trees [59], naive Bayes, discriminant analysis, ensembles, KNN, and SVM [28,43]. The many parameters and variants that make up each classifier were tuned in order to get the highest feasible performance and decrease the problem of overfitting to the greatest extent feasible. This was done by utilizing 5-fold cross-validation to produce an accurate estimation of the performance. Within each fold, distinct parts of the images are delegated to their respective sets. To determine the overall performance of the system, the results gained from every fold are aggregated overall, and this process is repeated five times, which is like the number of folds. To be more specific, the data for cancerous cases are randomly divided into three sets, which are referred to as the training dataset, the validation, as well as the testing set. The percentages of split between both the various divides are as follows: 70%, 15%, and 15% for the training set, the validation set, and the testing set, respectively.

The introduced learning model has been assessed using the well-known performance metrics including the model accuracy, specificity, sensitivity, false negative rate (FNR), area under the curve (AUC), false positive rate (FPR), F1-score, and Matthew's correlation coefficient [28,43]. Because accuracy is a key performance indicator since it indicates the proportion of right classifications compared to the total number of observations, it does

not differentiate between false-positive and false-negative errors. This is detrimental to the success of a CAD system, because a false-negative classification potentially has far more severe consequences than a false-positive classification. This issue is solved by the sensitivity measure, which indicates the proportion of malignant cases for which a proper diagnosis was made. However, specificity indicates the fraction of typical individuals whose diseases were correctly identified. Together, they provide the complete picture, allowing an observer to draw accurate comparisons between different systems. In a CAD system, for example, if systems use the same degree of accuracy, the framework with the greater level of sensitivity is preferable.

3.4. Execution Environment

The experimental study is executed via a PC with CPU of Intel(R) Core (TM) i7-10700KF @ 3.80 GHz, 32.0 GB RAM, six CPUs and one GPU of NVIDIA GeForce RTX 3060.

4. Results and Discussion

In this investigation, we are introducing a significant framework of a computer-aided diagnosing system for the classification of breast tumor lesions using cascaded stages of pre-trained CNNs and unconditional mixture model univariate-based FS techniques. The framework comprises four stages including the data preparation, FS using pre-trained CNNs, univariate-based FS, and Classification. In [18], we have introduced the use of pseudo-colored mapping for the submission of input images to pretrained CNNs. The resulting performance prompted us to implement them in this study. Pseudo-colored mapping has been performed to INbreast, and the pseudo-colored images are used to evaluate the adequacy of the newly developed CAD system.

Pre-trained CNNs have been utilized as the first stage of FS. The ROIs for each dataset were resized to the respective size of each network (224 × 224 for the VGG/GoogleNet and 227 × 227 for AlexNet) using bilinear interpolation with an anti-aliasing filter in an effort to meet specifications, preserve image quality, and keep them free of aliasing artifacts, as the input layer cannot be modified as an aspect of the transfer learning approach. Throughout training of pre-trained CNNs, it is necessary to set the learning parameters. A stochastic gradient descent with momentum (SGDM) optimizer was used in this study. This optimizer's learning rate was adjusted to 0.0001, the L2-regularization setting to 0.0005, the momentum term factor to 0.9, and the gradient threshold technique to L2-norm. The maximum number of training epochs permitted was 400, and the maximum mini-batch size allowed was 16. The selection of these training alternatives was based on the validation findings gained from the running trials. They were applied across all networks to enable a direct comparison of the networks' respective results. According to what have been revealed in [18], the topology of the network influences the number of features that can be recovered from it. Using AlexNet, GoogleNet, and VGG16, respectively, a total of 4096, 1024, and 4096 features have been retrieved from the data.

An educational edition of MATLAB 2020a and the ML toolbox were used to develop and execute this study. At 2.60 GHz, the computing system's microprocessor is an Intel® CoreTM i7-6700HQ quad core. It also includes 16 gigabytes of memory storage (RAM) and a gpu that supports CUDA (NVIDIA GeForce GTX 950M with 4 gigabytes of memory). However, despite the fact that the results of the studies are machine and environment particular, their relative values can still be used to compare different approaches.

4.1. Signal Profiles and Hypothesis Testing of the Extracted Key Features

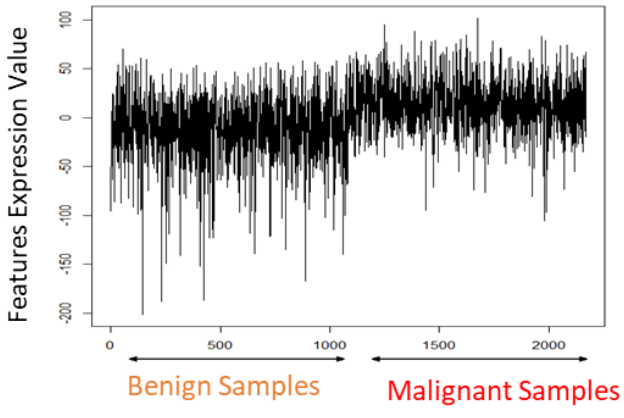
The extracted feature matrices from AlexNet, GoogleNet, and VGG16 have the following dimensions: 4096 × 2168, 1024 × 2168 and 4096 × 2168 respectively. The EFM is then submitted to the second stage of feature filtration. To determine the degree of similarity between the features retrieved from pre-trained CNNs and the ideal features, four univariate-based FS methods including the Pearson correlation coefficient, the cosine coefficient, the Euclidean distances, and the mutual information were used. The features are

sorted based on the similarities that have been measured, and the variables with the highest rankings are returned as key features, which are then exploited to classify breast lesions.

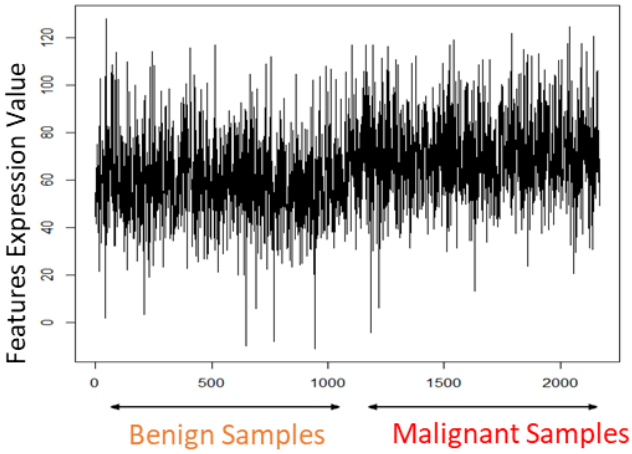
The significance of the extracted key features has been evaluated using the ANOVA F-statistics test [60], and it has been determined whether or not the means of their values in the benign and malignant samples are different from one another. The following metrics were used to calculate significance: the degree of freedom (DF), the t statistics (t), and the *p*-value of the retrieved features. Figures 6–9 depict the signal profiles for the top ten ranking extracted key features (upregulated and downregulated features) utilizing PCC, CC, ED, and MI on the EFM extracted from the AlexNet pre-trained CNN. This is done for the purpose of providing an illustration and inspection of the retrieved on/off features. Each feature is plotted in both normal and abnormal samples, and the *p*-value associated with that feature is appended to the same plot. The samples are shown along the X axis; there are 1084 samples for each of the disinfected samples, and the infected samples also have a total of 1084 samples. The value of the feature's expression is shown along the Y axis. The signal profile, in conjunction with the appropriate *p*-values of each on/off feature, ensures that the significance of these features for the classification of normal/up-normal breast lesions. Figure 7 shows that the majority of the features that were yielded by the ED feature selection do not have signal profiles that are comparable to those of the ideal key features and have insignificant *p*-values (> 0.0.5). But, as shown in Figures 6–9, the signal profiles and *p*-values of the extracted features from PCC, CC, and MI on the EFM extracted from AlexNet show that the returned key features are statistically significant (p-value < 0.05) and can be used to train the classification model. Figures 10 and 11 show the retrieved *p*-values of the top 50 ranked features (upregulated and downregulated) using all univariate-based FS methods on the EFM extracted from the GoogleNet and VGG16, respectively. The figures showing the results of statistical analysis of extracted on/off features from the GoogleNet and VGG16 reveal that features with high rankings from PCC, CC, and MI have significant *p*-values (0.05), whereas those with high rankings from ED feature selection have negligible *p*-values (> 0.05).

4.2. Classification Model Construction and Evaluation

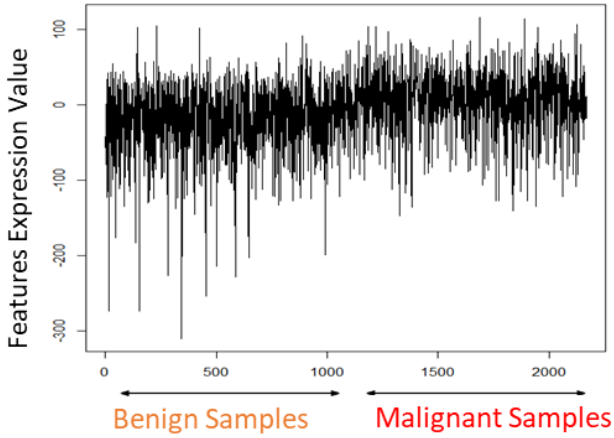
The top ranked significant features (20 features) are then submitted to six distinct families of classical classifiers as the final step in the proposed CAD framework. These classical classifier families include decision trees, discriminant analysis, SVM, KNN, naive Bayes, and ensemble. In order to test and evaluate how well the newly developed CAD system works, we have carried out five separate experiments. In each experiment, the features that have been extracted from the pre-trained CNNs (AlexNet, VGG, and GoogleNet) are fed to one method of the univariate filter-based FS methods (PCC, CC, ED, and MI) in order to determine the degree of similarity between the features that have been retrieved from pre-trained CNNs and the ideal features before they are submitted to the classical classifiers. This is done before the features are submitted to the classical classifiers. In the last phase of the experiment, a hybrid collection of the extracted features from all of the univariate FS approaches is utilized for the classification of breast tissues. Table 2 contains the results that were obtained from the conducted tests on the INbreast dataset. Only the results of the best performing models among the aforementioned ones are shown in Table 2. The application of a hybrid set of features given from all univariate FS method on the features that were derived from AlexNet has resulted in the best performance, as shown in Table 2 with a gray highlight color.



