Novel Algorithms for Early Cancer Diagnosis Using Transfer Learning with MobileNetV2 in Thermal Images

Swapna Davies[1] and Jaison Jacob[2]
[1,2] Department of Electronics and Communication Engineering,
Rajagiri School of Engineering and Technology, Kochi, India
Affiliated to APJ Abdul Kalam Technological University, Thiruvananthapuram, India
[E-mail : swapnad@rajagiritech.edu.in, jaison_jacob@rajagiritech.edu.in]
*Corresponding author: Swapna Davies

Abstract

Breast cancer ranks among the most prevalent forms of malignancy and foremost cause of death by cancer worldwide. It is not preventable. Early and precise detection is the only remedy for lowering the rate of mortality and improving the probability of survival for victims. In contrast to present procedures, thermography aids in the early diagnosis of cancer and thereby saves lives. But the accuracy experiences detrimental impact by low sensitivity for small and deep tumours and the subjectivity by physicians in interpreting the images. Employing deep learning approaches for cancer detection can enhance the efficacy. This study explored the utilization of thermography in early identification of breast cancer with the use of a publicly released dataset known as the DMR-IR dataset. For this purpose, we employed a novel approach that entails the utilization of a pre-trained MobileNetV2 model and fine tuning it through transfer learning techniques. We created three models using MobileNetV2: one was a baseline transfer learning model with weights trained from ImageNet dataset, the second was a fine-tuned model with an adaptive learning rate, and the third utilized early stopping with callbacks during fine-tuning. The results showed that the proposed methods achieved average accuracy rates of 85.15%, 95.19%, and 98.69%, respectively, with various performance indicators such as precision, sensitivity and specificity also being investigated.

1. Introduction

Breast cancer (BC) is mainly the fatal breast disease in females globally. Early and correct diagnosis is crucial for reducing disease burden, lowering morbidity and mortality rates, avoiding disfiguring surgery and improving the overall survival index by up to 95% [1]. If the tumour size is smaller than 10 mm, the chances of complete cure are 85% [2]. According to studies [3], different radiologists interpret mammograms differently at various points of observation. The interpretation error does not justify exclusive reliance on mammography as the only assessing method. As a result, researchers advise incorporating thermal imaging alongside breast inspection and mammography to monitor breasts, as mammography produces false negatives (FN) in the preliminary phases.

Thermography is an adjunct technique authorized by the United States Food and Drug Administration to assist in the secure, economical, and precise recognition of abnormalities in breast tissue [1] that has demonstrated the ability to identify and track thermal anomalies [4]. Moreover, thermography has certain benefits: ionizing radiation is not used in imaging, non-invasive and painless procedure making it better comfortable, and its equipment is easier to transport than mammography machines; the technology can help as a complimentary tool to boost sensitivity and accuracy, as well as make a preliminary selection of patients based on abnormal skin temperature to have mammography [5, 6].

The human body, like other objects, emits infrared signals under normal conditions [7]. The presence of a breast tumour enhances the temperature of the surrounding tissues due to high metabolic activity of cancerous cells and increased blood flow as tumour promotes generation of new blood vessels [8, 9]. Specialists commonly use asymmetric studies of healthy and sick breasts.

Deborah A. Kennedy et al. [10] described a method to identify BC that combines mammography and thermography. Their findings revealed that thermogram-based discernment only provided 84% sensitivity, mammography-based detection provided 91% sensitivity, and coupled mammogram and thermogram detection provided 95% sensitivity. Rafal Okuniewski et al. [11] reported a method to classify contours observed in thermographic breast images, captured by the Braster gadget to determine the presence of cancerous tumours. Four classifiers were used to categorise (SVM, Naive Bayes, RF and DT), and based on the outcomes, the RF classifier yielded the most favorable results.

Acharya et al. [12] presented a thermogram-based approach for detecting BC early. For each image, they extracted statistical parameters such as energy, mean, homogeneity and entropy. The SVM classifier is employed to identify tumours in photos. They employed 50 thermogram photos in their investigations (25 abnormal and 25 normal). Their findings revealed that the SVM classifier had an accuracy of 87.0%, a sensitivity of 86.82%, and a specificity of 91.52%. These findings, however, were produced with relatively little data. As a result, the findings cannot be generalised. Rajagopal et al. [13] identified BC in thermal images through an analysis of the asymmetry between the right and left breasts. They developed an automated segmentation approach that utilized the Projector Profiles Methodology to separate the right and left breasts. With minor adjustments, including standardizing the background, height, and eliminating noise, this technique holds promise for broader applicability. Sheeja V. Francis et al. [14] used a curvelet transform-based approach to extract statistical and textural information from breast thermograms. All characteristics were provided to SVM for automatic classification, with 90.91% accuracy.

Earlier, automated systems for analyzing thermograms rely on manually crafted features

that are given into a machine learning (ML) algorithm for classification [4, 15, 16, 17]. In the realm of AI, deep learning (DL) models have surpassed human performance in picture categorization since 2015 [18]. Consequently, these DL models show significant potential for the automated categorization of BC thermograms. CNN models with several layers are used in DL [19, 20]. During training of a neural network, the internal variables (synaptic weights) among its layers are iteratively adjusted to minimize the disparity between the output of the system and the intended outcome. The training procedure needs a multitude of labelled samples in order to fine-tune the network's internal parameters, enabling it to accurately predict and classify inputs [21]. Zhang, Pan, Chen, Wang [22] introduced a refined 9 layer CNN for BC identification and compared three activation functions: ReLU, leaky ReLU and parametric ReLU (PReLU). Also compared six distinct pooling techniques: average, max, stochastic, rank-based average, rank-based weighted and stochastic rank pooling, and concluded that stochastic rank pooling along with PReLU exhibited the best results with a sensitivity 93.4%, specificity 94.6% and an accuracy of 94.0%. Zhang et al. [23] utilized an 8-layer CNN employing batch normalization, dropout and stochastic rank pooling, and integrated with a graph convolutional network to analyze breast mammograms through a 14-way data augmentation approach. The achieved performance includes a sensitivity of 96.2±2.9%, specificity of 96.0±2.31%, and an accuracy of 96.1±1.6%.

Ekici and Jawzal [8] created algorithms to detect BC in breast thermograms, employing a CNN optimized by Bayes algorithm in their methodology. To address class imbalance, data augmentation techniques were applied to ensure equal representation of both cases. The accuracy of non-optimized CNN was determined to be 96.78%; however, upon optimising CNN settings using the Bayes algorithm, the accuracy increased to 97.6%. In a recent study by Sanchez-Cauce et al. [24], a creative way to identify breast cancer was introduced. Their approach integrates diverse perspectives of breast thermograms alongside clinical and personal data. The study implemented a multi-input CNN network tailored for different thermogram views. The inclusion of clinical data in these structures yielded impressive results, achieving 97% accuracy, 83% sensitivity, and 0.99 AUC-ROC.

Mambou SJ [25] proposed the utilization of Inception V3, the deep CNN model. They used the DMR database, which had 1062 thermal pictures, 75% of which were used for training and 25% for validation. For classification, they introduced SVM at the end of the CNN model and the learning rate is set to 10^{-4} and the epochs to 15. Although the researchers did not specify the accuracy rate, they did mention the proportion of illness detection. J. Wang, Khan, S. Wang, Zhang [26] employed SqueezeNet, incorporating fire modules and a complex bypass mechanism, to capture significant characteristics from mammography images. These features are subsequently harnessed for training an SVM. This model demonstrated noteworthy outcomes, achieving an accuracy of 94.1% and a sensitivity of 94.3%. To validate the robustness of the results, a 10-fold cross validation procedure was executed and the mean and standard deviation of several performance metrics were computed across numerous iterations.

According to the literature review on BC detection [27], several distinct technologies rooted in various modalities, such as mammography, ultrasound, MRI, thermography and others, are frequently explored for BC diagnosis. Among these, thermography stands out as a potentially impactful tool in the early stages of examining BC patients [1], often yields favourable results over others. Advancing beyond traditional ML, subsequent efforts have focused on the usage of DL approaches [28]. They use quantitative analytics to more accurately diagnose suspected lesions [29]. Transfer Learning (TL) approaches are commonly utilised for thermograms to solve the aforementioned difficulties. We use deep learning model on massive datasets and refine them with limited data sets in TL [30].

