Integrated IoT-based Healthcare System for the Early Detection of Breast Cancer Using Intelligent Diagnostic System

[1]Shruthishree S H, [2]Harshvardhan Tiwari and [3]Devaraj Verma
[1]Department of Information Science and Engineering, Faculty of Engineering and Technology,
JAIN (Deemed-to-be University), Bangalore, India.
[2]Centre for Incubation, Innovation, Research and Consultancy (CIIRC),
Jyothy Institute of Technology, Bangalore, India.
[3]Department of Computer Science and Engineering, Faculty of Engineering and Technology,
JAIN (Deemed-to-be University), Bangalore, India.
[1]sh.shruthi@jainuniversity.ac.in, [2]tiwari.harshvardhan@gmail.com, [3]c.devaraj@jainuniversity.ac.in

Correspondence should be addressed to Shruthishree S.H: sh.shruthi@jainuniversity.ac.in

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202303004
Received 20 June 2022; Revised form 06 October 2022; Accepted 10 December 2022.
Available online 05 January 2023.

Abstract – Breast cancer represents one of the leading cancer-related diseases worldwide, affecting mostly women after puberty. Even though the illness is fatal and kills thousands of people each year, it is mostly curative if found quickly. As a result, prompt and precise detection methods are critical to patient survival. Previously, doctors used manual detection systems for this objective. However, such techniques have been slow and frequently dependent on the physician's expertise. As technology advanced, these primitive methodologies were supplemented by computer-aided detection (CAD) algorithms. Deep learning is extremely common because of the massive development in large data, the Internet of Things (IoT), linked devices, and high-performance computers using GPUs and TPUs. The Internet of Things (IoT) has advanced recently, and the healthcare industry is benefiting from this growth. Sensors that gather data for required analysis are crucial components utilized in the Internet of Things. Physicians and medical staff will be able to carry out their tasks with ease and intelligence thanks to the Internet of Things. The proposed research focus on integrating Alexnet and ResNet101 for accurate prediction of Breast malignancy from mammogram data. This methodology will target the features more precisely than any other combination of the pre-trained model. Finally, to resolve the computational burden issue, the feature reduction ReliefF methodology is used. To demonstrate the proposed method, an online publicly released set of data of 750 BU images is used. For training and testing the models, the set of data has been further split into 80 and 20% ratios. Following extensive testing and analysis, it was discovered that the DenseNet-201 and MobileNet-v2 trained SVMs to have an accuracy of 98.39 percent for the original and augmented Mammo images online datasets, respectively. This research discovered that the proposed approach is efficient and simple to implement to assist radiographers and physicians in diagnosing breast cancer in females.

Keywords – Internet of Things, Sensors, Convolutional Neural Network, Mammogram, Alexnet, Inception, Robustness, Integration, Malignancy, Intelligent Diagnostic System.

I. INTRODUCTION

WHO states that breast cancer is the most frequent cancer among women. Breast cancer causes the highest number of female cancer-related deaths (570,000 in 2015 alone). Early breast cancer identification is necessary for optimal diagnosis and therapy. Cancer screening is the most frequently applied imaging models for the recognition and diagnosis of breast cancers. Every day, numerous screenings are performed, making it challenging to assess the pictures since a radiologist cannot do it quickly and accurately for thousands of images.

The invention of computer-aided diagnosis (CAD) technology [1] which has healthcare professionals in the early identification of cancer, represents a vital alternative. DL [2] is a relatively recent branch of machine learning. DL has

established an interest in up-to-date years in an extensive range of academic fields, including computer visualization, image analysis, and big data analytics. This strategy was able to obtain a high result across a wide variety of competitions, including the (ILSVRC) [3] is one of the most effective deep learning approaches, achieving exceptional results on demanding tasks including image analysis [4] visual object identification [5] and division [6]. This work aims to implement DCNN to identify breast cancer from mammograms (8000 mammography). We applied two DCNN models in the screening mammography CAD model. To evaluate our approach to large datasets, we compared our simulations to Stacked Autoencoders (SAE)[7].

Considering above stated key inferences, in this paper a highly robust deep hybrid featured machine learning model for breast cancer tissue classification (AlexResNet+) is developed.[8] As the name indicates, our proposed model employs two well-known deep learning methods AlexNet CNN and ResNet as feature extractor to retrieve optimal set of best features for further classification.[9] Retrieving the hybrid features we performed two-class classification using SVM with radial basis function (RBF) kernel. Employing a 10-fold cross validation-based classification our proposed AlexResNet+ model achieves the accuracy of 95.87%, precision 0.9760, sensitivity 1.0, specificity 0.9621, F-Measure 0.9878 and AUC of 0. 960. The overall proposed model is developed using MATLAB 2019b platform where the simulation results with DDSM dataset revealed that the proposed model outperforms major at-hand solutions towards breast cancer tissue classification.

In Section II, we discuss earlier research that classified breast cancer using a DCNN. Such research's goal was to detect and diagnose breast cancer. A DCNN is discussed in Section III. Section IV describes our DCNN model design. Our studies and their findings are described, Section V brings the study to a close by outlining some potential directions for future research.

II. RELATED WORK

The categorization of breast cancer using deep learning will be covered in this presentation. Breast cancer screening involves three stages: identification, evaluation, and final review. Cancer screening pictures are divided into many sorts of regions, such as bulk and calcification, at the first stage. The second stage focuses on removing characteristics from the target area. In the third step, you'll go through all the regions of interest and labeled as malignant or benign. Although there are many different types of characteristics that may be retrieved from the picture, only a small number of effective features can identify benign from malignant tumors, which would be the disadvantage of the old technique. The feature extraction approach, which is quite complex and time-consuming, is used to choose and compute the relevant characteristics.