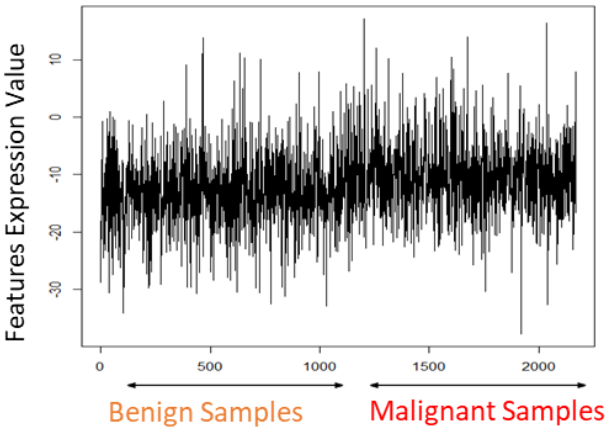
(**a**) $t = -17.975, DF = 2023.3, p\ value < 2.2 \times 10^{-16}$



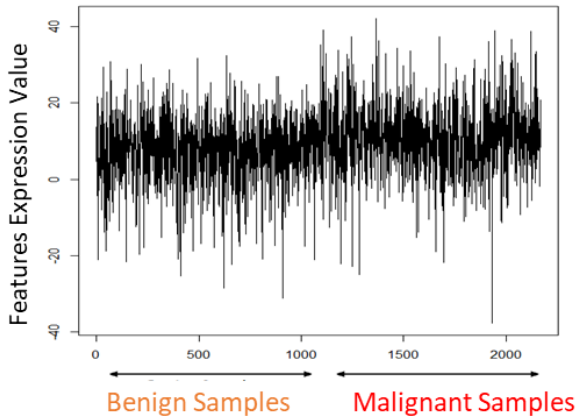
(**b**) $(t = -14.921, DF = 2165.8, p\ value < 2.2 \times 10^{-16})$



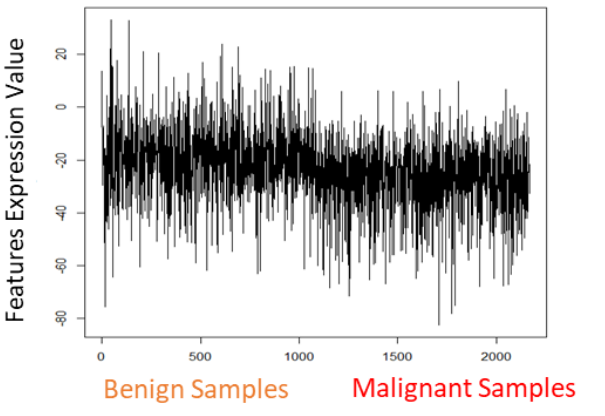
(**c**) $(t = -13.115, DF = 2120.5, p\ value < 2.2 \times 10^{-16})$



(**d**) $(t = -9.8024, DF = 2166, p\ value < 2.2 \times 10^{-16})$

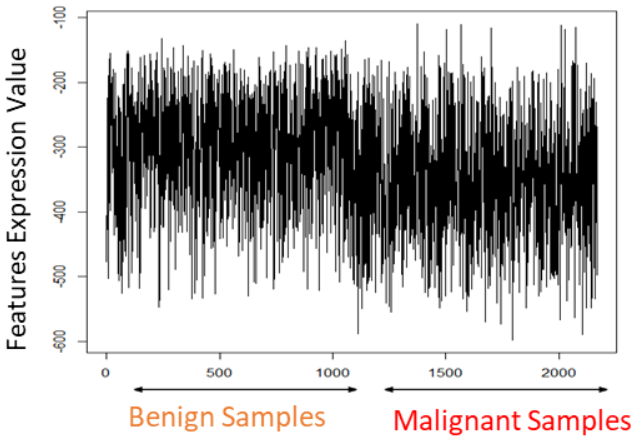


(**e**) (t = −9.3678, DF = 2165.5, p value < 2.2×10^{-16})

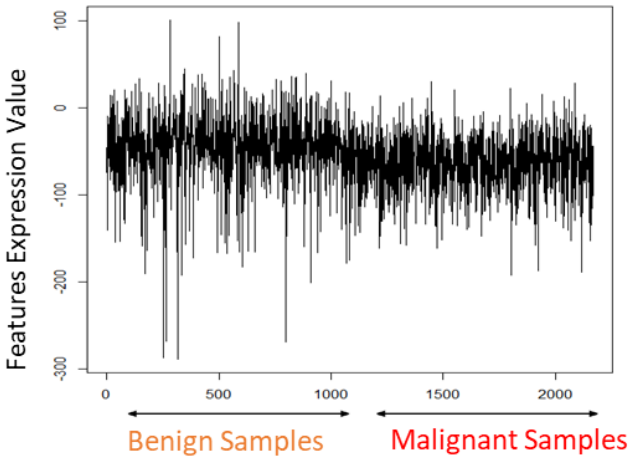


(**f**) (t = 13.732, DF = 2165.5, p value < 2.2×10^{-16})

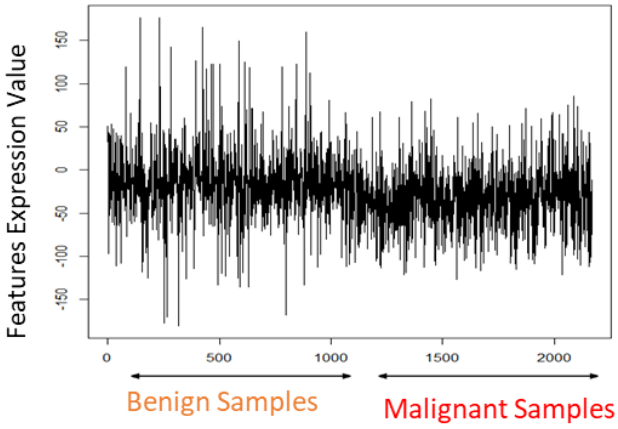
Figure 6. *Cont.*



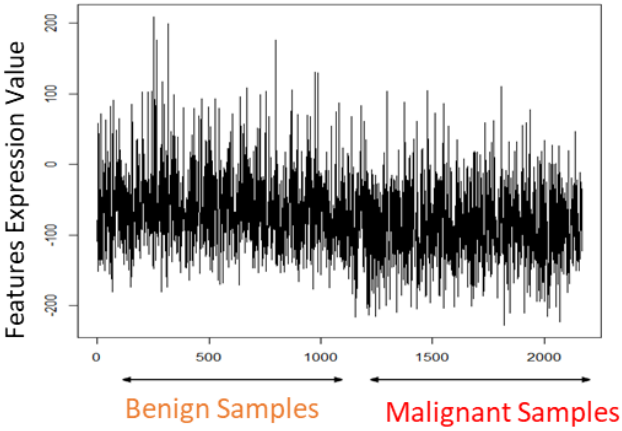
(**g**) (t = 13.078, DF = 2149.2, p value < 2.2×10^{-16})



(**h**) (t = 12.491, DF = 2086, p value < 2.2×10^{-16})

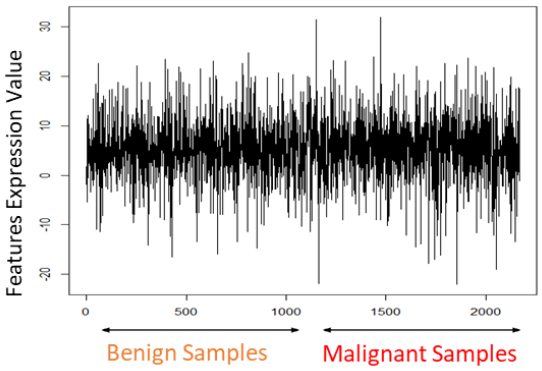


(**i**) (t = 11.425, DF = 2134.7, p value < 2.2×10^{-16})

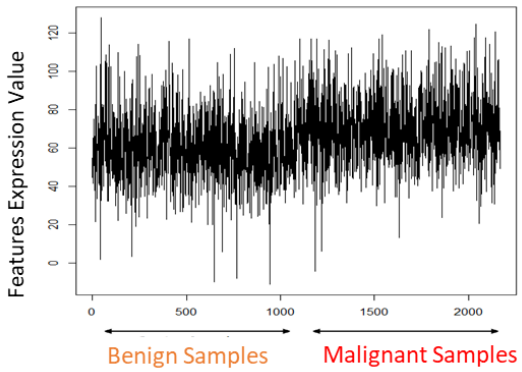


(**j**) (t = 10.694, DF = 2166, p value < 2.2×10^{-16})

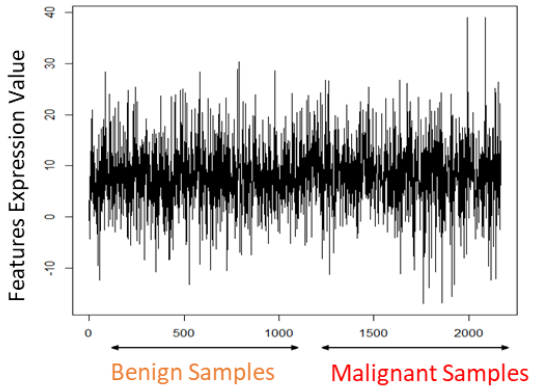
Figure 6. Signal profiles for the top ten ranking extracted key features (i.e., upregulated and downregulated features) utilizing PCC as well as the retrieved results from the ANOVA F-test, including the t-value, degree of freedom (DF), and p-value. The extracted upregulated and downregulated profile features are depicted in (**a**–**j**), respectively. (**a**) Top ranked upregulated feature no. 1. (**b**) Top ranked upregulated feature no. 2. (**c**) Top ranked upregulated feature no. 3. (**d**) Top ranked upregulated feature no. 4. (**e**) Top ranked upregulated feature no. 5. (**f**) Top ranked downregulated feature no. 1. (**g**) Top ranked downregulated feature no. 2. (**h**) Top ranked downregulated feature no. 3. (**i**) Top ranked downregulated feature no. 4. (**j**) Top ranked downregulated feature no. 5.



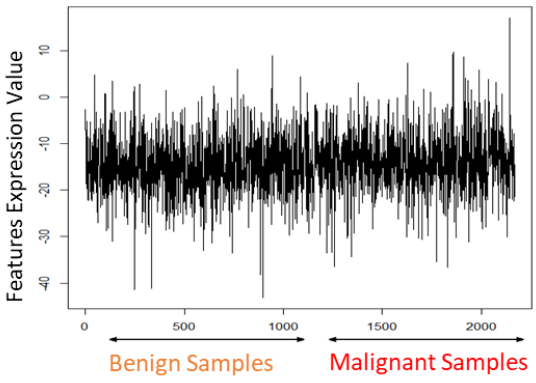
(**a**) (t = −1.7274, DF = 2149.6, p value = 0.08423)



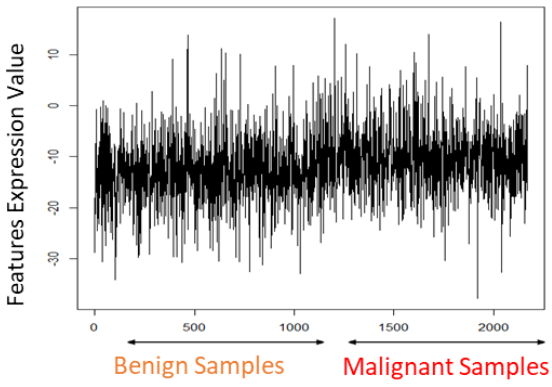
(**b**) (t = −0.31773, DF = 2149.4, p value = 0.7507)