In this paper, we focus on the role of thermography in early cancer diagnosis, provide our experimental findings on the use of transfer learning to diagnose breast cancer. We employ a MobileNetV2 [31] based approach that is efficient in terms of computational cost and memory. MobileNetV2 is an efficient architecture specifically crafted for mobile devices and embedded systems. We also demonstrate a comparison with other networks employed in mobile applications such as MobileNetV1, NasNetMobile and ShuffleNet. Additionally, we conducted a comparative analysis with contemporary techniques, which encompass the Extreme Learning Machine, Bag of Features with SVM and the stacked Autoencoder method. We also use strategies such as image pre-processing, augmentation of data, and the selection of key parameters (e.g. learning rate, batch size) to reduce over-fitting and increase the generalisation of TL technique for thermograms.

The subsequent sections of the paper is arranged as: Section 2 gives a review of thermography. Section 3 discusses about the methodology with dataset description and evaluation parameters. Section 4 provides the outcomes of the study and juxtaposes them with the use of cutting-edge networks in mobile applications as well as various contemporary techniques. Section 5 furnishes the conclusion.

2. Review of Thermography

Many innovations have been made to increase patient survival and faithful diagnosis of BC. Mammography is an imaging method that can identify growth of cancer around the 12th month when the tumour is larger than 1 cm in diameter and X-ray can flow through it, which has already metastasized in many cases. Mammograms can be painful and stressful because procedures frequently require compressing the breast tissue between two plates in order to improve contrast between non-cancerous and malignant cells [2].

To identify malignancies, ultrasound use sound waves. Because it does not use radiation, it is the best recommended approach for examining younger and pregnant women with thick breasts. It can tell the difference among solid masses and cysts. But it cannot distinguish malignancies in distal sites or detect micro-calcifications. The efficacy of ultra-sonography is determined by the physician evaluating the images [2]. It is typically used in association with mammography to pinpoint the precise area of suspicion [1].

Breast MRI is a non-invasive scanning method that employs a substantial 1.5T magnetic field to produce high-quality breast pictures [32]. It can detect even the smallest lesions that the other two methods cannot. However, it is an expensive exam. Because it commonly reports false positive (FP) diagnosis, its PPV is limited. It is unable to identify micro-calcifications. This test is not suggested for pregnant women because it uses a potent magnet and a contrast agent, which might cause allergic reactions [2]. Furthermore, limitations such as x-rays, expense, dense tissues at an early age, FP and FN rate motivated specialists and organizations to conduct substantial investigation into alternative approaches like thermography. Thermography can detect alterations in breast tissue temperature before the formation of a mass or tumor. This can potentially allow for earlier detection of abnormalities. Mammograms can be less effective in breasts with dense tissue because the tissue can mask abnormalities. Thermography may be more effective in such cases as it detects variations in temperature, not tissue density [33,34]. **Fig. 1** depicts photos of sick and healthy breasts.
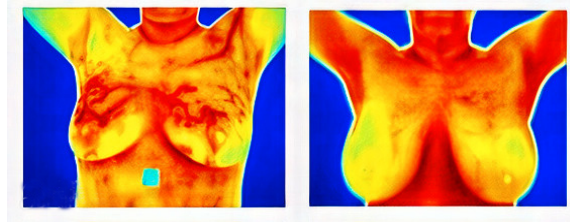
Fig. 1. Images of unhealthy and healthy breasts

2.1 Procedure for Breast Thermography

The human body generates some of its own thermal energy as IR radiation. This is the basic principle to use thermogram as a BC diagnostic method. Cancer tissues metabolise quicker than other tissues, and the heat generated in this process is transferred to the skin surface, indicating a probable malignancy or thermally active, rapidly growing tumour [35]. Furthermore, there is an enhanced provision of oxygen and nutrients to the tumour due to excessive regional vasodilation generated by nitric oxide emanating from the malignant lesion [36].

The thermal camera creates an image of the temperature change over the skin's surface. The temperature changes that characterise tissue metabolism are periodic [37] (approximately 24 hrs). Non-circadian rhythms associated with cancer cells are symptomatic of abnormality. Women with thermograms showing asymmetry are ten times more likely to develop breast cancer than those with thermograms exhibiting symmetry [38]. Thermal imaging of the breast was captured utilizing a FLIR SC-620 thermal camera. **Fig. 2** depicts the technique and the equipment utilised.
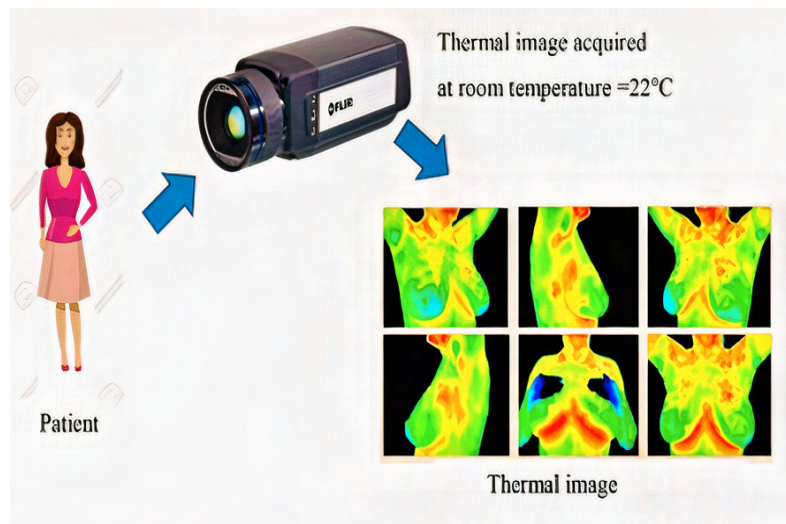

Fig. 2. Breast cancer diagnosis with thermography

2.2 Pre-Thermographic Monitoring and Patient Acclimatization

The patient is chilled before the scan to reveal hotspots caused by anomalies. The breast should not come in connection with any surfaces that have the potential to change their warmth.

Patients are recommended to avoid using ointments or scents, drinking more tea or coffee than usual, and exposing or treating their breasts. To achieve consistent findings, a protocol must be followed during thermography. During an examination, incandescent lighting should not be utilised since it emits radiation. The disrobed patient must sit for   10-20 minutes at rest in order to attain temperature equilibrium of the portions to be inspected [1]. After reaching a thermal stable phase, the subject is positioned in front of the IR camera, hands lifted just above head, and the afflicted breast is examined via three angles: frontal, medial, and lateral. **Fig. 3** depicts a patient's breast thermogram in various settings.
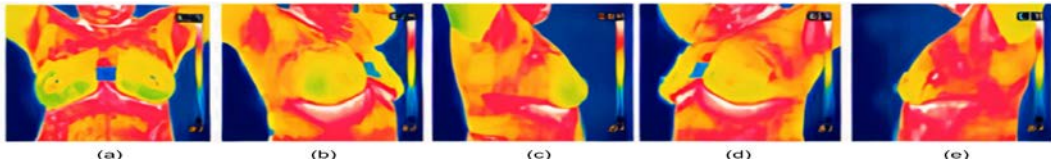

Fig. 3. Thermogram samples: Positions (a) Frontal, (b) 45° Right Lateral, (c) 90° Right Lateral, (d) 45° Left Lateral, (e) and 90° Left Lateral [39]

3. Methodology

The conceptual design of the presented technique for detecting breast cancer with deep CNN and thermal imaging is explained in depth in this section. A pre-trained CNN structure is employed to detect cancer from thermography image.