Numerous research has started to use deep learning for breast categorization because of the significant advancements it has achieved in picture identification,[10] the suggested combination of DL with ML takes out the properties of every mass using a DCNN and then identifies the extraction of attributes using the conventional ML approach. According to [11], mammograms CC and MLO pictures with known calcified and bulk (ROI) information might be used to train the network. The CNN classification is created by identifying attributes from each original map and its related segmentation maps, and it is then assessed using the two open data sources IN breast and DDSM. Over 90% is the ROC. Although this strategy has had great results, it doesn't have the same end-to-end availability or convenience as other ways.

The residual network may classify different levels without user interference. A DCN has also been suggested by [12] as a method for identifying and categorizing malignancies. [13] created a multi-view ResNet to distinguish between certain benign and mammograms utilizing the manual attributes of the mammograms, the automated parameters of classification, and the real picture of the input variables for comparability. ROC for the outcome was 0.80. [14] pre-trained 3 network topologies (CNN model, AlexNet, and GoogLeNet) using ImageNet information and then fine-tuned them based on their statistics (breast ROI). Data pre-processing uses data extension to enlarge the trained dataset. The dependability of training results for the 3 models was 0.6050, 0.8910, and 0.9300, respectively. Healthcare picture recognition and authentication may be done using transfer learning Fine-tuning employs the learned model's characteristics to initialize the networks and achieves a satisfactory result quickly.

III. PROPOSED METHODOLOGY

The overall suggested deep hybrid ML approach for breast cancer cell detection consists of 3 essential stages.

- Stage 1 involves data gathering and augmenting.
- Stage 2 involves extraction of features using AlexResNet101.
- Stage 3 includes feature selection technique.
- Stage 4 involves 2categories of SVM-RBF classification. The following sections covered those consecutive steps to follow in depth.

Data Gathering and Enhancement

To evaluate the efficacy of the suggested breast cancer cell recognition and categorization algorithm, It used a piece of familiar testing information known as the DDSM. Mammography information is gathered and standardized by South Florida University. Original mammary data created with an average dimension of 3000*4800 pixels (size) and a grade of 42 microns with 16 bits is represented by DDSM, which is gathered. There are 2,620 breast scans from mammographic in the DDSM dataset that have been divided into 43 different volumes. The benign and malignant tumors in this data have been distinguished and described by qualified radiographers. Including both benign and malignant lesions and cancers.

For data augmentation, we began by affinely transforming mammography pictures. This was done to eliminate the possibility of any insertion biases during identification or predictions using other morphology methods. Mammography photos were patched to supplement data. However, this method has the effect of picking out portions or fragments of a picture that have a similar structure but come from distinct classes of images. To transfer all microscopic mammogram images into a single space for easier statistical analysis, we normalized the quantity of bulk or stain data. Random color extensions were applied to each annotated picture (benign and malignant). Each original picture was down sampled to the 1024*768 pixel size for our suggested task. We produced 150*75-pixel crops using down-sampled images. Every mammography picture was described by 20 crops, with the crops being further encrypted into 20 descriptions when it was determined that the data acquired was sufficient. The collection of descriptions was then combined using 3-norm pooling to create a single description. We may extract the single descriptor mathematically by using equation (1).

$$C_{\text{pool}} = \left(\frac{1}{N} \sum_{i=1}^{N} (d_i)^p \right)^{\frac{1}{p}} \tag{1}$$

We set the value of p - 3 under. In this case, N represents the overall quantity of crops, and the description of each crop is shown as c_i and c_{pool} refers to the collective description for every mammogram. It's noteworthy that the p norm of vectors allows for the median value of p to be 1 and the greatest occurrence to be p to be ∞. The original mammogram picture produces a huge number of descriptors that assist construct an ideal collection of characteristics for categorization. After data preprocessing, we extracted deep features using AlexNet CNN and ResNet50. The next parts provide an in-depth discussion.

AlexResNet101 + Extraction of Feature

In this research, our primary emphasis was on combining the deep features that were collected from the AlexNet CNN and the ResNet50deep networks. In this case, our main goal is to extract and employ a combination of deep (AlexNet within 4096 kernels) and varied intensity of data (residuals ResNet50) to conduct more accurate and effective breast cancer detection. The following information is provided on these models of deep learning.

AlexNet is the first CNN framework that outperformed DL methods for item recognition and categorization. AlexNet CNNs were created to conduct diverse item categorization using the pre-trained model, its resilience allows it to be utilized as a transferrable learning approach for breast cancer bulk extraction of image features. Classical CNN gets 256-D, but we got 4096-D at the FC levels. This gives us extra data to make improved decisions In this suggested approach AlexNet CNNs architecture, we used 5 convolutional levels CONV [1,2,3,4,5] and 2 FC levels (FC-6 and FC-7). Of the three potential fully interconnected layers (FC-6, FC-7, and FC-8), FC-8 stood out because of its 1024-D properties. Conversely, FC-6 and FC-7 levels possessed 4096- D aspects, Which is greater than FC-8, and as a result, only FC-6 and FC-7 characteristics were taken into consideration for further categorization [6,7].

