Medical & Biological Engineering & Computing (2023) 61:2843–2857
https://doi.org/10.1007/s11517-023-02883-2

ORIGINAL ARTICLE

Hardware deployment of deep learning model for classification of breast carcinoma from digital mammogram images

Kayalvizhi R[1] · Heartlin Maria H[1] · Malarvizhi S[1] · Revathi Venkatraman[2] · Shantanu Patil[3]

Received: 21 October 2022 / Accepted: 3 July 2023 / Published online: 26 July 2023

Abstract
Cancer is an illness that instils fear in many individuals throughout the world due to its lethal nature. However, in most situations, cancer may be cured if detected early and treated properly. Computer-aided diagnosis is gaining traction because it may be used as an initial screening test for many illnesses, including cancer. Deep learning (DL) is a CAD-based artificial intelligence (AI) powered approach which attempts to mimic the cognitive process of the human brain. Various DL algorithms have been applied for breast cancer diagnosis and have obtained adequate accuracy due to the DL technology's high feature learning capabilities. However, when it comes to real-time application, deep neural networks (NN) have a high computational complexity in terms of power, speed, and resource usage. With this in mind, this work proposes a miniaturised NN to reduce the number of parameters and computational complexity for hardware deployment. The quantised NN is then accelerated using field-programmable gate arrays (FPGAs) to increase detection speed and minimise power consumption while guaranteeing high accuracy, thus providing a new avenue in assisting radiologists in breast cancer diagnosis using digital mammograms. When evaluated on benchmark datasets such as DDSM, MIAS, and INbreast, the suggested method achieves high classification rates. The proposed model achieved an accuracy of 99.38% on the combined dataset.

Keywords Breast cancer · FPGA · Deep learning · CNN · Artificial intelligence · Computer-aided diagnosis (CAD)

1 Introduction

In recent years, AI using DL has demonstrated its potential and efficacy in tackling a wide range of real-world issues. Convolutional neural networks (CNNs) are today's cutting-edge DL algorithms, having demonstrated excellent accuracy in handling real-time issues. CNNs outperform conventional algorithms in terms of speed, but they demand a large amount of resources and memory due to the high parameter count in the completely connected layers [1][1]. This is a computational task for general-purpose processors and consumes a lot of electricity. As a result of this problem, hardware accelerators such as GPU, FPGA, and application-specific integrated circuits (ASIC) are being used to enhance CNN throughput [1].

Owing to its increased throughput and memory bandwidth, GPUs are the most extensively utilised platforms for CNNs [3]. GPUs, on the other hand, necessitate a significant amount of power, which is another important performance evaluation factor in today's digital systems. Furthermore, ASIC design has displayed remarkable throughput while consuming little power [4], but at a significant development time and expense when compared to alternative options. As a result, FPGA accelerators are perfect alternatives that provide low power requirements as ASICs yet excellent throughput and customisability at a cheap cost. Especially when real-time healthcare applications are concerned, DL models with high throughput and accuracy are a huge necessity which is where FPGAs play a crucial role. In this

✉ Malarvizhi S
malarvig@srmist.edu.in

[1] Department of Electronics and Communication, SRM Institute of Science and Technology, Kattankulathur, Chennai 603203, India

[2] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Kattankulathur, Chennai 603203, India

[3] Department of Translational Medicine and Research, SRM Medical College Hospital and Research Centre, Kattankulathur, Chennai 60320, India

work, FPGA implementation of CNN is used to identify and classify breast cancer from mammograms.

Breast carcinoma is one of the most prevalent diseases in women worldwide. According to the Global Cancer Observatory (GCO), which is also affiliated with the World Health Organization (WHO), 47.8% of the world's population was diagnosed with breast cancer in 2020, which has the greatest incidence rate among the top ten cancer kinds in the female population [5]. Breast cancer is more likely to be detected early, which increases the likelihood of survival. As a result, regular screening is regarded as one of the most significant methods for assisting in the quick identification of this form of tumour. Since mammography is the most effective assessment procedure for first identification of breast cancer, it is widely used. It can identify many breast anomalies before symptoms emerge. The formation of calcifications and suspicious mass regions on digital mammograms is thought to be a primary indicator of breast cancer, which is why mammograms are prioritised in this research [6].

The following is how this paper is organised: Section 2 summarises the linked studies in breast cancer detection as well as the problems that emerge. Section 3 elaborates on the proposed software and hardware implementation, Section 4 explains the hyperparameter tweaking, Section 5 analyses the outcome, and Section 6 ends and addresses future work.

2 Related work

This study emphasises the pertinent and well-known breast cancer diagnosis methodologies that have been applied. This survey discusses conventional screening methods, mammogram forecasts, and several publicly available mammogram datasets. This portion also provides a quantified dataset-based assessment of DL models with highly preferred as well as commonly used public datasets; additionally, the section emphasises the set-backs and advantages of conventional CAD systems for breast cancer classification.