3.1 DL Model and Training

As computational power has dramatically increased and large new data sets are created on a regular basis, DL, a subset of ML, has become more popular in recent generations. Because of their capacity to run on the graphics processing unit, they have surpassed several classic ML algorithms in terms of modelling large data sets. CNN's main premise is that incoming data can be processed as images. This can result in fewer variables being used, resulting in quicker computation. Depending on the image, type of data, and intention, various study frameworks have been designed and implemented. The model training phase is time-consuming in image processing investigations using DL techniques. Furthermore, building a DL model necessitates a vast amount of data as well as processing time. To train and improve use case training, it is frequently beneficial to employ pre-trained network on large dataset that span day or week. To save training time, a pre-trained network such as MobileNet is used in this study.

3.1.1 MobileNet

ResNet and VGG architectures are generally large in size or demand a large number of mathematical computations, while achieving very high levels of accuracies in the ImageNet dataset, for several reasons such as deep architecture, feature hierarchy, high capacity, residual connections (ResNet) and regularization techniques. A deep CNN model is incredibly tough to train with limited medical data. Google in 2017 proposed MobileNet, a DL-based framework for mobile device that is analytically efficient, small in size, and delivers increased accuracy.
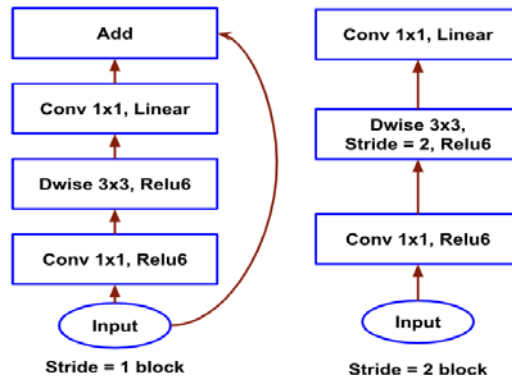
Fig. 4. Architecture of MobileNetV2

MobileNet is a deep CNN trained on ImageNet dataset of millions of photos, with the goal of performing finely on handheld devices. It has a distinct structure than conventional CNN structures, including links among the bottleneck layer. The architecture consists of 32 initial convolution layers subsequent to 19 bottleneck layers. **Fig. 4** depicts the MobileNetV2 model's features. The design has some benefits over previous DL systems. Training the model on limited datasets is difficult, and the perceptual classification process becomes prone to over-fitting. MobileNet's architecture and design, including depthwise separable convolutions and reduced model size allow it to have a smaller parameter count, which prevents over-fitting and leads to lower consumption of memory with a narrow error margin. Furthermore, this approach allows quick predictions due to less computational complexity while modularity in structure make it adaptable for experimentation and parameter optimisation [40].

3.1.2 Transfer Learning and Keras

Keras is a Python-centric advanced neural net API that can operate on top of Tensor Flow. Since the model requires less parameters to run, it is user-friendly and simple to apply. In Keras, there are two approaches to create a model: functional and sequential. Sequential models are built from layers in a linear stack. The process of learning is set in the sequential models, and three arguments are used: the Adam optimiser, the categorical cross-entropy loss function, and the accuracy measure. Then the data is fed into the training phase.

Building a CNN system from beginning takes an inordinate amount of time and is computationally expensive [41]. As a result, using a pre-learned CNN models on datasets is a common strategy. Transfer learning is implemented typically using a pre-trained model. A pre-trained framework is one that has undergone training using a huge quantity of data to resolve problems comparable to the one that is presently being solved. TL is applying this acquired knowledge. There are several pre-trained architectures accessible when using Keras API, including MobileNet, Xception, InceptionV3, VGG16, DenseNet, VGG19, and so on. Instead of starting from scratch, the available pre-trained model can be utilized. They must be updated by fine tuning it and retraining all or some of the layers with specified approaches. As a consequence, for the new categorization task, a different architectural approach is used.

3.1.3 Fine-tuning

Fine tuning is the most widely employed approach for TL. It entails keeping the knowledge gained from resolving one classification challenge and employing it to address a similar issue using a different approach. Latest research has emphasized the usage of pre-trained networks

rather than training from scratch when addressing classification problems. Pre-trained CNN architectures are designed, tested, and trained on extensive datasets that are more diverse than the available thermographic dataset. This makes these networks highly capable. In the fine-tuning process, the frontal layers of a pre-trained CNN are frozen, while concurrently training the last layers and the newly added classification layers. Typically the final layers of most CNNs are more specialized than the initial ones. Fine-tuning tries to alter these specialised features so that they operate with the recent information instead of overwriting the general information.

The use of certain massive network on mobile devices strains storage and processing capacity, lowering performance. To tackle this issue, we present a TL strategy [42] that uses a MobileNet-based framework with minimal parameters. As a consequence, they consume less storage capacity and produce better and rapid results than larger models. Two output classes are added into the last layers of a pre-trained MobileNetV2. Subsequently the output stages are fine-tuned, whilst the network's initial layers are frozen. Throughout the training phase, the model's progress can be monitored by observing the learnt features at various levels of MobileNet. It provides the flexibility to halt the initial training and lower the learning parameter size.

3.2 Proposed Methodology

We developed a deep learning-based solution that combines the MobileNet network to simplify and increase the thermography equipment accuracy. MobileNet is a CNN structure that has proven to be particularly effective and fast in biomedical segmentation when compared to other approaches. The suggested system is unique in that it uses the MobileNet network to automate the pre-processing and builds a DL model that uses the MobileNet output to categorize the provided thermogram (**Fig. 5**).
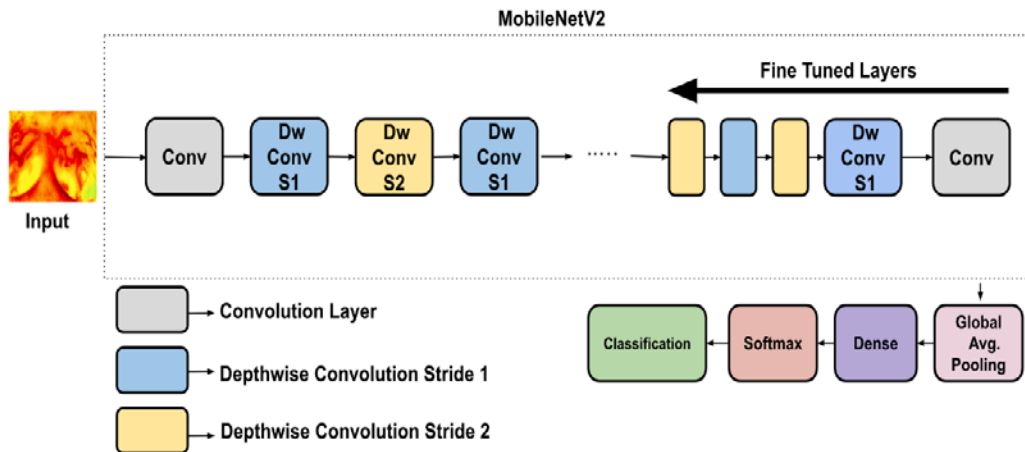

Fig. 5. Proposed Framework for breast cancer detection

In MobileNetV2, there are two types of blocks. The one with stride 1 is the residual block, while the other with stride 2 is the non-residual block, which is utilised for down-sizing [43]. The classification layer is one of the 154 layers in MobileNetV2. We utilised this model to transfer data from a previously completed related activity. Our suggested framework comprises of 154 pre-trained network layers, in addition to 2 extra layers, one at the start for pre-processing and another at the end for classification task.

Our method combines the CNN training steps, such as augmenting of data, TL, and fine-tuning, into a unified model using the MobileNetV2 architecture. To increase performance of the model, we presented one traditional and two novel fine-tuning approaches. We further employed data augmentation techniques including horizontal flipping and rotation to maintain the test data unique. Our concept is that during network training, the final layers are taught more, increasing the capability to generalise to the novel challenge whilst concealing the knowledge learnt from previous training of the initial layers, resulting in more effective outcome. Three methodologies were employed in this study to compare theoretical execution as listed below. The models were trained over a period of 15 epochs.

1. Baseline Model (Model 1)

The initial strategy was conventional, freezing the base model from MobileNetV2 by making it non-trainable. Transfer learning is performed to optimize the training process, using features extracted from ImageNet Dataset.