CONV

The convolutional layer, often known as CONV, is a combination of two separate filters (vertical and horizontal filters) that may extract and incorporate feature patterns from input pictures. There are [CONV (1)-96 CONV(2)-256, CONV(3)-384 ,CONV(4)-386 and CONV(5)-256] kernels among the neurons or kernel specifications at the CONV layer [28,29]. In the current issue of breast cancer screening mammography image features extraction, every neuron extracted an image representation that had the same set of weights (Wi) and bias (bi). These ranges were aimed at neurons in feature mapping identifying related features. CONV with various neurons (**Fig 1**) allows varying bias and weight levels to extract local characteristics. The final feature vector is retrieved as an output from the CONV layer, which filters the mammographic (augmented) pictures used as input. Using different neurons, zero-padding of 2, as well as a step of 4, we were able to produce successive features.

The 1st level of DL node was fed in the specified design as 224 * 224 size with 96 kernels (size 11 * 11 and stride 4 pic). 96 kernels equaled 96 image streams. The outcome of the first level was passed to the second layer neurons after local the response normalizing and max pooling. 256* 5*5*96 kernel filters were used in the second level. The third, fourth, and

fifth levels are not connected by a normalization layer. The third layer 384 kernel size is 3*3*256, the 4th level has 3*3*384 kernels.5 C ONV layers, two FC layers, 4096-D kernels.

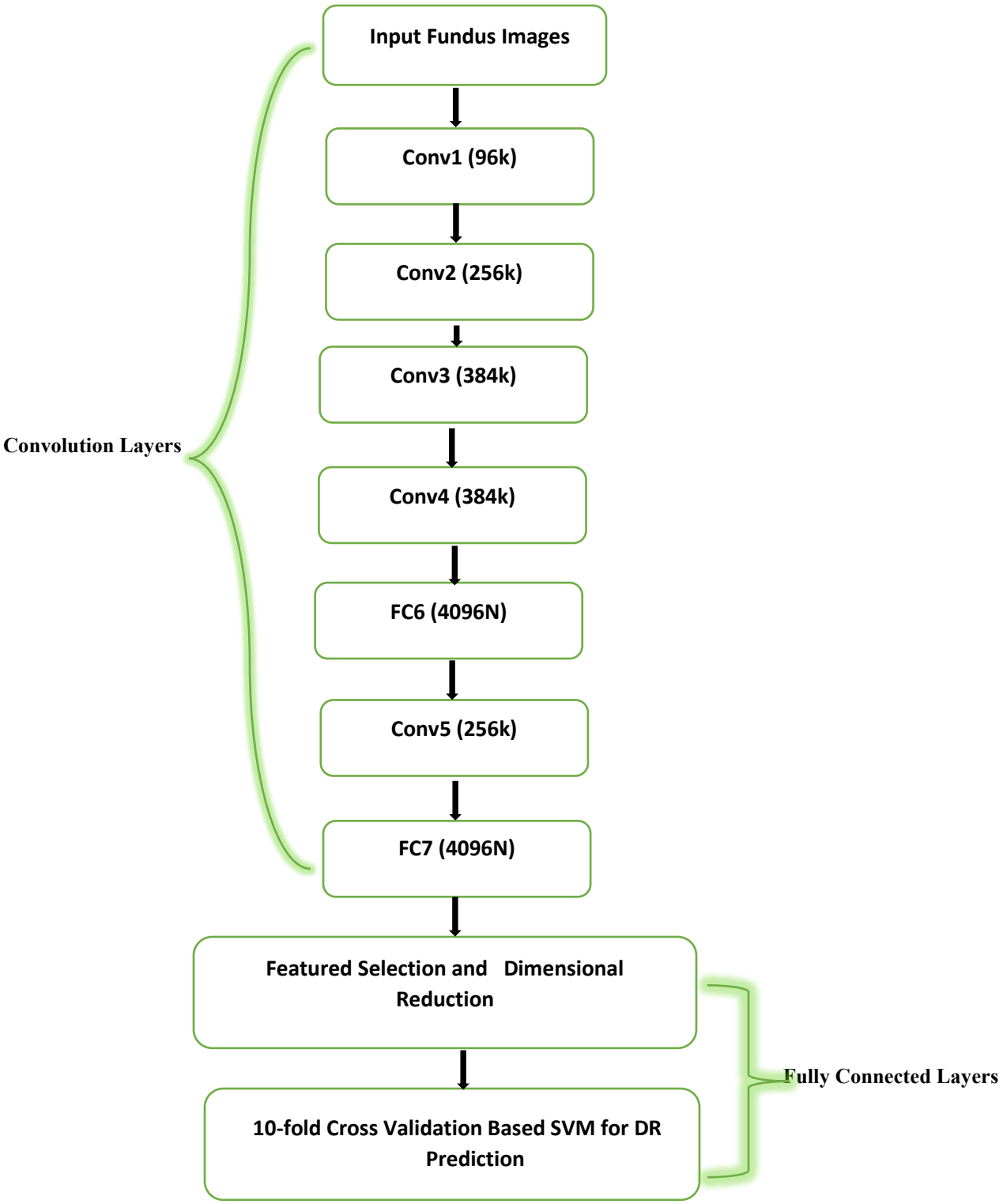

Fig 1. Convolution Neural Network architecture on AlexNet

Layer of Max-Pooling

In our suggested approach, we used Max-Pooling to progressively lower the positioning accuracy of each CONV feature map. The pooling layer reduces the variables and computations. Local averages and a subsampling approach are used to accomplish it. Additionally, it aids in preventing the over-fitting issue [30]. To get the translation-invariant interpretations from the input information, Max-pooling was used. It down sampled the feature representations using a continuous component and the greatest non-overlapping sub-region frequency. Max-pooling evaluates sparseness by removing non-maximal quantity in non-overlapping subdomains and upgrading the feature extractor to avoid irrelevant solutions. Producing dense latent encoding decreases the number of filters required to decrypt every pixel for restorations. Consequently, our proposed model is much more efficient in terms of computing. After every CONV layer, we used a 3x3 Max-pooling layer.