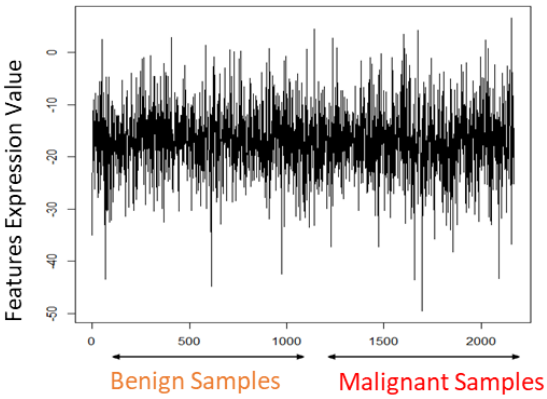
(**c**) (t = −2.7212, DF = 2162.2, p value = 0.006557)



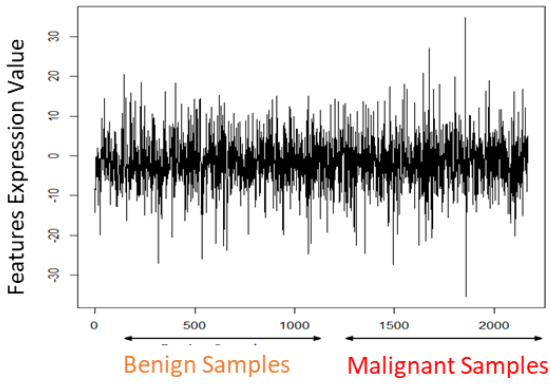
(**d**) (t = −3.2342, DF = 2165.2, p value = 0.0012)



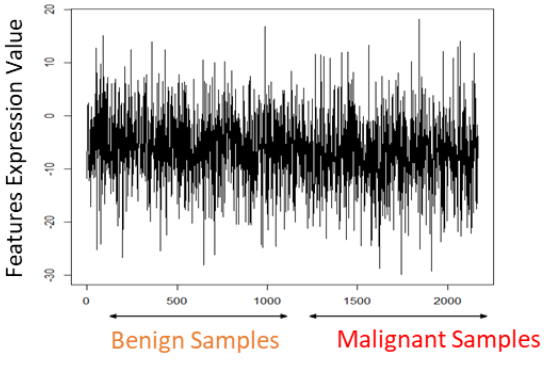
(**e**) (t = −9.8024, DF = 2166, p value < 2.2×10^{-16})



(**f**) (t = 1.7193, DF = 2154.1, p value = 0.08571)

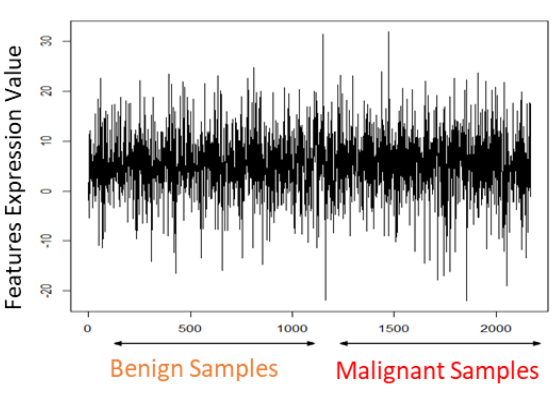


(**g**) (t = −0.31773, DF = 2149.4, p value = 0.7507)

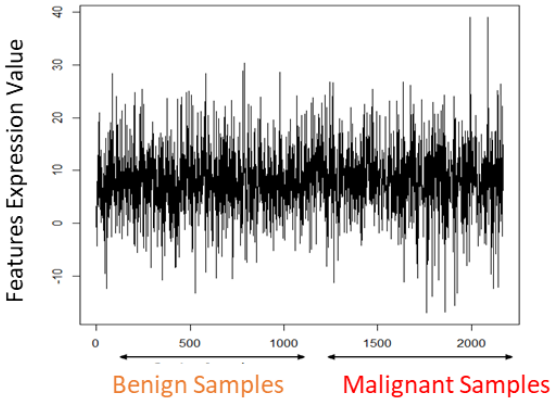


(**h**) (t = 6.2974, DF = 2161.4, p value= 3.65×10^{-10})

Figure 7. *Cont.*

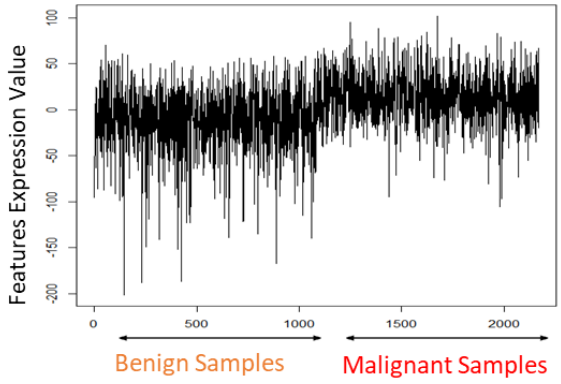


(**i**) (t = −1.7274, DF = 2149.6, p value = 0.08423)

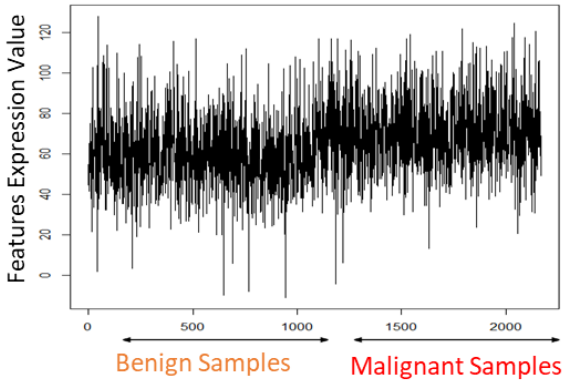


(**j**) (t = −2.7212, DF = 2162.2, p value = 0.007)

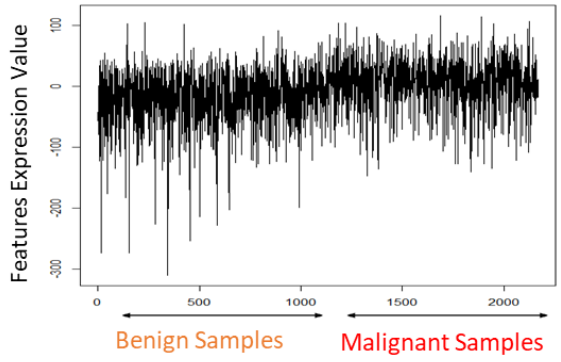
Figure 7. Signal profiles for the top ten ranking extracted key features (i.e., upregulated and downregulated features) utilizing Euclidean distance as well as the retrieved results from the ANOVA F-test, including the t-value, degree of freedom (DF), and p-value. The profiles in (**a**–**e**) depict the upregulated features, whereas the profiles in (**f**–**j**) show the downregulated features. (**a**) Top ranked upregulated feature no. 1. (**b**) Top ranked upregulated feature no. 2. (**c**) Top ranked upregulated feature no. 3. (**d**) Top ranked upregulated feature no. 4. (**e**) Top ranked upregulated feature no. 5. (**f**) Top ranked downregulated feature no. 1. (**g**) Top ranked downregulated feature no. 2. (**h**) Top ranked downregulated feature no. 3. (**i**) Top ranked downregulated feature no. 4. (**j**) Top ranked downregulated feature no. 5.



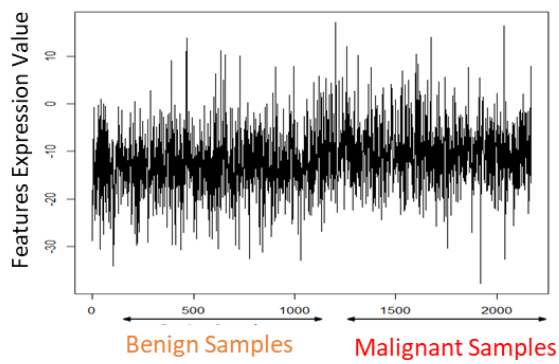
(**a**) (t = −17.975, DF = 2023.3, p value < 2.2×10^{-16})



(**b**) (t = −14.921, DF = 2165.8, p value < 2.2×10^{-16})

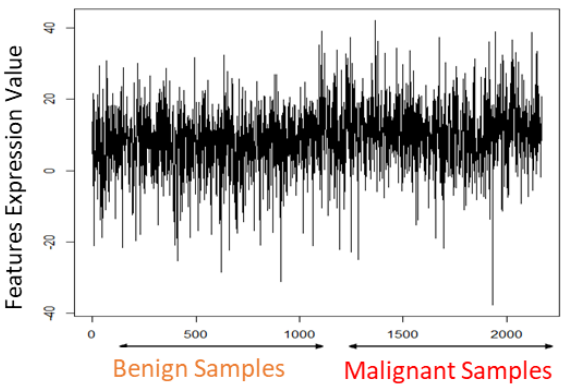


(**c**) (t = −13.115, DF = 2120.5, p value < 2.2×10^{-16})

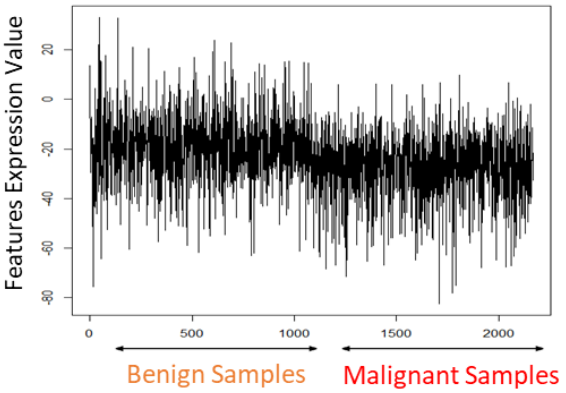


(**d**) (t = −9.8024, DF = 2166, p value < 2.2×10^{-16})

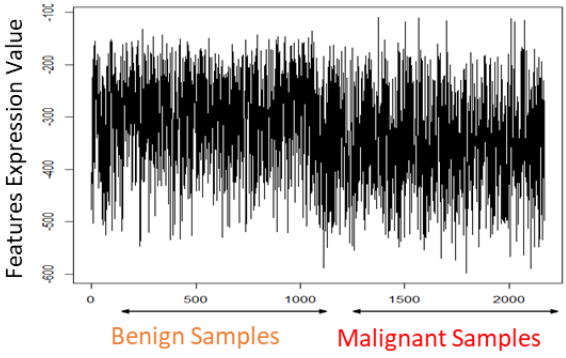
Figure 8. *Cont*.