2. Fine Tuned Model (Model 2)

The second method does the fine-tuning, the final 35 layers of MobileNetV2 are opened and the remaining 119 layers being frozen. A stepping algorithm fine-tunes the model, re-training the model on the final layers with a very low learning rate. Adapting the learning rate to traverse these layers in smaller steps, can result in capturing finer details and higher accuracy.

3. Fine Tuned Model with early stopping (Final Proposed Model or Model 3)

The fine-tuning process significantly improves the model performance. To even do better, one can train longer and in that case, to ensure that the model does not train for more epochs than necessary, we have to implement callbacks for early stopping. The third technique used a specified progressive mathematical algorithm by early stopping, saving model as the best fine-tuned model on monitoring validation loss to be minimum. Early stopping is performed to avoid unwanted iterations and obtain optimized result. We also implement callbacks for model checkpoint in order to preserve the optimum model when found. This allows to resume training from the optimal weights for example in the event of unforeseen incidents causing a kernel crash. We continuously monitor the validation loss and once it reaches a point where there are no more improvements, i.e. it cannot be further minimized, the training process is halted. Model checkpoint monitor validation loss and preserve the optimum model from all epochs.

We conducted experiments using 3 different scenarios and analysed the numeric outcomes and suggested that when using raw data without extensive preprocessing steps, the proposed technique delivers promising results.

3.3 Dataset Description

The dataset used in this research was taken from the Database for Mastology Research (DMR) [44]. This is a breast thermogram database that is accessible online. In the literature, the DMR data set was the most widely utilised data set. The dataset contains breast thermal images from a total of 56 individuals, of 640×480 pixels. 37 of the 56 participants are sick, while 19 are healthy [45, 46]. The age of the participants is not mentioned in the DMR dataset. The information was obtained from http://visual.ic.uff.br/dmi/. Breasts of diverse forms and sizes can be found in these photographs. Images are transformed to grey scale before being pre-processed and categorised using a neural network.
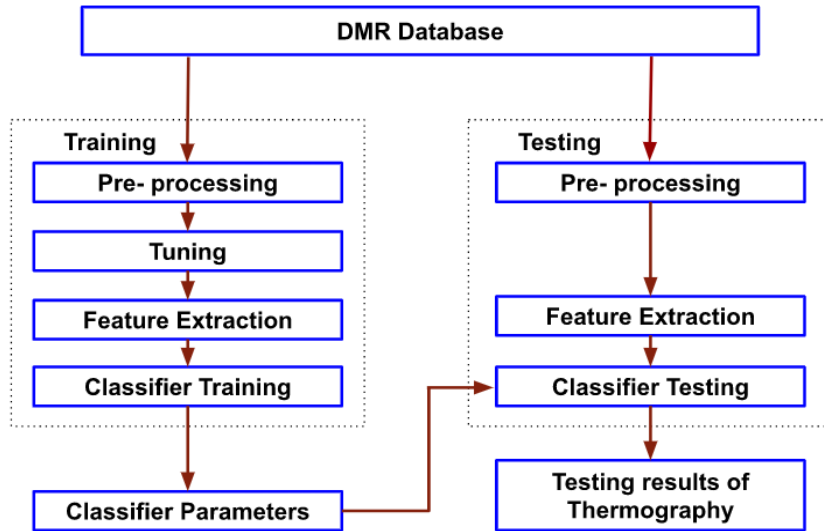
Fig. 6. Flow chart of the suggested approach

As illustrated in **Fig. 6**, the analysis in this work was performed by first pre-processing the thermogram, feature extraction, identifying the relevant important characteristics, and finally using classifier for tumour diagnosis. The process of classification consists of two operations: training and testing. As a result, the previously collected characteristics are given into the classifier to determine if the thermal images are healthy or aberrant. It is a critical phase in pattern recognizing algorithms for categorising and evaluating measurable properties. The confusion matrix was used in the classification step to convey an overall view of how the classifier completed the classification process.

3.4 Evaluation Metrics

Classification metrics assess the model's performance and quantify how excellent or terrible the categorization is. Finally, the average accuracy and average Kappa index for each configuration are used to assess system performance [47]. The Kappa index is a quantitative tool for determining the degree of consistency or repeatability among two data sets; it ranges from -1 to 1. Cohen's Kappa was used. It is a reliable statistics for assessing intra- and inter-rater accuracy. Other prominent classifier measures used for fair evaluation are accuracy, recall, and AUC-ROC[48]. The AUC-ROC, which normally ranges between 0.5 and 1, measures the model's ability to differentiate across classes. It evaluates the model's diagnosis performances in the medical setting; values closer to 1 suggest good test findings, while values closer to 0.5 indicate bad test results [24].

$$\text{Precision} = \frac{TP}{(TP+FP)} \qquad (1)$$

$$\text{Specificity} = \frac{TN}{(TN+FP)} \qquad (2)$$

$$\text{Recall/Sensitivity} = \frac{TP}{(TP+FN)} \qquad (3)$$

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+FN+TN)} \qquad (4)$$

$$F_1 - \text{Score} = 2 * \frac{(recall*precision)}{(recall+precision)} \qquad (5)$$

TP, TN, FP, and FN stand for true positives, true negatives, false positives, and false negatives, respectively. TP and TN signify instances where the model correctly identifies positive and negative cases. FP and FN occur when the model inaccurately labels negative instances as positive and positive instances as negative, respectively.

3.5 Hyper Parameters

Hyper-parameters are features that govern the network's structure as well as parameters that govern the training process of the network (e.g. rate of learning, number of epochs, decay, batch size etc.). One of the foremost important hyper parameter impacting CNN effectiveness is learning rate (LR). SGD optimizers are commonly employed in deep learning models. SGD variants include RMS Prop and Adam. All of these optimizers use LR to regulate and minimise a model's losses. However, if the LR is excessively high, it can generate undesired diverging behaviour and damage the model's pre-trained weights [49]. Over a mini-batch, a batch normalisation layer standardises input variables. It minimises network initialization sensitivity and accelerates training. A larger batch size signifies that the dataset is well captured, signifying that our modeling will progress faster and with fewer diversity while attempting to find the optimum minimum in gradient. Hyper parameters are chosen in an iterative process, refining the choices based on experimentation and monitoring metrics such as accuracy, precision and recall. Fine-tuning is done multiple times guided by improvements in these metrics.

4. Results and Discussion

Experimental settings, results of classification and comparative analysis of the final proposed model with other networks used in mobile applications, and also with various up-to-date techniques are detailed in this section.

4.1 Experimental Settings

Python is used to create the model. Python is beneficial for building ML and DL because it includes a substantial quantity of libraries, including Keras, TensorFlow, NumPy, Pandas, Scikit-learn, and many others. The model was trained on thermograms using the lively procedure from DMR-IR dataset. There are 1522 thermal pictures (762 abnormal and 760 healthy) in the dataset. It was split into 1293 and 229 thermal image training and test sets. **Fig. 7** shows some photos from the data set. Because the supplied dataset was small, the models memorised it instead of training it, causing the model to over-fit the data and to perform poorly on upcoming datasets. We rotated and flipped patches 90°, 180°, and 270°, zoomed, and randomly scaled them. Because tumours may develop in a variety of directions and sizes, such data increases the number of eligible training samples. Hence, augmentation procedures have little effect on the masses' basic disease. As a result, we had
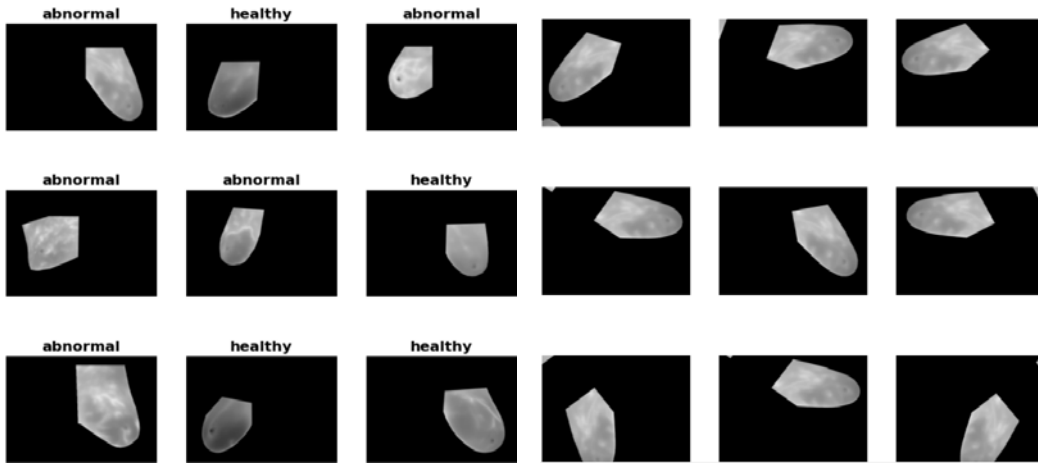two variations of augmented images for every initial image, **Fig. 8**.