Layer of ReLU

Normalization ReLU was included as a convolution operation. A non-linear component operator that behaves as a level is included in the ReLU layer. Our model includes three ReLU layers. ReLU recovers the neuron's outputs from the inputs y. q(x) as x if x>0 and (δ×x) if x<=0. Specifically, if the negative elements can be avoided by conducting multiplied with a slope of 0.01 or by setting it to 0, the result will indicate as δ. To make our model work as ReLU function q(x) =max(0, x), we activated that zero barrier and provided δ=0.

In AlexNet-traditional CNN's architecture, there are 8 layers total, including 5 CNN layers and fully - connected levels. **Fig 1** provides an overview of the AlexNet CNN that has been suggested **Table 1**.

Table 1. Classifier-Centric Performance Assessment with AlexResNet+ Feature

Deep Features	Performance with AlexResNet+Features				
	Accuracy (%)	Precision	Recall (Sensitivity)	F-Measure	Specificity
SVM-Linear	92.31	0.9411	1.000	0.9696	0.8235
SVM-Polynomial	95.44	0.9565	1.000	0.9777	0.9412
SVM-RBF	95.87	0.9760	1.000	0.9878	0.9621

In our suggested approach, enhanced pictures were given directly to 96-neuron to AlexNet. In this case, every CONVs level produced unique attributes that underwent feature scaling and mean reduction. The results were then resized and sent to the next layers after undergoing additional processing.

Layer of FC

In this case, FC levels function as the final level(s) of the AlexNet CNNs and perform in-depth processing. We use the final visual features for the SVM classifier rather than the FC layer, which is the classification step in conventional AlexNet. Technically, this layer translates the previous layer's neuron, the CONV layer's vectors, to every connected neuron. As a result, a 1D feature representation is created, which may then be utilized to further categorize data. We employed (FC6 and FC7) levels, which supplied 4096-dimension features, for any further categorization. We fused FC6 characteristics. The final one-dimensional characteristic obtained using AlexNet CNN became known as FeatAlexNet consequently.

ResNet50

The residual network is often referred to as ResNet. ResNet deep model uses reconstructed CONV levels to learn residue functions. Unlike standard deep learning techniques, residue networks, notably ResNet, may be improved to recover more variety and depth of information (features) [14]. In the usual application, a RB is attached for every CONV inside the format of an "impedance connectivity" to accomplish identity mapping. CONV output is combined with the bypass branching result, and the resultant is sent to the following block. ResNet's network design developed from VGGNet, except for the usage of the aforementioned "shorted connection." Each CONV in this situation contains little kernel elements of size 3*3. We respect the following guidelines for design:

- Guideline 1: To achieve identical output ports, the layer must have the same quantity of filters.
- Guideline 2: Reducing an image representation by 50% doubles the quantity of filters, making it analytically more effective and less time-consuming at every level.
- Guideline 3: Depending on the level size, the ranges from 34 to 152, the ResNet architecture(s) may have varying depths.

In our ResNet deep learning model, mammograms were used to obtain high-dimensional features. Pre-activation enhanced ResNet50, known for its ability in "training by understanding" using residual variables, was utilized in our network. We built the ResNet model with several training algorithms so that it could analyze numerous mammographic mass pictures simultaneously for feature extraction and learning. Here, ResNet50 extracted various characteristics from various perspectives or enhanced input photos.

It resized or modified pictures to fit 150753 pixels. Since being training on breast mammogram characteristics, ResNet50 showed multi-layer morphological operations and transmission. We retrained the inputs by inserting a residue block denoted as to maintain greater performance and quick convergence as shown equation (2).

$$x = F(y\{W_i\}) + y \tag{2}$$

In Eqn (2), y and x are the input and output matrices of the recursive feature. The residual mapping data between the input pictures to be learned is stated in this case by F(x, W i). The operational framework of residue pattern development is shown in **Fig 2**. As shown, ResNet50 was used as a shortcut addition to solve the loss function without the requirement for additional elements or increased computational work.

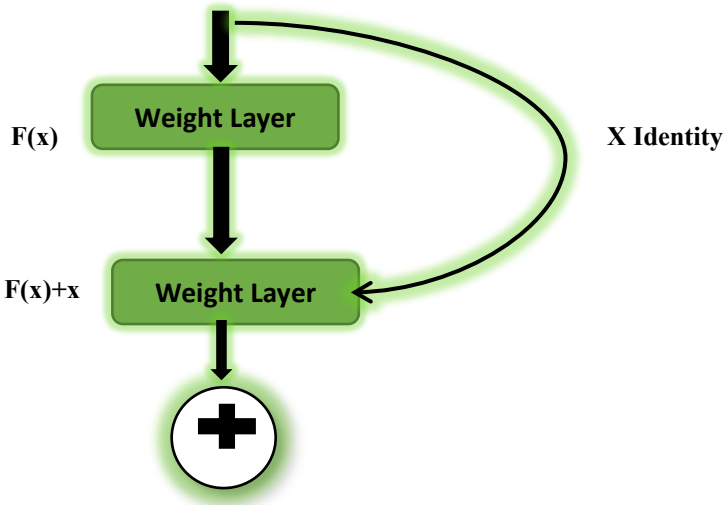

Fig 2. Block-of-identity

Our suggested ResNet model batch-normalized input pictures. As the first 3 layers, we included Batch Normalization. It normalized each input channel (we used three) throughout a mini batch. To enhance training speed while lowering initiation sensitivities, we applied Batch Normalization among CONV and non-linearity. By splitting the mini-batch standard error and removing the mini-batch median, it was possible to normalize the action potentials of every channel. The input was then moved by a trainable offset value and scaled by a trainable weighting factor.