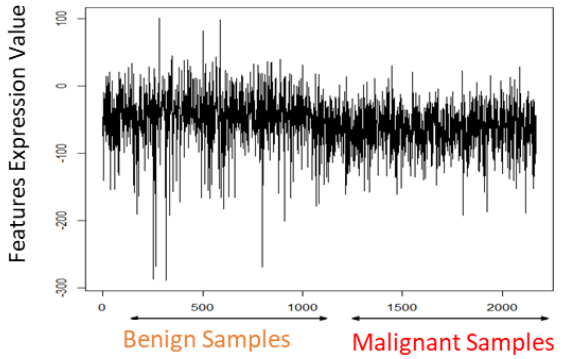
(**e**) ($t = -9.3678$, $DF = 2165.5$, p value $< 2.2 \times 10^{-16}$)



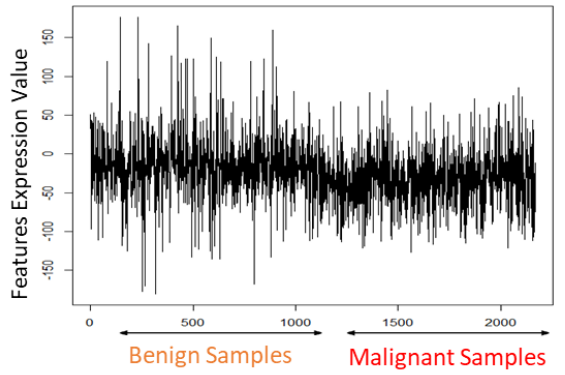
(**f**) ($t = 13.732$, $DF = 2165.5$, p value $< 2.2 \times 10^{-16}$)



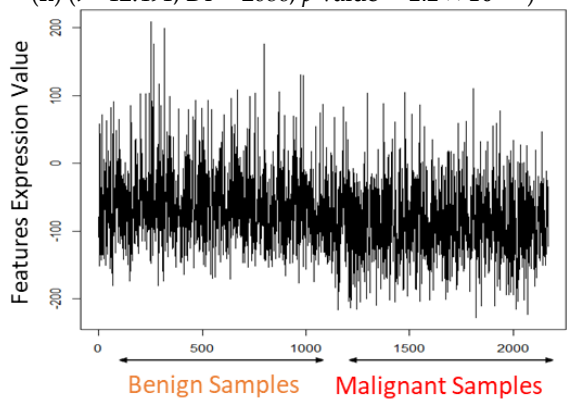
(**g**) ($t = 13.078$, $DF = 2149.2$, p value $< 2.2 \times 10^{-16}$)



(**h**) ($t = 12.491$, $DF = 2086$, p value $< 2.2 \times 10^{-16}$)

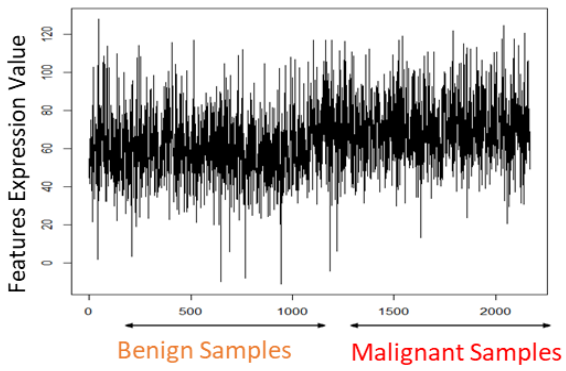


(**i**) ($t = 11.425$, $DF = 2134.7$, p value $< 2.2 \times 10^{-16}$)

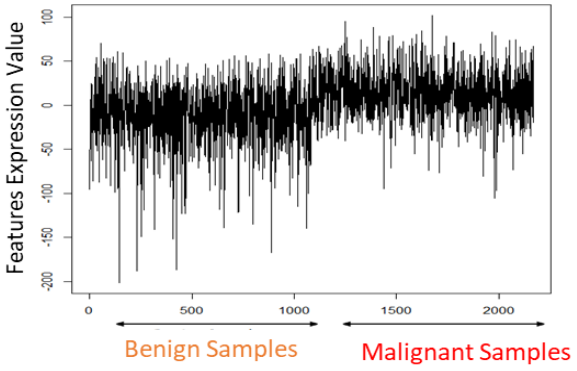


(**j**) ($t = 10.694$, $DF = 2166$, p value $< 2.2 \times 10^{-16}$)

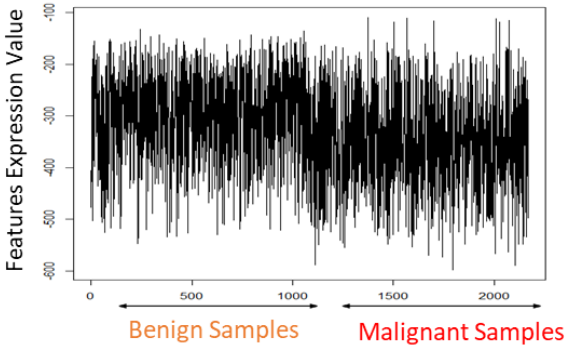
Figure 8. Top ten signal profiles identified key features (i.e., upregulated and downregulated features) using CC feature selection and ANOVA F-test results, including t-value, degree of freedom (DF), and p-value. The profiles in (**a**–**e**) depict the upregulated features, whereas the downregulated features are presented in (**f**–**j**). (**a**) Top ranked upregulated feature no. 1. (**b**) Top ranked upregulated feature no. 2. (**c**) Top ranked upregulated feature no. 3. (**d**) Top ranked upregulated feature no. 4. (**e**) Top ranked upregulated feature no. 5. (**f**) Top ranked downregulated feature no. 1. (**g**) Top ranked downregulated feature no. 2. (**h**) Top ranked downregulated feature no. 3. (**i**) Top ranked downregulated feature no. 4. (**j**) Top ranked downregulated feature no. 5.



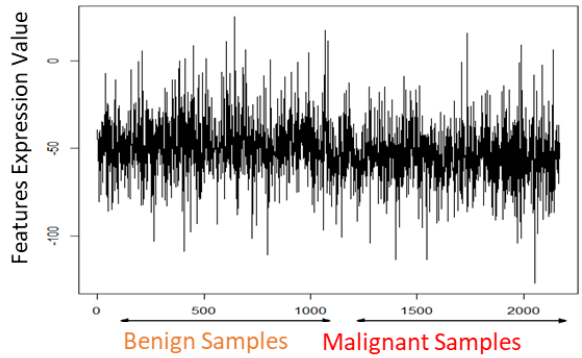
(**a**) ($t = -14.921$, $DF = 2165.8$, p value $< 2.2 \times 10^{-16}$)



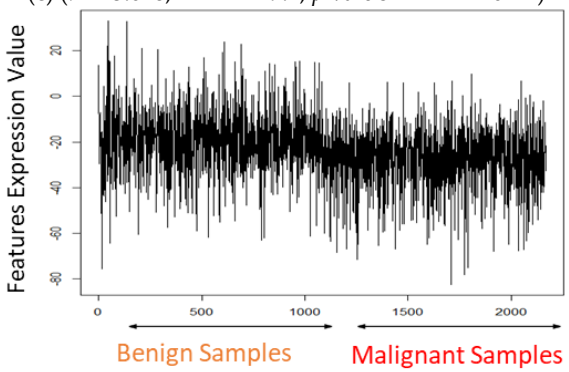
(**b**) ($t = -17.975$, $DF = 2023.3$, p value $< 2.2 \times 10^{-16}$)



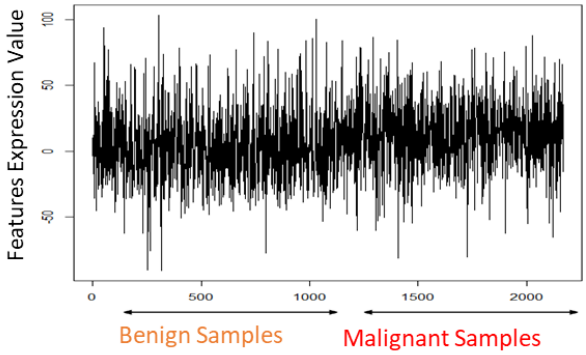
(**c**) ($t = 13.078$, $DF = 2149.2$, p value $< 2.2 \times 10^{-16}$)



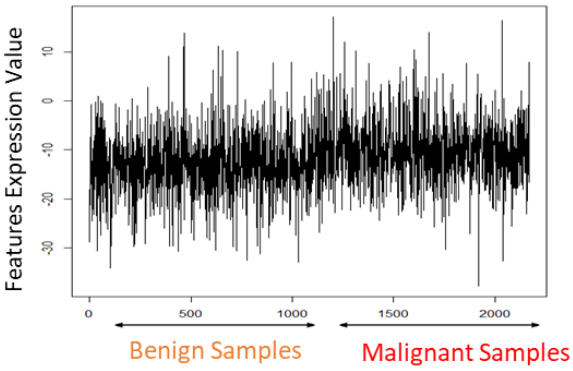
(**d**) ($t = 10.676$, $DF = 2163.5$, p value $< 2.2 \times 10^{-16}$)



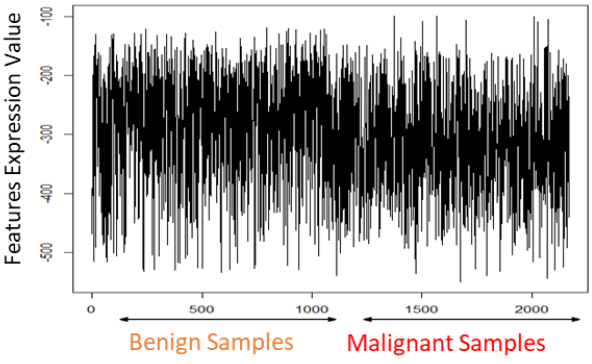
(**e**) ($t = 13.732$, $DF = 2165.5$, p value $< 2.2 \times 10^{-16}$)



(**f**) ($t = -9.0855$, $DF = 2160.2$, p value $< 2.2 \times 10^{-16}$)

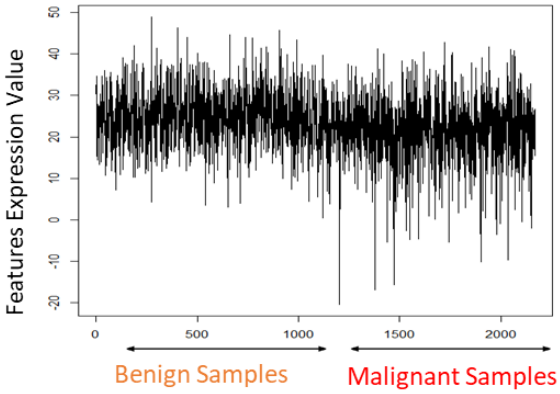


(**g**) ($t = -9.8024$, $DF = 2166$, p value $< 2.2 \times 10^{-16}$)

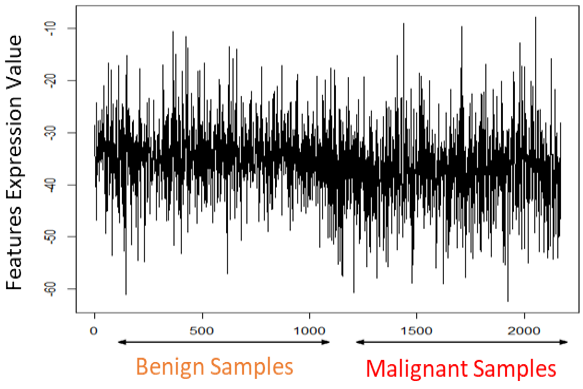


(**h**) ($t = 9.7528$, $DF = 2162.2$, p value $< 2.2 \times 10^{-16}$)