Fig. 7. Sample images of the data set   **Fig. 8.** Data-set Augmentation

The count of abnormal and healthy images in the dataset is shown in **Fig. 9**. 'Healthy' depict normal thermograms, while 'abnormal' represent regions with cancerous development.
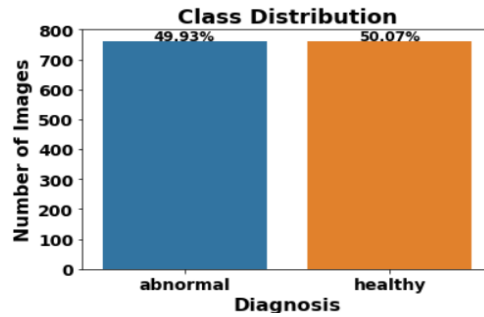

Fig. 9. Image count

Fig. 10 depicts the baseline TL model utilizing MobileNetV2 architecture. To train the models, DL methodologies necessitate a strong GPU hardware and a significant volume of training samples. By re-designing the final few layers of the network and fine tuning the model, TL can be employed to pre-trained CNN model for limited data-sets to tackle this problem. If relevant hyper-parameters are modified and efficient fine-tuning procedures are utilised, the CNN model can attain high accuracy. To minimize loss, an optimization algorithm is used.

The Adam optimizer [50] is an adaptive rate learning optimizer that takes little memory. We used a GAP 2D layer to transform the feature vectors into actual predictions. Periodic learning was utilized to raise the accuracy in image categorization results. Before fine tuning, the learning rate was 0.001, and after fine tuning, it was 0.0001. The batch size for fitting the model was set to 32, and the initial number of epochs was set at 5. The value for the drop-out was set to 0.2. The loss function employed is binary cross entropy. To fine-tune the structure, we must initially unfreeze the baseline framework, followed by freezing all layers preceding a specific layer; in this instance, all layers preceding the 120th layer. Changing the learning rate to run over these layers in fewer increments can result in finer details and greater accuracy.
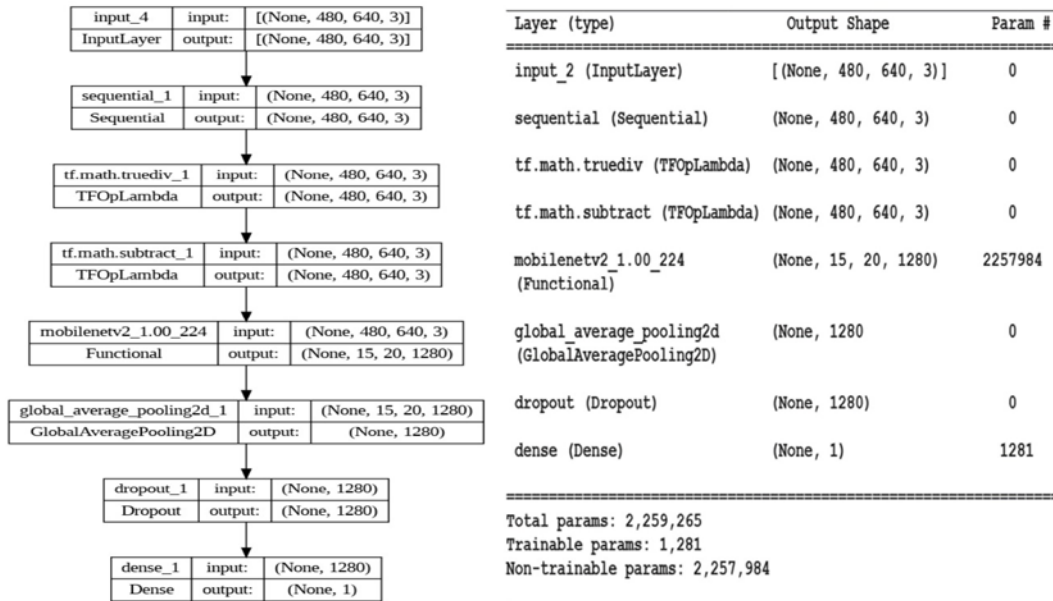
Fig. 10. Baseline Model Parameters utilizing MobileNetV2

4.2 Classification Results

Table 1 exhibits the models' specificity, accuracy, precision, sensitivity, F1-score, Cohen's kappa scores and AUC-ROC. The maximum accuracy rate attained will be used to select the best outcomes.

Table 1. Performance comparison of the suggested models

Classifier Evaluation Index	Baseline Model (Model 1)	Fine Tuned Model (Model 2)	Final Proposed Model
Accuracy (%)	85.15	95.19	98.69
Specificity (%)	79.13	1	98.3
Sensitivity (%)	91.23	90.4	99.1
Precision (%)	81.25	1	98.3
AUC-ROC (%)	85.18	95.18	98.69
F1-Score	85.95	94.93	98.69
Cohen's Kappa	70.3	90.39	97.8

A confusion matrix is a visualization method for assessing classification model performance. It displays successfully and wrongly classified samples, together with the real results from the data. The four confusion matrix parameters, abbreviated as TP, FP, TN, and FN, are crucial in measuring the performance of any classifier. **Fig. 11** depicts the confusion matrices associated with each model.
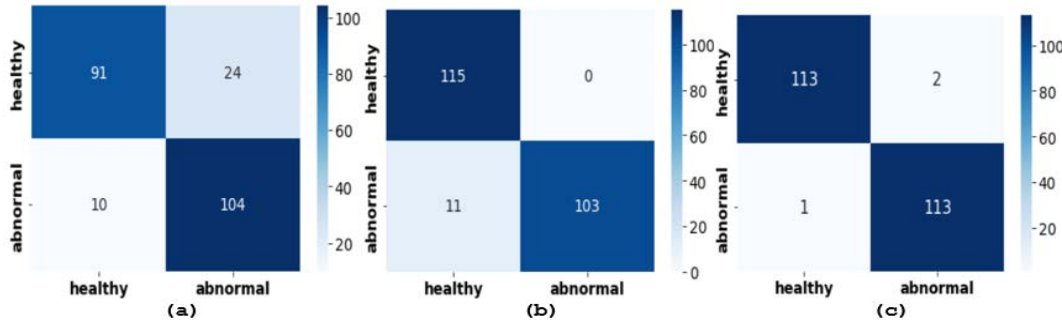
Fig. 11. Calculated confusion matrices of models (a) baseline model (b) fine tuned model
(c) final proposed model

Despite the fact that the model remains identical during the pre-tuning step, the outcomes varied slightly due to the arbitrary split of data-set and the limited amount of epochs. In Keras, we used three functions: model checkpoint, which saved our top model's units weights; Callbacks, which saved the model after each epoch; and Early stopping monitor, which reduced over-fitting and halted the process of training when the model halted learning. The model was evaluated using both training and validation datasets. The training set is utilized to train the model. Meanwhile, the validation set is utilised to assess model performance. An epoch specifies how many times the algorithms examine at the complete data set. When the algorithm has viewed all of the samples within the data-set, an epoch comes to an end. Training with a smaller epoch will result in under-fitting the data, whereas training with a larger epoch will result in over-fitting the data. The number of training epochs is set at 5. The system can train more to enhance accuracy by expanding the number of epochs, incorporating extra layers, and training the data-set.