Therefore, we redesigned ResNet50 to keep important properties with less processing. We only retrieved the final deep features FeatResNet to be utilized for two-class classifier using the SVM-RBF ML process, even though standard deep learning methods like AlexNet and ResNet50 employ a final dense network with Softmax function to accomplish categorization. The complete implementation design of the ResNet50 model that was employed with concealed units can be seen in Figure 3. We employed "residual blocks" for each stacked layer, as shown above, to finally recover multi-layered characteristics for use in further categorization.

Feature Fusion

We combined the characteristics from ResNet50 FeatResNet and AlexNet CNN after collecting them to create a 1D feature representation. In the model we suggested, both sets of features were added together before categorization. Thus, we were able to (3). Feat $_{\text{DeepHybrid}}$ = conc [Feat $_{\text{AlexNet}}$, Feat $_{\text{ResNet}}$] The following section provides details of the SVM-RBF model.

Two-Class Categorization Using SVMs

It is used to classify the patterns in machine learning. It is ideal for classification applications such as text classification, detection and tracking, image analysis, and other classification-related tasks based on computation time and resilience. SVM adapts from input data and acts as a non-probabilistic binary classification. Categorize inputs and decreases overfitting over

unseen instances via systemic reduction risk. Here, the SVM is a portion of the learning database that returns the border line ranges between two classes with distinct characteristics or patterns, known as hyper-places. To accomplish two-class categorization, we used the equation (3).

$$x' = wi * \phi(y) + bi \tag{3}$$

The non-linear transformation is stated in $\phi(Y)$ (3), where its focus is on determining the right weight wi and bias factor bi. By incrementally lowering the correlation of equation (4) in (3), x' is predicted.

$$R_{reg}(x') = C * \sum_{i=0}^{l} \gamma(x_i' - x_i) + \frac{1}{2} * \| wi \|^2 \tag{4}$$

In (4), C specifies a penalty factor and, a target value γ Using, the weighted values are determined in the equation (5).