Figure 9. *Cont.*

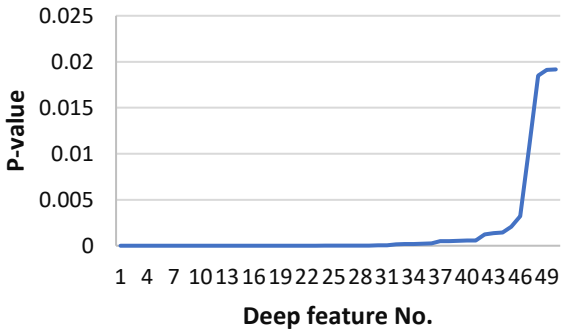


(**i**) ($t = 10.678$, $DF = 2100.4$, p value $< 2.2 \times 10^{-16}$)

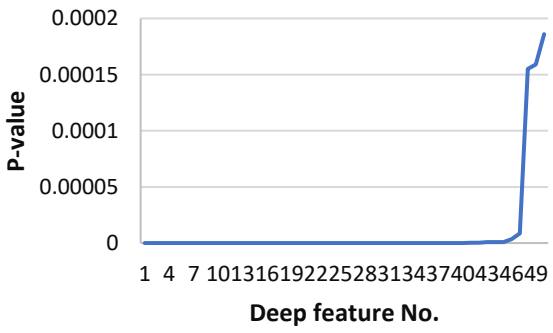


(**j**) ($t = 9.4907$, $DF = 2087.8$, p value $< 2.2 \times 10^{-16}$)

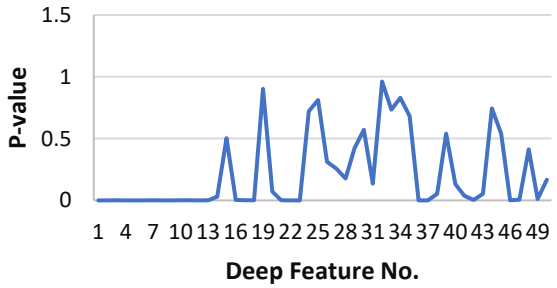
Figure 9. Top ten signal profiles identified key features (i.e., upregulated and downregulated features) using the mutual information FS method and ANOVA F-test results, including *t*-value, degree of freedom (*DF*), and *p*-value. The profiles of the upregulated and the downregulated features are depicted in (**a**–**e**) and (**f**–**j**), respectively. (**a**) Top ranked upregulated feature no. 1. (**b**) Top ranked upregulated feature no. 2. (**c**) Top ranked upregulated feature no. 3. (**d**) Top ranked upregulated feature no. 4. (**e**) Top ranked upregulated feature no. 5. (**f**) Top ranked downregulated feature no. 1. (**g**) Top ranked downregulated feature no. 2. (**h**) Top ranked downregulated feature no. 3. (**i**) Top ranked downregulated feature no. 4. (**j**) Top ranked downregulated feature no. 5.



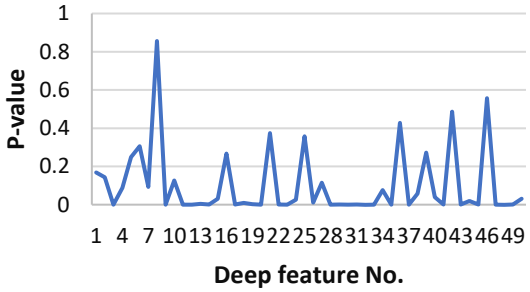
(**a**)



(**b**)



(**c**)



(**d**)

Figure 10. *Cont.*

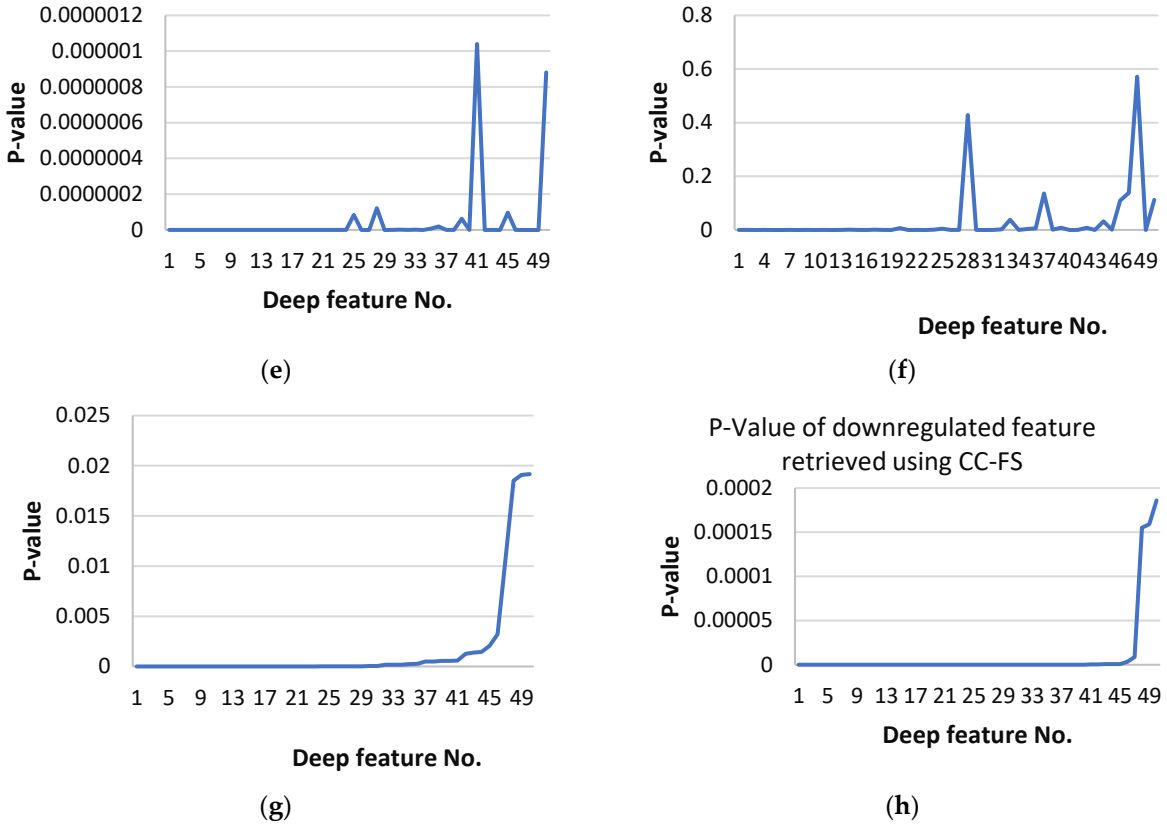


Figure 10. The *p* values of the top 50 upregulated and downregulated features that were retrieved using a univariate-based feature selection approach on the extracted features matrix (EFM) retrieved via the GoogleNet. The *p* values of both the extracted upregulated and the downregulated deep features retrieved from all univariate-based feature selection (PCC, ED, MI, and CC) are shown. (**a**) *p* Value of up-regulated feature retrieved using PCC-FS. (**b**) *p* Value of downregulated feature retrieved using PCC-FS. (**c**) *p* Value of upregulated feature retrieved using ED-FS. (**d**) *p* Value of upregulated feature retrieved using ED-FS. (**e**) *p* Value of upregulated feature retrieved using MI-FS. (**f**) *p* Value of downregulated feature retrieved using MI-FS. (**g**) *p* Value of upregulated feature retrieved using CC-FS. (**h**) *p* Value of downregulated feature retrieved using CC-FS.