A graph of the model's validation and training loss and accuracy is shown in **Fig. 12(a)**. The greatest training accuracy in this study is 85.15%, whereas the highest validation accuracy is 84.8%. This suggests that on the new data, this model is anticipated to execute with accuracy of 85.15%. The validation accuracy reduces marginally as the training accuracy raises from epoch 4 to 5. This suggests that the system is better fitting the training set while significantly decreasing its ability to forecast new data, indicating the start of over-fitting of data. A loss function is utilized to enhance the DL algorithm's efficiency. At epoch 0, the training loss keeps diminishing, indicating that the framework is learning to recognise all of the training data-set. The graph shows that the minimal value for training loss is 41% and the minimal reading for validation loss is 41.8%. The training loss for 5th epoch is 42%, and the validation loss is 42.54%. The validation loss is more than the training loss, indicating that the data have begun to over-fit.
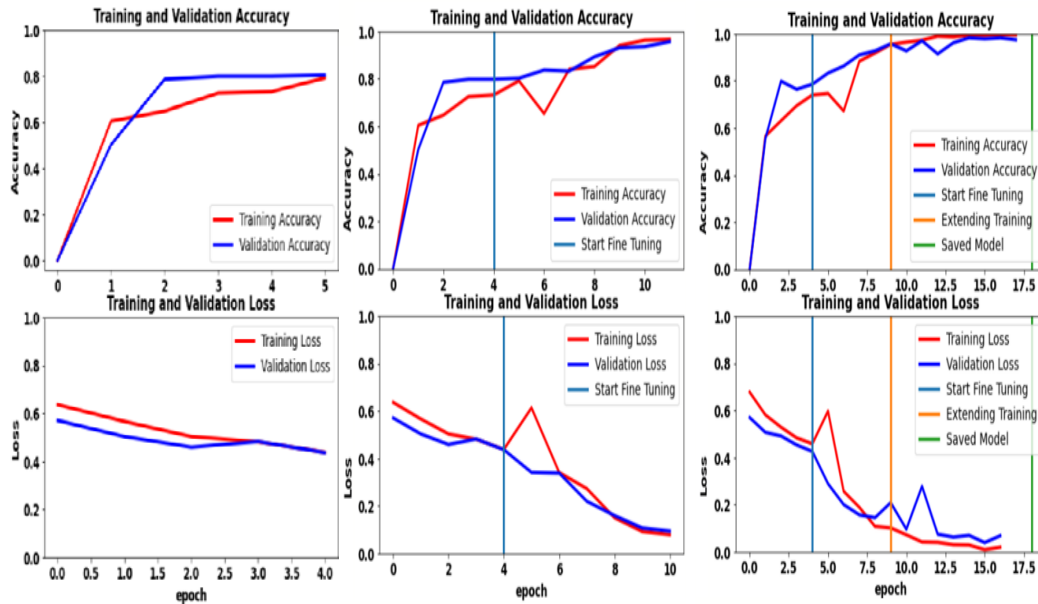
Fig. 12. Training Process (a)Baseline model (b)Fine tuned model (c)Final Proposed Model

The fine-tuned training and validation accuracy and loss are shown in **Fig. 12(b)** and **Fig. 12(c)**. In the plots, the blue vertical line represents the fourth epoch, where fine-tuning began. For a total of 15 epochs, we trained the classification layer introduced to the baseline model for 5 epochs and fine-tuned the entire framework for an extra 10 epochs. It is obvious that utilising Model 2 to fine-tune enhanced validation accuracy from 85.15% to around 95.19%. Validation accuracy changed significantly during the start of the fine-tuning process. It has helped to enhance validation accuracy and minimise validation losses by

employing a slower learning rate. It eventually calmed down to more reasonable rates in later epochs. It was also discovered in **Fig. 12(c)** that the epoch of the preserved model is set to 17. The colored vertical line represents different phases of step fine-tuning. After fine-tuning, validation accuracy improved significantly in earlier epochs. In the training process, curves of accuracy and loss exhibit non-smooth behavior due to several factors such as variability in the training data, randomness introduced by small mini-batch size and stochastic nature of optimization algorithms. Such non-smooth curves are not a cause for concern, but more critical is whether the network converges to a satisfactory performance level and whether there is a consistent trend of improvement on the validation data.

In the first technique, a MobileNetV2 baseline model was created. With weights acquired from the ImageNet dataset, all convolutional layers were made to be non-trainable. This baseline model had an accuracy score of 85.15%. Weights of the final convolutional layer are fine tuned in our second excursion. This MobileNetV2 based fine-tuned TL model outperformed the base model, properly forecasting 95.19% of the sample photos. As can be observed, the fine-tuning method considerably enhances our model's performance. To improve even further, we can train for extended durations; but to ensure that the model does not train for more epochs than necessary, we must provide callbacks for early stopping. We will additionally provide model checkpoint callbacks in order to preserve the optimum model when it is found. This enables the training to recommence from the best weights, for example, if the kernel fails due to unforeseeable reasons. The accuracy of the model was 98.69%. The results

of the fine-tuning reveal that the suggested fine-tuning mechanism outperformed the classical fine-tuning methods in terms of accuracy. Model 3 outperformed the others, with recall, mean accuracy, F1 scores and precision of 99.1%, 98.69%, 98.69 % and 98.3% correspondingly.

The evaluation metrics are computed with ADAM, RMS Prop and SGD optimizing algorithms and the final proposed model performed the best with ADAM optimizer and portrayed in **Fig. 13**. SGD is a derivative of gradient descent. Instead of conducting calculations on the entire dataset, SGD only computes solely on a randomly chosen subset of data examples and gives the worst performance with the final proposed model. ADAM optimizer outperforms RMS Prop as it optimizes adaptive learning rate from estimates of the first and second moments of the gradients whereas RMS Prop uses only the second moment.
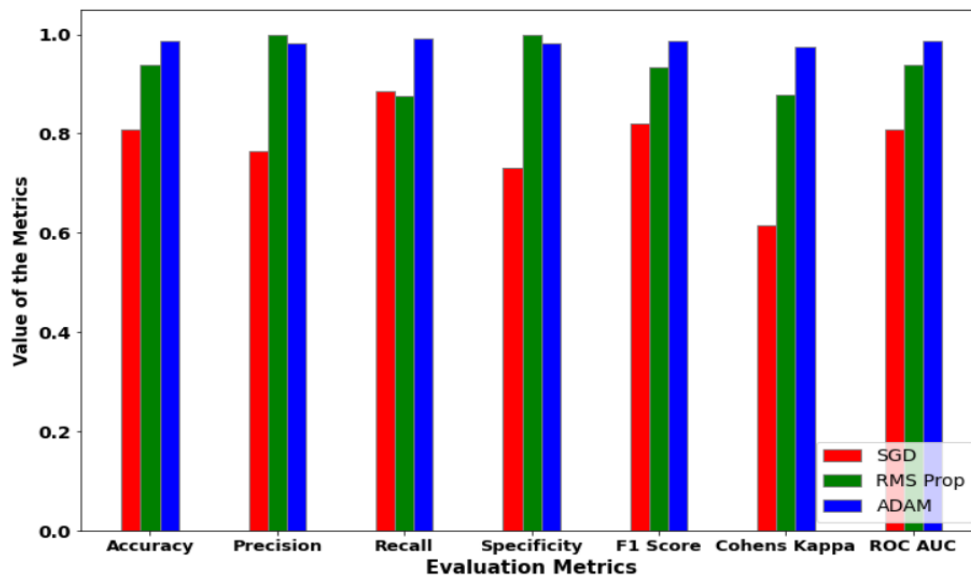

Fig. 13. Efficiency comparison of the final proposed model with SGD, RMS prop and ADAM optimizers