$$wi = \sum_{j=1}^{l} (\alpha 1_j - \alpha 1_j^*)\phi(y_j) \tag{5}$$

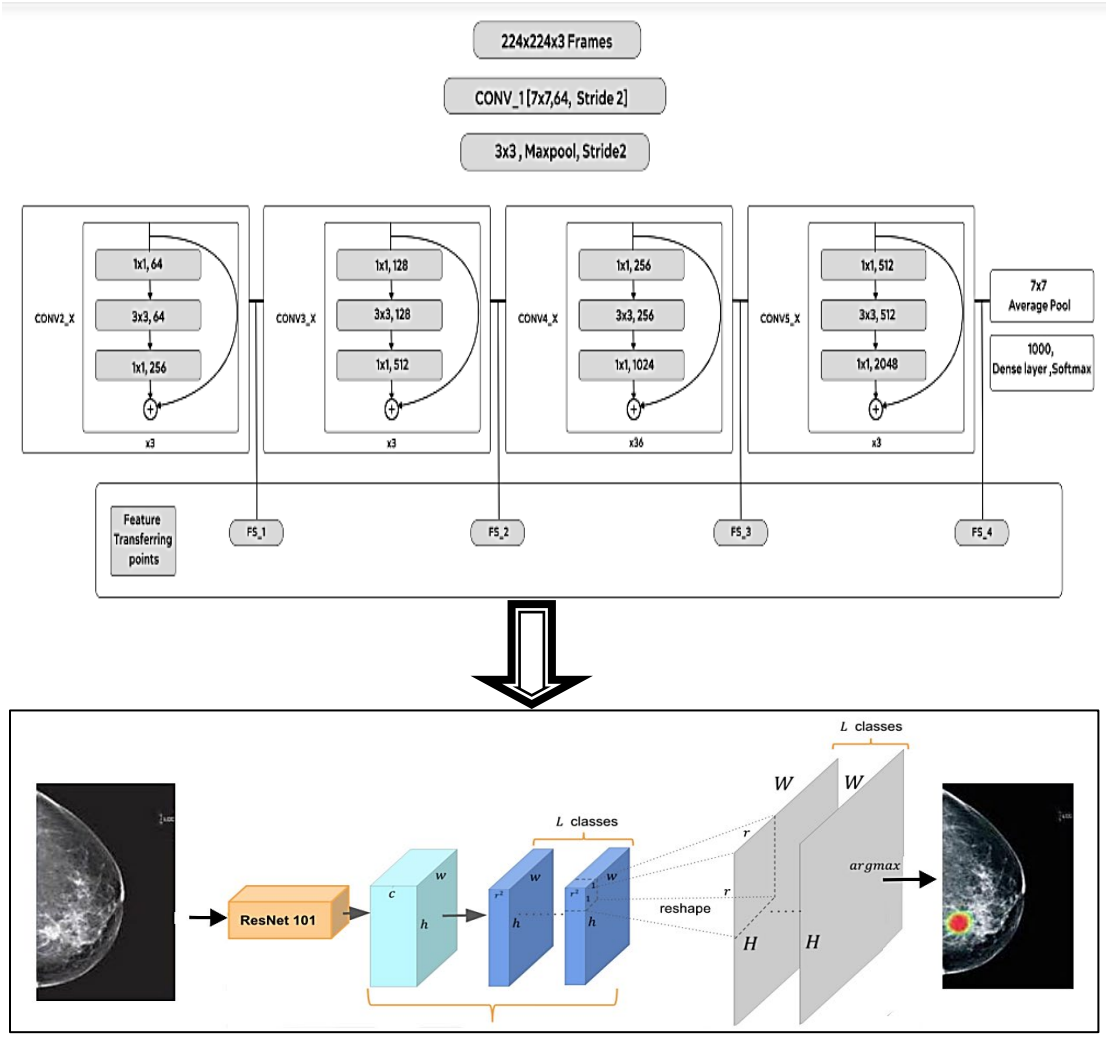

Fig 3. Architecture of ResNet50's Connections and Multi-Layered Extraction of Features Extraction

In (7),$k(_{Xj,\ x)}$ the kernel parameter is shown. There are generally three main kernel factors: linear, polynomials, and RBF. We used SVM with several basic functions, including linear, polynomials, and RBF, in our suggested model. We tested for randomized test cases or pictures while learning over the retrieved characteristics. Each test input or chest mammogram picture was evaluated as benign or malignant by SVM throughout this procedure. Following this overview, we will examine the overall simulation findings and related implications. The Architecture of ResNet50's connections and multi-layered extraction of features extraction is shown in **Fig 3**.

In the following equation, $\alpha 1$ and $\alpha 1^{*}$represent Lagrangian multiplier, which is constantly non-zero. The SVM's ultimate output will be as shown in equation (6).

$$\begin{aligned} x' &= \sum_{j=1}^{l} \left(\alpha 1_j - \alpha 1_j^*\right)\phi\left(y_j\right) * \phi(y) + bi \\ &= \sum_{j=1}^{l} \left(\alpha 1_j - \alpha 1_j^*\right) * K\left(y_j, y\right) + bi \end{aligned} \tag{6}$$

The (6) $\phi(y)$ is the non-linear transformation state, with a focus on finding the appropriate weight wi and bias value bi. X' is calculated in equation (7) by lowering the correlation.

$$R_{reg}\,(X') = C * \sum_{i=0}^{l} \gamma(X_i' - X_i) + \frac{1}{2} * \| wi \|^2 \tag{7}$$

ReliefF Techniques

ReliefF can evaluate the attribute quality in classification situations with high correlation. They have a broader perspective through using local knowledge acquired from various settings. ReliefF classifies features by weight and age, with the most active appearing first. As shown in Figure 4, other factors have a virtually lesser impact on the final. Therefore, we may choose the ten most effective features based on their weighted age and exclude variables that add unnecessary computation time to the model [6–8].

Original ReliefF uses two classes to handle noisy and defective data. It is used to determine each feature's accessibility, which may then be employed to choose features with the highest feature selection ratings. ReliefF similarly selects an element T randomly (Stage 3), after that, k scans the neighborhood for membership of the same class, also referred to as the closest hit results of H. We utilize the number of immediate neighbors as three indicate a real number scalar (Stage 4) as well as an integer value.

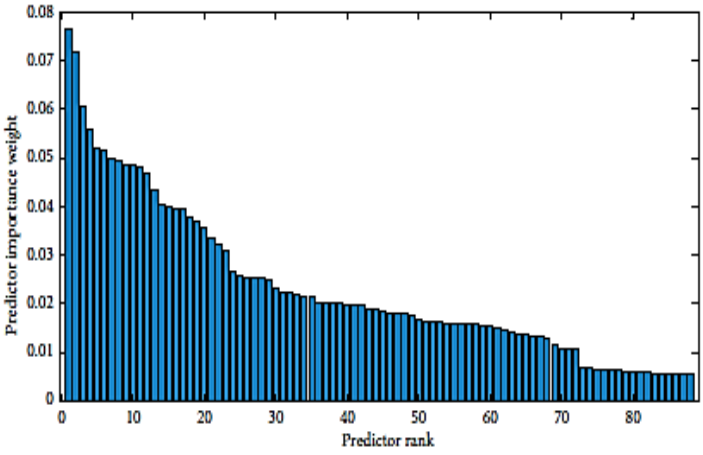

Fig 4. Ranking Attributes by Relevance (of Weight)

In this way, k-closest neighbors provide the result of one to each class, also known as closest misses N(T) values (Stages 5 and 6). It also changes the V[B] class estimation for every feature based on, hits H, T and misses N(T) (Stages 7 and 8). The RelieffF process is shown in **Algorithm 1**.