Mughal et al. identified the abnormal breast boundary area using an entropy filter-based texture image generation technique. Furthermore, they extracted and refined the ROI using a mathematical morphological function. The extracted equation features were classified using SVM, decision tree, KNN, and bagging tree classifiers. It was observed that SVM produced the best results, with accuracy of 96.9% for DDSM and 97.5% for MIAS, respectively [7].

For certain studies, the extreme learning machine (ELM) [8], a sort of NN, in addition to only one concealed layer was used. Mohanty Figlu et al. [9] developed a mass identification computer-aided diagnosis that showed high precision with a limited number of features. Mammograms were classified as normal or aberrant, and their method determined whether the tumour was benign or cancerous. They validated their suggested technique using the DDSM, MIAS, and BCDR datasets. The chaotic maps and weight ideas are applied to the swarm algorithm in their method to select the optimal feature set and change the characteristics of the ELM algorithm. They employed principal component analysis (PCA) for feature reduction and a modified learning strategy based on ELM for classification. For normal and abnormal categorisation, their approach attained an accuracy of 99.62% for MIAS and 99.92% for DDSM. Despite the fact that their model can classify images in real time, the manually cropped ROIs are regarded as a flaw in such an automatic CAD system.

Transfer learning has recently been recognised as a key to enhancing learner model success. It is defined as the ideology of moving information acquired from one job to another. Transfer learning is now vastly used in most major CAD systems to address the issue of inadequate data while also lowering the processing cost and time required for training the models [10]. Ansar et al. suggested a MobileNet-based architectural model capable of categorising mammogram masses as malignant or benign with reasonable performance when compared to state-of-the-art architectures. The proposed technique detects masses in mammography by first classifying them as carcinoma or normal using a convolutional neural network and then passing the cancerous ones to a pretrained model for classification. Their model did well, with DDSM precision of 86.8% [11].

Although previous efforts dependent on ML have shown remarkable results, ML has shortcomings in terms of error susceptibility, which is a crucial aspect when it comes to real-time deployment. However, the issue of negative transfer is one of the most significant constraints to transfer learning. This setback is overcome by DL.

Dhungel Neeraj et al. developed a CAD tool with little user participation for bulk identification, segmentation, and classifying mammographic images. For bulk identification, they chose a random forest and a succession of DL models and hypothesis refining. They categorised using a DL model that had been priorly trained with hand-curated feature values and was validated on the INbreast dataset. The algorithm identified more than 90% of the masses with only one false-positive rate per image, and the model had a classification sensitivity of 0.98 [12].

Shen et al. developed a NN that detects tumours in breast images using adversarial learning. The completely connected network takes the desired region as input to create a pixel-by-pixel heatmap, with each pixel showing whether the pertinent input pixel is linked with a mass lesion [13].

Ribli Dezs et al. [14] built a system that can identify, locate, and categorise abnormalities in mammograms using Faster R-CNN [15]. They employed the DDSM in their study, and as a result of the low quality of digitised mammograms, they linked the pixel values with optical density, and then resized the pixel values to 0–255 range. They discovered that higher quality photos produce better outcomes using their experiment model. For testing, they implemented INbreast dataset, and for training, they employed a proprietary dataset in addition to the DDSM dataset. The model's last layer classifies the masses as benign or cancerous, and it also provides a bounding box for every discovered mass. Without any human assistance, the technology finds and diagnoses malignant or benign tumours on a mammogram. The suggested technique achieves the best classification performance using INbreast dataset, with an AUC of 0.95.

DL has eliminated the need for hand-curated features by automatically learning the most pertinent attributes to use when performing a specific job. However, DL requires pricy GPUs to perform training and testing at an accelerated speed. To overcome this drawback, an FPGA is incorporated in the proposed DL system to efficiently classify mammogram images in real time.

Studies conducted on FPGA acceleration for CNNs used for different tasks are briefed as follows:

FPGA acceleration for CNNs used for various jobs has been studied. These studies have revealed that FPGA acceleration for CNN has resulted in reduced computational complexity of the convolutional layers and to improve resource efficiency [16]. Few of these studies have been discussed as follows:

Bing Liu et al. suggested a CNN accelerator based on the Xilinx ZYNQ 7100 hardware architecture that accelerates both conventional and depth-wise separated convolution for CNNs used in a variety of disciplines. The authors also claim that the FPGA's highly restricted resources, coupled with CNN's large number of factors and processing complexity, present major design challenges. According to the results of the experiments, the accelerator created in this article can achieve 17.11GIGA operations per second (GOPS) for 32bit floating point [17].

A hardware accelerator architecture for CNNs was suggested by Zhiqiang Liu et al. The suggested standard design allows for the rapid creation of 2D and 3D CNN accelerators with cutting-edge throughput, delay, and energy economy. Despite the lower performance density compared to customised accelerators depending on the Winograd algorithm, this design is more adaptable, allowing for the acceleration of different CNN models without reconfiguring the FPGA. Finally, at a clock frequency of 100 MHz, the accelerator attained a processing rate of 17 giga floating point operations per second (GFLOPS) which is a statistic used to assess the computational efficiency system, such as a CPU or GPU [18].