Comparison with other pre-trained CNNs used in mobile applications such as MobileNetV1, NasNetMobile and ShuffleNet [43] are also furnished. MobileNetV1 utilizes depthwise separable convolutions, consisting of depthwise convolution and pointwise convolution to decrease the volume of calculations to perform convolution. Pointwise group convolutions and channel shuffle are the main highlights of ShuffleNet. NASNet is built on reinforcement learning to find the optimum CNN structure. Residual invertions and linear constriction points in MobileNetV2 enhances its performance compared to other cutting-edge mobile CNN architectures. The accuracy achieved by MobileNetV1, NasNetMobile, ShuffleNet and final proposed model are 95.2%, 96.5%, 97.7% and 98.69% respectively. Only in specificity and precision final proposed model shows minimal performance deterioration. **Fig. 14** illustrates that final proposed model with MobileNetV2 performing the best compared to other networks.
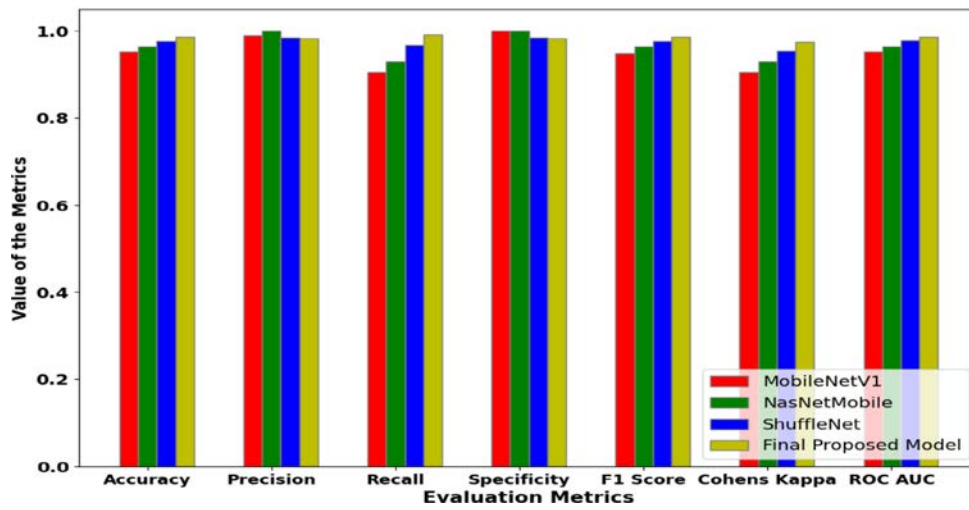
Fig. 14. Performance comparison of the final proposed model with MobileNetV1, NasNetMobile
and ShuffleNet [51]

Table 2 provides a comparison with state-of-the-art techniques, including Extreme
Learning Machine (ELM) [52], Bag of Features (BoF) with SVM [53] and Stacked
Autoencoder (SAE) [54]. The final proposed model surpasses all the other models in terms of
performance.

Table 2. Performance comparison with state-of-the-art techniques

Classifier / Evaluation Index	Extreme Learning Machine	Stacked Autoencoder	Bag of Features and SVM	Final Proposed Model
Accuracy (%)	95.7	87.5	90.6	98.69
Specificity (%)	93.5	93.5	91.1	98.3
Sensitivity (%)	98.0	82.4	90.1	99.1
Precision (%)	93.6	93.7	91.0	98.3
AUC-ROC (%)	98.1	91.6	93.7	98.69
F1-Score	95.8	87.69	90.6	98.69
Cohen's Kappa	91.5	75.1	82.0	97.8

5. Conclusion

Using DL approaches, we examined various strategies for detecting breast cancer in
thermographic pictures. Transfer learning was utilised to create and verify a DL model
grounded in a pre-trained CNN MobileNetV2. The proposed method was evaluated and
validated using a publicly available dataset (DMR-IR). The use of transfer learning and
MobileNetV2 with fine tuning in deep learning for the early diagnosis of cancer using
thermography has been demonstrated to be a promising approach. It enables the development
of a highly accurate algorithm for the assessment of cancer in a much shorter time and with
fewer data than the traditional methods. With an accuracy rate of up to 98.69%, this approach
has proven to be an effective technique for detecting breast cancer. The facts and gaps derived
from previous trials imply that strong scientific investigation in this subject is urgently

required to transform healthcare quality and connect multi-disciplinary breakthroughs. Despite the low cost and non- invasive characteristics of this technique, it would be worthwhile to pursue additional research in this area. The developed model can be realized in portable and economical devices for massive screening to reduce global BC mortality, is planned in future. CNN networks uniquely for early cancer diagnosis from thermal images will be also framed in future expansion.

References

[1] Kandlikar, Satish G., et al., "Infrared imaging technology for breast cancer detection–Current status, protocols and new directions," *International Journal of Heat and Mass Transfer*, vol. 108, pp. 2303-2320, 2017. Article (CrossRef Link)

[2] Sathish, Dayakshini, et al., "Medical imaging techniques and computer aided diagnostic approaches for the detection of breast cancer with an emphasis on thermography-a review," *International journal of medical engineering and informatics*, vol. 8, no. 3, pp. 275-299, 2016. Article (CrossRef Link)

[3] Loizidou, Kosmia, Rafaella Elia, and Costas Pitris, "Computer-aided breast cancer detection and classification in mammography: A comprehensive review," *Computers in Biology and Medicine*, 153, 106554, 2023. Article (CrossRef Link)

[4] Milosevic, Marina, Dragan Jankovic, and Aleksandar Peulic, "Thermography based breast cancer detection using texture features and minimum variance quantization," *EXCLI journal*, 13, 1204, 2014. Article (CrossRef Link)

[5] Garcia, Evelyn M., et al., "Evolution of imaging in breast cancer," *Clinical obstetrics and gynecology*, 59.2, 322-335, 2016. Article (CrossRef Link)

[6] Omranipour, Ramesh, et al., "Comparison of the accuracy of thermography and mammography in the detection of breast cancer," *Breast Care*, 11.4, 260-264, 2016. Article (CrossRef Link)

[7] Mashekova, Aigerim, et al., "Early detection of the breast cancer using infrared technology–A comprehensive review," *Thermal science and engineering progress*, 27, 101142, 2022. Article (CrossRef Link)

[8] Ekici, Sami, and Hushang Jawzal, "Breast cancer diagnosis using thermography and convolutional neural networks," *Medical hypotheses*, 137, 109542, 2020. Article (CrossRef Link)

[9] Ibrahim, Abdelhameed, Shaimaa Mohammed, and Hesham Arafat Ali, "Breast cancer detection and classification using thermography: a review," in *Proc. of The International Conference on Advanced Machine Learning Technologies and Applications (AMLTA2018)*, 496-505, 2018. Article (CrossRef Link)

[10] Kennedy, Deborah A., Tanya Lee, and Dugald Seely, "A comparative review of thermography as a breast cancer screening technique," *Integrative cancer therapies*, 8.1, 9-16, 2009. Article (CrossRef Link)

[11] Okuniewski, Rafał, et al., "Contour classification in thermographic images for detection of breast cancer," *Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments*, Vol. 10031, 2016. Article (CrossRef Link)

[12] Acharya, U. Rajendra, et al., "Thermography based breast cancer detection using texture features and support vector machine," *Journal of medical systems*, 36, 1503-1510, 2012. Article (CrossRef Link)

[13] Prasad, Keerthana, and K. Rajagopal, "Segmentation of Breast Thermogram Images for the Detection of Breast Cancer–A Projection Profile Approach," *Journal of Image and Graphics*, 3.1, 47-51, 2015. Article (CrossRef Link)

[14] Francis, Sheeja V., et al., "Breast cancer detection in rotational thermography images using texture features," *Infrared Physics & Technology*, 67, 490-496, 2014. Article (CrossRef Link)

[15] Krawczyk, Bartosz, and Gerald Schaefer, "Breast thermogram analysis using classifier ensembles and image symmetry features," *IEEE Systems Journal*, 8.3, 921-928, 2014. Article (CrossRef Link)

[16] Lashkari, AmirEhsan, Fatemeh Pak, and Mohammad Firouzmand, "Full intelligent cancer classification of thermal breast images to assist physician in clinical diagnostic applications," *Journal of medical signals and sensors*, 6.1, pp.12-24, 2016. Article (CrossRef Link)

[17] Kakileti, Siva Teja, Aman Dalmia, and Geetha Manjunath, "Exploring deep learning networks for tumour segmentation in infrared images," *Quantitative InfraRed Thermography Journal*, 17.3, 153-168, 2020. Article (CrossRef Link)