Algorithm 1. ReliefF Algorithm Process

Input data: feature and value of Class
Outcome data: a list of evaluations for the features' building blocks.
Stage 1: weight V[B]: = 0.0;
Stage 2: For j:=1 to m do begin
Stage 3: Select an instance ri randomly;
Stage 4: Find the k-nearest hits hj;
Stage 5: For every class C class(ri) do
Stage 6: From class C find k Closet misses nj (c);
Stage 7: For B: =1 to b
Stage 8: V[B] =V[B] − diff(B, ri, hj)(n, k)kj =1+
Stage 9: [(d)1 − p(class ri diff(b, rihj))]kj = 1(n, k)D≠ class ri
Stage 10: End all

Resnet101 Description

ResNet_101 system utilizes residual blocks, through which gradient can pass directly, to prevent grades from becoming zero following chain constraint applications. Total convolution layer in ResNet-101 are 104. Additionally, it is made up of 33 layers total blocks, 29 of which directly use the output of the previous block, which is what is meant by residual blocks above. These residual links are then used as the first operand of the summation provider that is applied at the end of every block to obtain the input for the subsequent blocks. The output of the preceding block is used in the remaining four blocks' convolution, which has a filter size of 1x1 and an amble of 1, then a batch normalization layer, which accomplishes the normalization operation, and finally the summing up operator, whose output is transmitted to the preceding block's output. Varying dense block depths occurs in the architecture.

IV. RESULT ANALYSIS

Standard Simulation Results

We may assess the reliability of the component ranking factors using measures like Spearman's correlation factor (rho). The average of the 7(7-1)/2 bilateral similarity for every method has been calculated. The value is displayed in Table 1, in which it can be observed that SVM-RFE is the more reliable ranking technique (0.4750), whereas LR-Wrapper is rather unreliable (0.0310).

Table 2. Stability of a seven-Rank Set Measured by Average Bilateral Similarity Using Spearman's Ranking Correlation Factor (). Best Results in Grey

ρ	**Pearson**	**Relief**	**RF**	**LR-Wrapper**	**SVM -RFE**	**SVM-Wrapper**
	0.0180	0.0530	0.4120	0.0300	0.4750	0.0660

One such method we've used is the Jaccard indices, which are useful for analyzing the consistency of the top-k feature lists inside a smaller feature set. Table 2 shows the Jaccard value for feature subsets ranging from 10 to 124 as well as the mean in the final row. These findings enable us to conclude that the embedding techniques are more stable than the wrapped approaches, which are less so. RF provides the most reliable classifier; however, its performance is lower than other techniques.

If the study is based on one measure, we have no means of determining how to compare the ranks offered by the various algorithms. Firstly, which models offer comparable predictions, and secondly, which predictor is more reliable throughout a variety of k values that should be comparable. Table 3's outcomes are difficult to comprehend.

Table 3. Using the Median Bilateral Similarity with the Jaccard Value for Various Values of K, The Stability of A Collection of 7 top-k Lists is Evaluated, The Ideal Outcome on A Grey Backdrop.

K terms	Pearson	Relief	RF	LR-Wrapper	SVM -RFE	SVM-Wrapper
10	0.2570	0.1890	0.3300	0.1930	0.1800	0.1930
20	0.2030	0.2480	0.5560	0.2150	0.2810	0.2150
30	0.2280	0.2480	0.4410	0.2270	0.3510	0.2270
40	0.3290	0.2730	0.4660	0.2730	0.3980	0.2730
47	0.3750	0.3170	0.4800	0.3210	0.4420	0.3210
50	0.3980	0.3450	0.5970	0.3430	0.4610	0.3440
60	0.4670	0.4290	0.7910	0.4020	0.5120	0.4020
70	0.5780	0.4960	0.8150	0.4830	0.5850	0.4830
80	0.7030	0.5660	0.9460	0.5420	0.6390	0.5420
90	0.7680	0.6360	0.8620	0.6230	0.6920	0.6230
100	0.8120	0.7120	0.9460	0.7250	0.8010	0.7250
110	0.8510	0.8130	0.8760	0.8230	0.8880	0.8230
120	0.9430	0.9400	0.9440	0.9460	0.9580	0.9460
124	1	1	1	1	1	1
Mean for K [1 to 124]	0.550	0.4750	0.6550	0.4810	0.5420	0.4720

Simulation Results in Graphics

Fig 5 (a) illustrates how the steady growth varies depending on the value of k. It seems that the RF and Pearson techniques provide the most consistent performance, but SVM-stability RFEs are quite poor for limited values of k. The most significant characteristics cannot be extracted only the techniques can be executed once. An excellent choice is to combine the rankings to get a more accurate one. Wrapper methods are also quite erratic. To compare feature selections, we employ MDS [1, 2]. As a result of using six different techniques and launching every method 7 times, the tests we conducted may be described as a collection of 46 data. This defines a 124-dimensional space. The MDS is used to project these locations in two dimensions. The distance between the locations is determined by Spearman's rank correlation, and the pressure criteria are normalized by the total squared divergences.

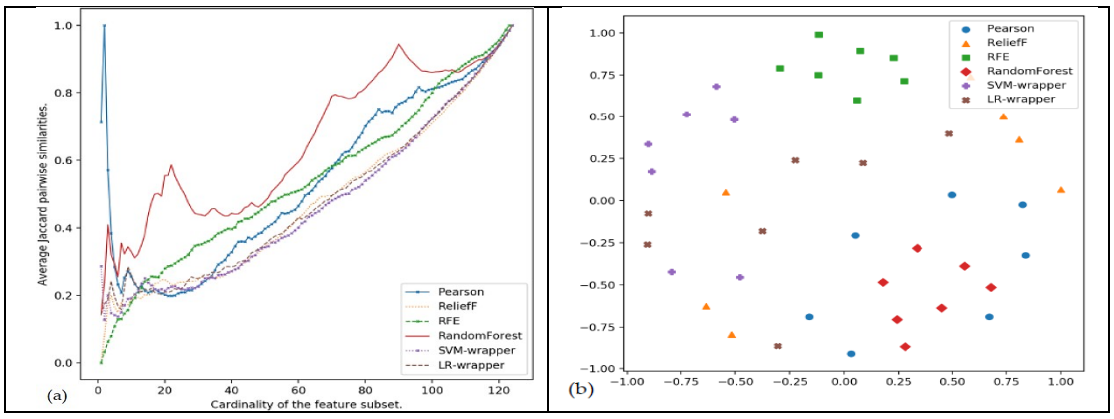

Fig 5. Illustration of feature selector reliability: (a) Jaccard value for Extracted Features (b) MDS plot of the Characteristic