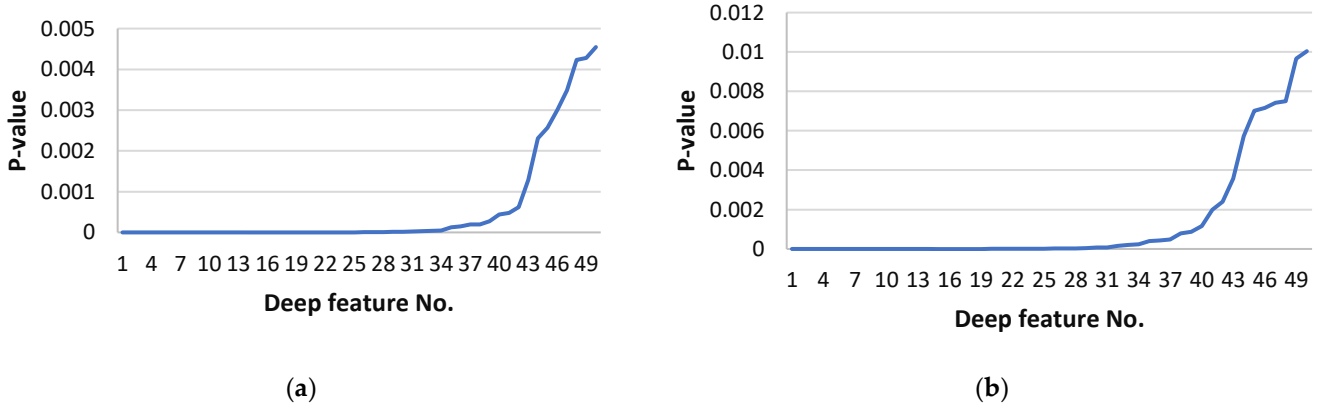


Figure 11. *Cont.*

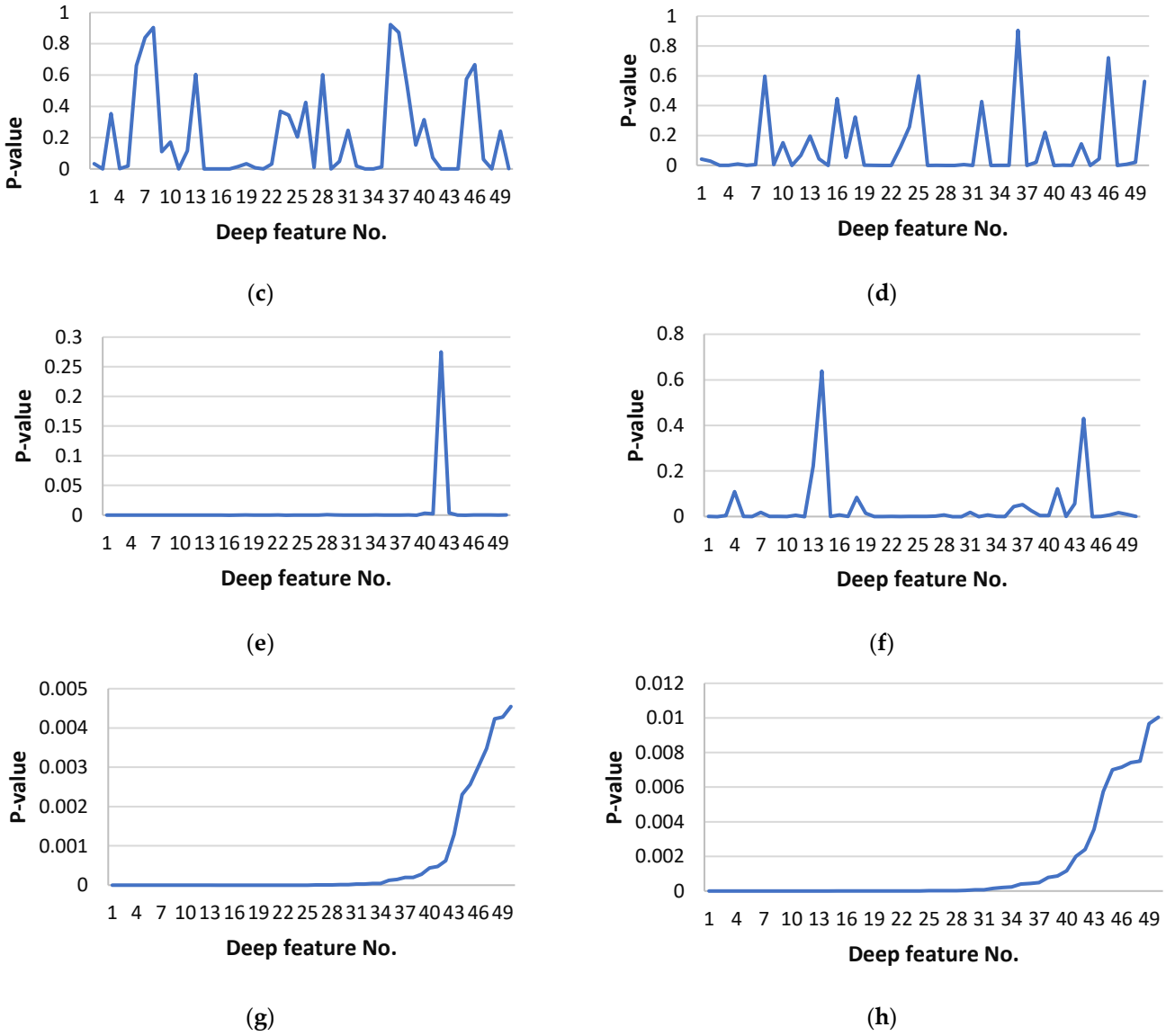


Figure 11. The *p*-values of the top 50 upregulated and downregulated features that were retrieved using a univariate-based feature selection approach on the EFM retrieved via the VGG16. The *p*-values of both the extracted upregulated and the downregulated deep features retrieved from all univariate-based feature selection (PCC, ED, MI, and CC) are shown. (**a**) *p*-value of upregulated feature retrieved using PCC-FS. (**b**) *p*-value of downregulated feature retrieved using PCC-FS. (**c**) *p*-value of upregulated feature retrieved using ED-FS. (**d**) *p*-value of upregulated feature retrieved using ED-FS. (**e**) *p*-value of upregulated feature retrieved using MI-FS. (**f**) *p*-value of downregulated feature retrieved using MI-FS. (**g**) *p*-Value of upregulated feature retrieved using CC-FS. (**h**) *p*-Value of downregulated feature retrieved using CC-FS.

Table 2. Evaluation performance of the different classical classifiers trained on features retrieved using all univariate-based feature selecting techniques on the extracted feature matrix (EFM) that was retrieved from pertained CNN. The ensemble (sub-space KNN) classifier is used for this evaluation.

Pre-Trained CNN for Feature Extraction	Univariate Feature Selection	TP	TN	FP	FN	Accuracy (%)	Sensitivity(%)	Specificity (%)	Precision (%)	False Negative Rate (%)	False Positive Rate (%)	AUC	Matthew's Correlation Coefficient	F1-Score (β = 1)
AlexNet	PCC	1051	1074	10	33	98.00	96.96	99.08	99.06	3.04	0.92	1	0.9605	0.9800
VGG16		1050	1066	18	34	97.60	96.86	98.34	98.31	3.14	1.66	1	0.9521	0.9758
GoogleNet		1017	1042	42	67	95.00	93.82	96.13	96.03	6.18	3.87	0.98	0.8997	0.9491
AlexNet	CC	1050	1072	12	34	97.90	96.86	98.89	98.87	3.14	1.11	0.99	0.9578	0.9786
VGG16		1035	1063	21	49	96.80	95.48	98.06	98.01	4.52	1.94	1	0.9357	0.9673
GoogleNet		1018	1037	47	66	94.80	93.91	95.66	95.59	6.09	4.34	0.99	0.8959	0.9474
AlexNet	ED	763	1004	80	321	81.50	70.39	92.62	90.51	29.61	7.38	0.95	0.6462	0.7919
VGG16		760	1009	75	324	81.60	70.11	93.08	91.02	29.89	6.92	0.94	0.6493	0.7921
GoogleNet		957	1018	66	127	91.10	88.28	93.91	93.55	11.72	6.09	0.98	0.8233	0.9084
AlexNet	MI	1054	1065	19	30	97.70	97.23	98.25	98.23	2.77	1.75	1	0.9548	0.9773
VGG16		1018	1057	27	66	95.70	93.91	97.51	97.42	6.09	2.49	0.99	0.9148	0.9563
GoogleNet		999	1040	44	85	94.00	92.16	95.94	95.78	7.84	4.06	0.98	0.8816	0.9394
AlexNet	Hybrid	1063	1073	11	21	98.50	98.06	98.99	98.98	1.94	1.01	1	0.9705	0.9852
VGG16		1043	1059	25	41	97.00	96.22	97.69	97.66	3.78	2.31	0.99	0.9392	0.9693
GoogleNet		1006	1044	40	78	94.60	92.80	96.31	96.18	7.20	3.69	0.99	0.8917	0.9446

5. Comparing the Performance & Conclusions

Transfer learning (pre-trained deep learning CNN models including AlexNet, VGG16, and GoogleNet) is a technique in which a model trained on one problem is used in some way on another related problem. Transfer learning has the advantage of shortening the training time for CNN models, which reduces computational complexity. Pre-trained CNN, on the other hand, suffers from the feature dimensionality curse (FDC). We present a framework, in this study, that includes a second stage of feature selection to select only the most important key features for precise breast cancer classification. The unconditional mixture model is a univariate filter-based FS method proposed for selecting significant features that can work without taking into account the classifier. As a result, they are extremely efficient in terms of computing. Our introduced CAD system is hypothesized to improve in performance through the application of a second stage of feature selection on the extracted set from the Pre-trained CNN. To test this, we compared the system's efficacy with and without the second stage of feature filtration. Figure 12 depict the accuracy, and sensitivity of the system achieved by using the standalone pre-trained CNNs and the proposed framework. As shown in Figure 12, the proposed framework improves the CAD's performance in terms of accuracy and sensitivity for all types of employed AlexNet, VGG16, and GoogleNet.

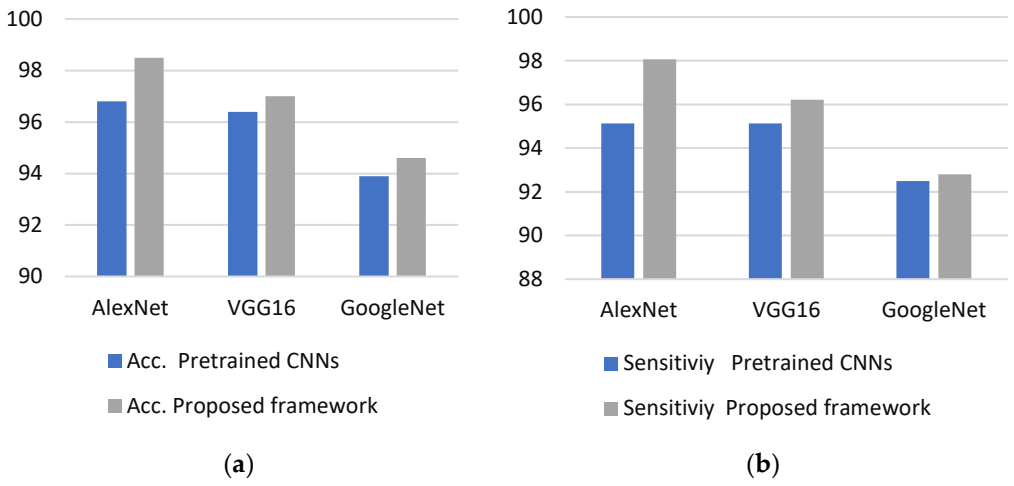


Figure 12. The system's accuracy, and sensitivity achieved by using the standalone pre-trained CNNs and the proposed framework. (**a**) Accuracy. (**b**) Sensitivity.

To explore the impact of input size on the suggested approach's performance, the training time for each CNN model is recorded using both INbreast and mini-Mias datasets. The compression of the execution training time for all deep learning models is depicted in Figure 13.

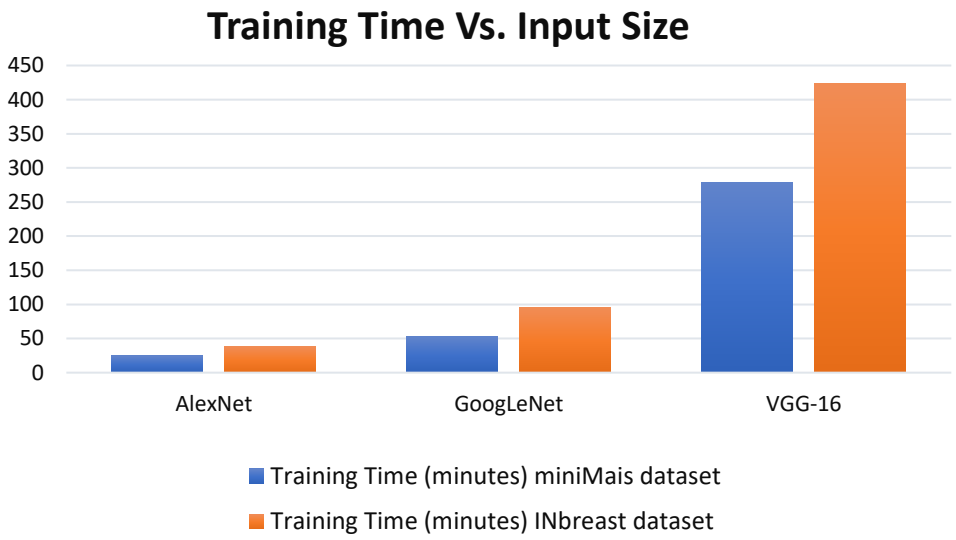


Figure 13. The system's training execution time recorded by using the standalone pre-trained deep learning CNN models of the proposed framework.