[18] He, Kaiming, et al., "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. of the IEEE international conference on computer vision*, 2015. Article (CrossRef Link)

[19] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton, "Deep learning," *nature*, 521.7553, 436-444, 2015. Article (CrossRef Link)

[20] Wainberg, Michael, et al., "Deep learning in biomedicine," *Nature biotechnology*, 36.9, 829-838, 2018. Article (CrossRef Link)

[21] Ben-Nun, Tal, and Torsten Hoefler, "Demystifying parallel and distributed deep learning: An in-depth concurrency analysis," *ACM Computing Surveys (CSUR)*, 52.4, 1-43, 2019. Article (CrossRef Link)

[22] Zhang, Yu-Dong, et al., "Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling," *Journal of computational science*, 27, 57-68, 2018. Article (CrossRef Link)

[23] Zhang, Yu-Dong, et al., "Improved breast cancer classification through combining graph convolutional network and convolutional neural network," *Information Processing & Management*, 58.2, 102439, 2021. Article (CrossRef Link)

[24] Sánchez-Cauce, Raquel, Jorge Pérez-Martín, and Manuel Luque, "Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data," *Computer Methods and Programs in Biomedicine*, 204, 106045, 2021. Article (CrossRef Link)

[25] Mambou, Sebastien Jean, et al., "Breast cancer detection using infrared thermal imaging and a deep learning model," *Sensors*, 18.9, 2799, 2018. Article (CrossRef Link)

[26] Wang, Jiaji, et al., "SNSVM: SqueezeNet-Guided SVM for Breast Cancer Diagnosis," *Computers, Materials & Continua*, 76.2, 2201-2216, 2023. Article (CrossRef Link)

[27] Yao, Xiaoli, et al., "A comparison of mammography, ultrasonography, and far-infrared thermography with pathological results in screening and early diagnosis of breast cancer," *Asian Biomedicine*, 8.1, 11-19, 2014. Article (CrossRef Link)

[28] Rahman, Hameedur, et al., "Efficient Breast Cancer Diagnosis from Complex Mammographic Images Using Deep Convolutional Neural network," *Computational intelligence and neuroscience*, vol. 2023, 2023. Article (CrossRef Link)

[29] Lee, June-Goo, et al., "Deep learning in medical imaging: general overview," *Korean journal of radiology*, 18.4, 570-584, 2017. Article (CrossRef Link)

[30] Aidossov, N., et al., "Evaluation of Integrated CNN, Transfer Learning, and BN with Thermography for Breast Cancer Detection," *Applied Sciences*, 13.1, 600, 2023. Article (CrossRef Link)

[31] Howard, Andrew G., et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017. Article (CrossRef Link)

[32] Sree, S. Vinitha, et al., "Breast imaging systems: a review and comparative study," *Journal of Mechanics in Medicine and Biology*, 10.01, 5-34, 2010. Article (CrossRef Link)

[33] Nogales, Alberto, Fernando Perez-Lara, and Alvaro J. Garcia-Tejedor, "Enhancing breast cancer diagnosis with deep learning and evolutionary algorithms: A comparison of approaches using different thermographic imaging treatments," *Multimedia Tools and Applications*, 1-17, 2023. Article (CrossRef Link)

[34] Brioschi, Gabriel Carneiro, et al., "The Socioeconomic Impact of Artificial Intelligence Applications in Diagnostic Medical Thermography: A Comparative Analysis with Mammography in Breast Cancer Detection and Other Diseases Early Detection," in *Proc. of MICCAI Workshop on Artificial Intelligence over Infrared Images for Medical Applications*, 1-31, 2023. Article (CrossRef Link)

[35] Bronzino, Joseph D., ed., *Medical devices and systems*, CRC press, 2006. Article (CrossRef Link)

[36] Anbar, Michael, et al., "Detection of cancerous breasts by dynamic area telethermometry," *IEEE Engineering in Medicine and Biology Magazine*, 20.5, 80-91, 2001. Article (CrossRef Link)

[37] Keith, Louis G., Jaroslaw J. Oleszczuk, and Martin Laguens, "Circadian rhythm chaos: a new breast cancer marker," *International journal of fertility and women's medicine*, 46.5, 238-247, 2001.

[38] Dey, Ankita, Ebrahim Ali, and Sreeraman Rajan, "Bilateral symmetry-based abnormality detection in breast thermograms using textural features of hot-regions," *IEEE Open Journal of Instrumentation and Measurement*, 2023. Article (CrossRef Link)

[39] DMR-IR. [Online]. Available: http://visual.ic.uff.br/dmi

[40] Huo, Hua, YaLi Yu, and ZhongHua Liu, "Facial expression recognition based on improved depthwise separable convolutional network," *Multimedia Tools and Applications*, 82, 18635-18652, 2023. Article (CrossRef Link)

[41] Alzubaidi, Laith, et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of big Data*, 8, 1-74, 2021. Article (CrossRef Link)

[42] Shorten, C., and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J Big Data*, 6 (1), 1–48, 2019. Article (CrossRef Link)

[43] Sandler, Mark, et al., "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2018. Article (CrossRef Link)

[44] PROENG dataset. [Online]. Available: http://visual.ic.uff.br/en/proeng/thiagoelias/

[45] Madhavi, Vijaya, and Christy Bobby Thomas, "Multi-view breast thermogram analysis by fusing texture features," *Quantitative InfraRed Thermography Journal*, 16.1, 111-128, 2019. Article (CrossRef Link)

[46] Silva, L. F., et al., "A new database for breast research with infrared image," *Journal of Medical Imaging and Health Informatics*, 4.1, 92-100, 2014. Article (CrossRef Link)

[47] Gupta, Sunil, et al., "Comparing the performance of machine learning algorithms using estimated accuracy," *Measurement: Sensors*, 24, 100432, 2022. Article (CrossRef Link)

[48] Orozco-Arias, Simon, et al., "Measuring performance metrics of machine learning algorithms for detecting and classifying transposable elements," *Processes*, 8.6, 638, 2020. Article (CrossRef Link)

[49] Lee, Rebecca Sawyer, et al., "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, 4.1, 1-9, 2017. Article (CrossRef Link)

[50] Kingma, Diederik P., and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. Article (CrossRef Link)

[51] Munadi, Khairul, et al., "A deep learning method for early detection of diabetic foot using decision fusion and thermal images," *Applied Sciences*, 12.15, 7524, 2022. Article (CrossRef Link)

[52] Melekoodappattu, Jayesh George, and Perumal Sankar Subbian, "Automated breast cancer detection using hybrid extreme learning machine classifier," *Journal of Ambient Intelligence and Humanized Computing*, 14.5, 5489-5498, 2023. Article (CrossRef Link)

[53] Ayadi, Wadhah, et al., "A hybrid feature extraction approach for brain MRI classification based on Bag-of-words," *Biomedical Signal Processing and Control*, 48, 144-152, 2019. Article (CrossRef Link)

[54] Nayak, Dillip Ranjan, et al., "A deep autoencoder approach for detection of brain tumor images," *Computers and Electrical Engineering*, 102, 108238, 2022. Article (CrossRef Link)

Swapna Davies received her M.E. degree in Electronics and Telecommunication Engineering from University of Mumbai, India. She is currently pursuing a Ph.D. while also serving as an Assistant Professor in Rajagiri School of Engineering and Technology, Kochi, affiliated to APJ Abdul Kalam Technological University, Thiruvananthapuram, India. Her research interests include deep learning, neural networks and embedded systems.

Jaison Jacob received his B.E. degree in Electronics from Babasaheb Ambedkar Marathwada University, Aurangabad, India, M.E. degree in Applied Electronics from Karunya University, Coimbatore, India, and his Ph.D. degree from Cochin University, Kochi, India. He is a professor at Rajagiri School of Engineering & Technology and doctoral supervisor at APJ Abdul Kalam Technological University, Thiruvananthapuram, India. His research interests focus on machine learning, spectrum sensing in cognitive radio, image processing and micro controller based system design.