This two-dimensional representation is then used to provide a pair of positions to every other algorithmic output (x, y). The similarity between feature selectors may be observed in Figure 5(b). Regarding stability, it can be noted that the values from SVM-wrapper, SVM-RFE, and RF are more tightly grouped compared to those from Pearson, ReliefF, and LR-wrapper. It is important to look at the techniques' stability together with their ability to predict the right class. This is important because experts need data on the most significant risk factors and protective factors rather than just information on the more trustworthy methods.

V. CONCLUSION

This Research project is to create a computational framework for classifying mammogram pictures using the MIAS databases. It is crucial for early identification of breast cancer to improve patient survival rates or lower patient mortality rates. This paper establishes the computational paradigm for automated breast cancer detection utilizing MIAS mammogram processing. Each procedure, which consists of multiple steps, is carried out correctly. Pre-processing steps are applied to the original noisy mammogram pictures. This research develops a unique hybrid depth feature-based machine learning approach for breast cancer tissue detection and classification. Unlike traditional techniques, the suggested model combines DL and ML techniques to optimize effectiveness. The proposed method uses two well-known deep learning models, AlexNet and ResNet50, to extract deep features. This study proposed that the purposeful combination of deep features might aid with more accurate acquisition of breast cancer tissue pattern learning and consequent classification. In this relationship, the different characteristics were taken from AlexNet and shorted ResNet50, the former of which was built with 5 convolutional layers and 3 fully connected layers and the latter as a simulation study. The fused characteristic was learned over SVM using RBF kernel function to classify benign or malignant mammography images. In this paper, resilience is evaluated using both a scalar measure and a graphic technique. Using an MDS projection, we can observe that SVM-RFE and Randomized Forests are the most secure techniques, whereas Pearson, Relief, SVM-wrapper, and LR-wrapped are unstable. In spite of the improved results, further work may be done to use more effective machine learning algorithms in the future to improve precision.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author(s) declare(s) that they have no conflicts of interest

References

[1]. J. Tang, R. M. Rangayyan, J. Xu, I. E. Naqa and Y. Yang, “Computer Aided Detection and Diagnosis of Breast Cancer with Mammography: Recent Advances”, IEEE Journal of and Health Informatics 13 (2): 236251, 2009.
[2]. G. E. Hinton, S. Osindero and Y. W. The “A fast learning algorithm for deep belief nets.” Journal of Neural Computation 18 (7): 15271554, 2006.
[3]. Basheer, S., Anbarasi, M., Sakshi, D.G. and Vinoth Kumar V., “Efficient text summarization method for blind people using text mining techniques”, Int J Speech Technol, 23, 713–725 (2020). Doi.10.1007/s10772-020-09712-z
[4]. Olga Russakovsky, Jia Deng, Hao Su, “ImageNet Large Scale Visual Recognition Challenge”, International Journal of Computer Vision 115 (3): 211252, 2005.
[5]. Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, “Gradient based learning applied to document recognition”, Proceedings of the IEEE 86 (11), 1998.
[6]. Kouser, R. R., Manikandan, T., & Kumar, V. V., “Heart Disease Prediction System Using Artificial Neural Network, Radial Basis Function and Case Based Reasoning”, Journal of Computational and Theoretical Nanoscience, 15(9), 2810–2817, 2018. Doi.10.1166/jctn.2018.7543
[7]. M. Radovic, O. Adarkwa, and Q.Wang, “Object Recognition in Aerial Images Using Convolutional Neural Networks”, Journal of Imaging 3 (2), 21, 2017.
[8]. J. Long, E. Shelhamer and T. Darrell, “Fully convolutional networks for semantic segmentation”, IEEE Journal of Transactions on Pattern Analysis and Machine Intelligence 39 (4), 2017.
[9]. Ahmed, S. T., Kumar, V. V., Singh, K. K., Singh, A., Muthukumaran, V., & Gupta, D., “6G enabled federated learning for secure IoMT resource recommendation and propagation analysis. Computers and Electrical Engineering, 102, 108210, 2022. Doi.10.1016/j.compeleceng.2022.108210
[10]. T.Kooi., “Large scale deep learning for computer aided detection of mammographic lesions”, Med. Image Anal. 35, 303–312 (2017).
[11]. A. Akselrod-Ballin, “A region based convolutional network for tumor detection and classification in breast mammography”, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, pp. 197–205, 2016.
[12]. Sarah S. Aboutalid, “Deep Learning to Distinguish Recalled but Benign Mammography Images in Breast Cancer Screening”. In: Clinical Cancer Research, 1-11, 2018.
[13]. K. Kira and L. A. Rendell, “A practical approach to feature selection,” in Machine Learning Proceedings, pp. 249–256, 1992.
[14]. R. J. Urbanowicz, M. Meeker, W. La Cava, R. S. Olson, and J. H. Moore, “Relief-based feature selection: introduction and review,” Journal of Biomedical Informatics, vol. 85, pp. 189–203, 2018.