The total images (i.e., labelled ROI) used for this study in terms of benign and malignant categories is 576 from mini-MIAS and 1095 from INbreast, respectively. We can conclude that the training process takes a longer time as the size of the input image dataset increases, but this only has a negligible effect on the overall system complexity because it only needs to be done once, before the system is deployed to its final destination. The classification of abnormal breast lesions is one of the more well-known uses of machine learning in medicine. Numerous investigations and attempts have been made to develop an accurate CAD system for such applications, and these studies and efforts have been published as part of a body of work. Table 3 compares the classification performance of the proposed system to that of many advanced breast cancer detection systems. This comparison is conducted in order to evaluate the effectiveness of the newly developed system. Using a standard machine learning technique, the primary objective of these exploratory tests was to develop a method for identifying breast tissue based on texture. Due to their poor precision and sensitivity, the outcomes of these approaches are unsuitable for precise mass classification. Compared to previous work utilizing standard ML techniques [43,61] and deep learning-based approaches [38,45,62–67], the CAD system developed utilizing ROI classification produced promising results. The suggested CAD framework outperformed the system developed by Zhang et al. [64] on the INbreast dataset in terms of accuracy and sensitivity, and the recovered accuracy and sensitivity are comparable with our previously reached values in [18] in comparison to our past work on the INbreast. As long as we have solved the FDC problem, which is a different problem rather than the previous published work in [18]. Although the previous study in [15] and the current one are different and presented to solve two different issues, the final classification evaluation results are comparable and almost the same (around 98%). The slight difference among both results could be due to the randomness of the deep learning model in fine-tuning and optimizing the trainable (weights and biases) during the training process. It is common and known that at each training time of the deep learning models the results might be slightly changed. In terms of overall performance, the proposed CAD system performed well and provided promising and comparable evaluation results as summarized in Table 3.

Table 3. Comparison results against the state-of the-art breast cancer deep learning classification models.

Feature Extraction Approach	Classifier	Dataset	Sensitivity (%)	Accuracy(%)	Reference
Microcalcification detection using digital mammograms	Support-vector machines	INbreast	92	-	[68]
Deep neural networks are used for feature extraction and breast lesion classification.	CNN	INbreast	-	96.8	[62]
End-to-end CAD system for breast mass segmentation and classification	YOLO classifier	INbreast	95.6	89.9	[45,66]
AlexNet, ResNet-18, ResNet-50, and ResNet-10 deep feature fusion	Support-vector machines	CBIS-DDSM miniMIAS	98 99	97.9 97.4	[69]
Transfer learning via GoogleNet, AlexNet, VGG-16	GoogLeNet AlexNet VGG-16	miniMIAS	98.3 98.3 98.7	98.3 98.3 98.3	[43]
Gist, HOG, SIFT, LBP, ResNet, VGG, and DenseNet were used to extract and fuse features.	SVM, XGBoost, Naïve Bayes, k-NN, DT, AdaBoosting	CBIS-DDSM INbreast	98.6 57.2	90.9 87.9	[64]
Transfer learning via Inception-v2, and GoogleNet	XGBoost	DDSM	99.7	92.8	[67]
Deep feature fusion using GoogleNet, VGG-16, VGG-19, and ResNet-50.	Pre-trained CNNs	CBIS-DDSM miniMIAS	98	96.6	[63]
Transfer learning and LR-PCA are used to select features from pseudo colored images.	LR-PCA and Transfer Learning	miniMIAS INbreast	99.60 98.28	98.80 98.60	[18]
Pre-trained CNNs and a Univariate filter-based approach are used to choose features in a cascaded FS architecture.	Cascaded pre-trained CNNs and a univariate filter-based approach	INbreast	98.06	98.50	Proposed Approach

To conclude this study, we proposed a cascaded feature selection framework for selecting nonredundant important features for the classification of breast lesions in digital mammograms. This research builds on our prior work using transfer learning to classify breast cancer. The features are selected using two cascaded FS stages. In the initial stage, pre-trained CNNs such as AlexNet, VGG, and GoogleNet are utilized to extract shallow and deep features in the region of interest in breast mammograms. The usage of the univariate-based FS paradigm in the second stage strengthens the selection of nonredundant relevant characteristics in the returned set of features from the first stage, which can assist overcome the dimensionality curse and multicollinearity difficulties in the retrieved deep features. The features are ordered based on similarities, and the top variables, 20 features, are used to classify breast lesions. The retrieved key features have been analyzed using signal profiles and hypothesis testing (ANOVA F-statistics test). Hence, it has been calculated whether or not there is a significant difference between the median values of benign and malignant samples. The importance of these features for the classification of normal/up-normal breast lesions is ensured by the signal profile in conjunction with the proper *p*-values of each on/off feature. The on/off features were utilized to train six different families of classical classifiers: decision trees, discriminant analysis, SVM, KNN, naive Bayes, and ensemble. There have been fifteen classification experiments. In each experiment, the features extracted from pre-trained CNNs (AlexNet, VGG, and GoogleNet) are fed into one of the univariate filter-based FS methods (PCC, CC, ED, and MI) to determine the degree of similarity between the pre-trained CNN features and the ideal features before they are submitted to the classical classifiers. The best performance was obtained by applying

a hybrid set of features from all univariate FS methods. The performance of our CAD system's proposed framework was compared to that of similar systems reported in the literature, and the results of the comparison revealed that the proposed one outperforms all other systems.

Author Contributions: Conceptualization, N.A.S., G.A., S.M. and Y.M.K.; methodology, Y.M.K. and N.A.S.; software, Y.M.K. and N.A.S.; validation, N.A.S.; formal analysis, N.A.S., G.A., S.M., M.A.A.-a. and Y.M.K.; investigation, N.A.S., G.A., S.M., M.A.A.-a. and Y.M.K.; resources, N.A.S.; data curation, Y.M.K. and M.A.A.-a.; writing—original draft preparation, N.A.S.; writing—review and editing, N.A.S., G.A., S.M., M.A.A.-a. and Y.M.K.; visualization, N.A.S.; supervision, Y.M.K.; project administration, S.M.; funding acquisition, S.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R196), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to acknowledge the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R196), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

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

Abbreviations

We've added the abbreviations for more convenience, listing each acronym with its related term in the article.

CNN	Convolutional Neural Network
FS	Feature Selection
CAD	Computer-Aided Diagnosis
FDC	Feature Dimensionality Curse
PCA	Principal Component Analysis
SVM	Support Vector Machine
CLAHE	Contrast-Limited Adaptive Histogram Equalization
PCC	Pearson Correlation Coefficient
CC	Cosine Coefficient
ED	Euclidean Distances
MI	Mutual Information
EFM	Extracted Feature Matrix

References

1. Breast Cancer. Available online: https://www.who.int/news-room/fact-sheets/detail/breast-cancer (accessed on 15 August 2021).
2. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [CrossRef] [PubMed]
3. Mao, Y.-J.; Lim, H.-J.; Ni, M.; Yan, W.-H.; Wong, D.W.-C.; Cheung, J.C.-W. Breast Tumour Classification Using Ultrasound Elastography with Machine Learning: A Systematic Scoping Review. *Cancers* **2022**, *14*, 367. [CrossRef] [PubMed]
4. Kolb, T.M.; Lichy, J.; Newhouse, J.H. Comparison of the Performance of Screening Mammography, Physical Examination, and Breast US and Evaluation of Factors that Influence Them: An Analysis of 27,825 Patient Evaluations. *Radiology* **2002**, *225*, 165–175. [CrossRef] [PubMed]
5. Cheng, H.; Shi, X.; Min, R.; Hu, L.; Cai, X.; Du, H. Approaches for automated detection and classification of masses in mammograms. *Pattern Recognit.* **2006**, *39*, 646–668. [CrossRef]

6. Atteia, G.; Alhussan, A.A.; Samee, N.A. BO-ALLCNN: Bayesian-Based Optimized CNN for Acute Lymphoblastic Leukemia Detection in Microscopic Blood Smear Images. *Sensors* **2022**, *22*, 5520. [CrossRef]
7. Ayatollahi, A.; Afrakhteh, S.; Soltani, F.; Saleh, E. Sleep apnea detection from ECG signal using deep CNN-based structures. *Evol. Syst.* **2022**, *322*, 1–16. [CrossRef]
8. Custode, L.L.; Mento, F.; Afrakhteh, S.; Tursi, F.; Smargiassi, A.; Inchingolo, R.; Perrone, T.; Demi, L.; Iacca, G. Neuro-symbolic interpretable AI for automatic COVID-19 patient-stratification based on standardised lung ultrasound data. *J. Acoust. Soc. Am.* **2022**, *151*, A112–A113. [CrossRef]
9. Samala, R.K.; Chan, H.; Hadjiiski, L.; Helvie, M.A. Risks of feature leakage and sample size dependencies in deep feature extraction for breast mass classification. *Med. Phys.* **2021**, *48*, 2827–2837. [CrossRef]
10. Taylor, M.E.; Kuhlmann, G.; Stone, P. Transfer Learning and Intelligence: An Argument and Approach. *Front. Artif. Intell. Appl.* **2008**, *171*, 326.
11. Parisi, G.I.; Kemker, R.; Part, J.L.; Kanan, C.; Wermter, S. Continual lifelong learning with neural networks: A review. *Neural Netw.* **2019**, *113*, 54–71. [CrossRef]
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2012**, *25*, 84–90. [CrossRef]
13. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *IEEE Comput. Soc.* **2016**, *2016*, 770–778.
15. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.
16. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 9 December 2016; pp. 2818–2826.
17. Huynh, B.; Drukker, K.; Giger, M. MO-DE-207B-06: Computer-Aided Diagnosis of Breast Ultrasound Images Using Transfer Learning From Deep Convolutional Neural Networks. *Med. Phys.* **2016**, *43*, 3705. [CrossRef]
18. Samee, N.A.; Alhussan, A.A.; Ghoneim, V.F.; Atteia, G.; Alkanhel, R.; Al-Antari, M.A.; Kadah, Y.M. A Hybrid Deep Transfer Learning of CNN-Based LR-PCA for Breast Lesion Diagnosis via Medical Breast Mammograms. *Sensors* **2022**, *22*, 4938. [CrossRef]
19. Samee, N.M.A. Classical and Deep Learning Paradigms for Detection and Validation of Key Genes of Risky Outcomes of HCV. *Algorithms* **2020**, *13*, 73. [CrossRef]
20. Chan, J.Y.-L.; Leow, S.M.H.; Bea, K.T.; Cheng, W.K.; Phoong, S.W.; Hong, Z.-W.; Chen, Y.-L. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics* **2022**, *10*, 1283. [CrossRef]
21. Samee, N.A.; Solouma, N.H.; Kadah, Y.M. Detection of biomarkers for Hepatocellular Carcinoma using a hybrid univariate gene selection methods. *Theor. Biol. Med. Model.* **2012**, *9*, 34. [CrossRef]
22. Mutar, M.; Goyani, M.; Had, A.M.; Mahmood, A.S. Pattern of Presentation of Patients With Breast Cancer in Iraq in 2018: A Cross-Sectional Study. *J. Glob. Oncol.* **2019**, *5*, 1–6. [CrossRef]
23. Coleman, C. Early Detection and Screening for Breast Cancer. *Semin. Oncol. Nurs.* **2017**, *33*, 141–155. [CrossRef]
24. Sardanelli, F.; Fallenberg, E.M.; Clauser, P.; Trimboli, R.M.; Camps-Herrero, J.; Helbich, T.H.; Forrai, G. Mammography: An update of the EUSOBI recommendations on information for women. *Insights Into Imaging* **2016**, *8*, 11–18. [CrossRef]
25. Hickman, S.E.; Woitek, R.; Le, E.P.V.; Im, Y.R.; Luxhøj, C.M.; Aviles-Rivero, A.I.; Baxter, G.C.; MacKay, J.W.; Gilbert, F.J. Machine Learning for Workflow Applications in Screening Mammography: Systematic Review and Meta-Analysis. *Radiology* **2022**, *302*, 88–104. [CrossRef] [PubMed]
26. Yoon, J.H.; Kim, E.-K. Deep Learning-Based Artificial Intelligence for Mammography. *Korean J. Radiol.* **2021**, *22*, 1225–1239. [CrossRef] [PubMed]
27. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
28. Atteia, G.; Samee, N.A.; Hassan, H.Z. DFTSA-Net: Deep Feature Transfer-Based Stacked Autoencoder Network for DME Diagnosis. *Entropy* **2021**, *23*, 1251. [CrossRef]
29. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In *International Conference on Artificial Neural Networks*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2018; pp. 270–279.
30. Yemini, M.; Zigel, Y.; Lederman, D. Detecting Masses in Mammograms Using Convolutional Neural Networks and Transfer Learning. In Proceedings of the 2018 IEEE International Conference on the Science of Electrical Engineering in Israel, ICSEE 2018, Eilet, Israel, 12–14 December 2018; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2019.
31. Hasan, K.; Aleef, T.A.; Roy, S. Automatic Mass Classification in Breast Using Transfer Learning of Deep Convolutional Neural Network and Support Vector Machine. In Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, 5–7 June 2020; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2020; pp. 110–113.

32. Abd-Elsalam, N.M.; Fawzi, S.A.; Kandil, A.H. Comparing Different Pre-Trained Models Based on Transfer Learning Technique in Classifying Mammogram Masses. In Proceedings of the 2020 30th International Conference on Computer Theory and Applications (ICCTA), Alexandria, Egypt, 12–14 December 2020; pp. 54–59.
33. Falconi, L.G.; Perez, M.; Aguilar, W.G. Transfer Learning in Breast Mammogram Abnormalities Classification With Mobilenet and Nasnet. In Proceedings of the 2019 International Conference on Systems, Signals and Image Processing (IWSSIP), Ojisek, Croatia, 5–7 June 2019; pp. 109–114.
34. Falconi, L.; Perez, M.; Aguilar, W.; Conci, A. Transfer Learning and Fine Tuning in Mammogram Bi-Rads Classification. In Proceedings of the Proceedings—IEEE Symposium on Computer-Based Medical Systems, Rochester, MN, USA, 28–30 July 2020; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2020; Volume 2020, pp. 475–480.
35. Samala, R.K.; Chan, H.-P.; Hadjiiski, L.M.; Helvie, M.A.; Cha, K.H.; Richter, C.D. Multi-task transfer learning deep convolutional neural network: Application to computer-aided diagnosis of breast cancer on mammograms. *Phys. Med. Biol.* **2017**, *62*, 8894–8908. [CrossRef]
36. Wimmer, M.; Sluiter, G.; Major, D.; Lenis, D.; Berg, A.; Neubauer, T.; Buhler, K. Multi-Task Fusion for Improving Mammography Screening Data Classification. *IEEE Trans. Med. Imaging* **2021**, *41*, 937–950. [CrossRef]
37. Wang, Z.; Li, M.; Wang, H.; Jiang, H.; Yao, Y.; Zhang, H.; Xin, J. Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion with CNN Deep Features. *IEEE Access* **2019**, *7*, 105146–105158. [CrossRef]
38. 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]
39. Ma, J.; Yuan, Y. Dimension reduction of image deep feature using PCA. *J. Vis. Commun. Image Represent.* **2019**, *63*, 102578. [CrossRef]
40. Alzubaidi, L.; Al-Shamma, O.; Fadhel, M.A.; Farhan, L.; Zhang, J.; Duan, Y. Optimizing the Performance of Breast Cancer Classification by Employing the Same Domain Transfer Learning from Hybrid Deep Convolutional Neural Network Model. *Electronics* **2020**, *9*, 445. [CrossRef]
41. Moreira, I.C.; Amaral, I.; Domingues, I.; Cardoso, A.; Cardoso, M.J.; Cardoso, J.S. INbreast: Toward a Full-field Digital Mammographic Database. *Acad. Radiol.* **2012**, *19*, 236–248. [CrossRef]
42. Kadah, Y.; Farag, A.; Zurada, J.; Badawi, A.; Youssef, A.-B. Classification algorithms for quantitative tissue characterization of diffuse liver disease from ultrasound images. *IEEE Trans. Med. Imaging* **1996**, *15*, 466–478. [CrossRef] [PubMed]
43. Alhussan, A.A.; Samee, N.M.A.; Ghoneim, V.F.; Kadah, Y.M. Evaluating Deep and Statistical Machine Learning Models in the Classification of Breast Cancer from Digital Mammograms. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 10. [CrossRef]
44. Feature Selection in Computer Aided Diagnostic System for Microcalcification Detection in Digital Mammograms. Available online: https://ieeexplore.ieee.org/document/5233466 (accessed on 18 June 2022).
45. Al-Antari, M.A.; Al-Masni, M.; Park, S.-U.; Park, J.; Metwally, M.K.; Kadah, Y.M.; Han, S.-M.; Kim, T.-S. An Automatic Computer-Aided Diagnosis System for Breast Cancer in Digital Mammograms via Deep Belief Network. *J. Med. Biol. Eng.* **2017**, *38*, 443–456. [CrossRef]
46. Yadav, G.; Maheshwari, S.; Agarwal, A. Contrast Limited Adaptive Histogram Equalization Based Enhancement for Real Time Video System. In Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2014, Delhi, India, 24–27 September 2014; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2014; pp. 2392–2397.
47. Fujita, H. AI-based computer-aided diagnosis (AI-CAD): The latest review to read first. *Radiol. Phys. Technol.* **2020**, *13*, 6–19. [CrossRef]
48. Hssayeni, M.D.; Saxena, S.; Ptucha, R.; Savakis, A. Distracted Driver Detection: Deep Learning vs Handcrafted Features. In Proceedings of the IS and T International Symposium on Electronic Imaging Science and Technology, Burlingame, CA, USA, 29 January–2 February 2017; Volume 29, pp. 20–26.
49. Yoo, H.; Kim, H.; Lee, J.L.; Lee, S. Convolution layer with nonlinear kernel of square of subtraction for dark-direction-free recognition of images. *Math. Model. Eng.* **2020**, *6*, 147–159. [CrossRef]
50. Abdelhafiz, D.; Yang, C.; Ammar, R.; Nabavi, S. Deep convolutional neural networks for mammography: Advances, challenges and applications. *BMC Bioinform.* **2019**, *20*, 281. [CrossRef]
51. Blum, A.L.; Rivest, R.L. Training a 3-node neural network is NP-complete. *Neural Netw.* **1992**, *5*, 117–127. [CrossRef]
52. Saeys, Y.; Inza, I.; Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **2007**, *23*, 2507–2517. [CrossRef]
53. Hira, Z.M.; Gillies, D.F. A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Adv. Bioinform.* **2015**, *2015*, 1–13. [CrossRef]
54. Alhussan, A.A.; AlEisa, H.N.; Atteia, G.; Solouma, N.H.; Seoud, R.A.A.A.A.; Ayoub, O.S.; Ghoneim, V.F.; Samee, N.A. ForkJoinPcc Algorithm for Computing the Pcc Matrix in Gene Co-Expression Networks. *Electronics* **2022**, *11*, 1174. [CrossRef]
55. Samee, N.A.; Osman, N.H.; Seoud, R.A.A.A.A. Comparing MapReduce and Spark in Computing the PCC Matrix in Gene Co-expression Networks. *Int. J. Adv. Comput. Sci. Appl.* **2021**, *12*, 2021. [CrossRef]
56. Meyer, P.E.; Lafitte, F.; Bontempi, G. minet: A R/Bioconductor Package for Inferring Large Transcriptional Networks Using Mutual Information. *BMC Bioinform.* **2008**, *9*, 1–10. [CrossRef] [PubMed]

57. Abeel, T.; Helleputte, T.; Van de Peer, Y.; Dupont, P.; Saeys, Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics* **2009**, *26*, 392–398. [CrossRef] [PubMed]
58. Daub, O.C.; Steuer, R.; Selbig, J.; Kloska, S. Estimating mutual information using B-spline functions—An improved similarity measure for analysing gene expression data. *BMC Bioinform.* **2004**, *5*, 118. [CrossRef]
59. Nawaz, A.; Abbas, Y.; Ahmad, T.; Mahmoud, N.F.; Rizwan, A.; Samee, N.A. A Healthcare Paradigm for Deriving Knowledge Using Online Consumers' Feedback. *Healthcare* **2022**, *10*, 1592. [CrossRef]
60. Lu, Z.; Yuan, K.-H. Welch's t Test. In *Encyclopedia of Research Design*; Sage: Thousand Oaks, CA, USA, 2010; pp. 1620–1623.
61. Jian, W.; Sun, X.; Luo, S. Computer-aided diagnosis of breast microcalcifications based on dual-tree complex wavelet transform. *Biomed. Eng. Online* **2012**, *11*, 96. [CrossRef]
62. Xu, J.; Li, C.; Zhou, Y.; Mou, L.; Zheng, H.; Wang, S. Classifying Mammographic Breast Density by Residual Learning. *arXiv* **2018**, arXiv:1809.10241.
63. Khan, H.N.; Shahid, A.R.; Raza, B.; Dar, A.H.; Alquhayz, H. Multi-View Feature Fusion Based Four Views Model for Mammogram Classification Using Convolutional Neural Network. *IEEE Access* **2019**, *7*, 165724–165733. [CrossRef]
64. Zhang, H.; Wu, R.; Yuan, T.; Jiang, Z.; Huang, S.; Wu, J.; Hua, J.; Niu, Z.; Ji, D. DE-Ada*: A novel model for breast mass classification using cross-modal pathological semantic mining and organic integration of multi-feature fusions. *Inf. Sci.* **2020**, *539*, 461–486. [CrossRef]
65. Al-Antari, M.A.; Han, S.-M.; Kim, T.-S. Evaluation of deep learning detection and classification towards computer-aided diagnosis of breast lesions in digital X-ray mammograms. *Comput. Methods Prog. Biomed.* **2020**, *196*, 105584. [CrossRef] [PubMed]
66. Al-Masni, M.A.; Al-Antari, M.A.; Park, J.-M.; Gi, G.; Kim, T.-Y.; Rivera, P.; Valarezo, E.; Choi, M.-T.; Han, S.-M. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Comput. Methods Prog. Biomed.* **2018**, *157*, 85–94. [CrossRef] [PubMed]
67. Song, R.; Li, T.; Wang, Y. Mammographic Classification Based on XGBoost and DCNN With Multi Features. *IEEE Access* **2020**, *8*, 75011–75021. [CrossRef]
68. Liu, X.; Mei, M.; Liu, J.; Hu, W. Microcalcification detection in full-field digital mammograms with PFCM clustering and weighted SVM-based method. *EURASIP J. Adv. Sign. Process.* **2015**, *2015*, 73. [CrossRef]
69. Ragab, D.A.; Sharkas, M.; Attallah, O. Breast Cancer Diagnosis Using an Efficient CAD System Based on Multiple Classifiers. *Diagnostics* **2019**, *9*, 165. [CrossRef]