These studies have shown significant resource utilisation and acceleration of CNNs using FPGAs. In this view, FPGA-based CNN acceleration is performed in the proposed study.

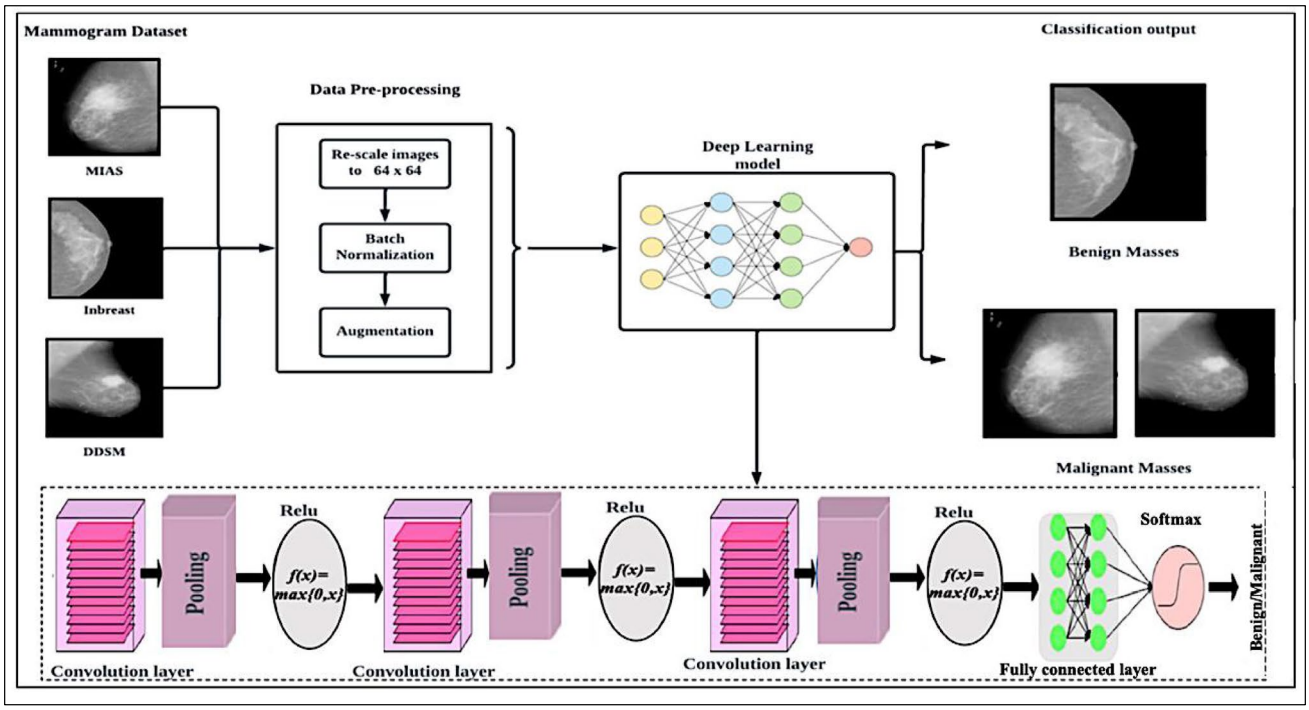


Fig. 1 Proposed framework of the DL classifier

3 Proposed method

The proposed system has been discussed in this section. First, the mammography datasets needed to develop the framework is briefed. Following that, the entire development process is presented, including the pre-processing phase, the CNN training and testing stage, and the hardware acceleration phase. The entire workflow of the proposed system is shown in Fig. 1 where three images one from each dataset are shown and its respective classification is seen as classification output. Image from INbreast is classified as benign and others as malignant.

3.1 Benchmark datasets used for the study

Three benchmark datasets are used for this study, namely, DDSM, MIAS, and INbreast which are discussed in brief as follows:

3.1.1 DDSM

DDSM is composed of 2620 digitised film mammogram exams divided into 43 parts. Because each breast side was captured from two angles, namely, right and left mediolateral, right and left cranio-caudal. Therefore, each case has four breast mammographic images. At the pixel level, the dataset also gives ground truth and kinds of suspicious regions. Figure 2 represents a DDSM sample from the 4 mammogram views of a single patient.

3.1.2 MIAS

The collection contains 322 digital videos as well as ground truth marks for every existing anomaly. Sample images with mass and no mass from the MIAS dataset are presented (Fig. 3). The photos are freely available on the University of Essex's Pilot European Image Processing Archive (PEIPA) [19].

3.1.3 INbreast

In the INbreast dataset, out of 115 instances, cancer was detected in 90 cases. The dataset contains four distinct types of breast diseases.

The database contains scans of CC and MLO perspectives recorded in DICOM format [20]. Figure 4 represents mammogram images of various densities from the INbreast dataset.

3.2 Pre-processing

The inadequate contrast of mammogram images is one of the concerns in mammography. This makes it challenging for radiologists to evaluate results. Furthermore, the

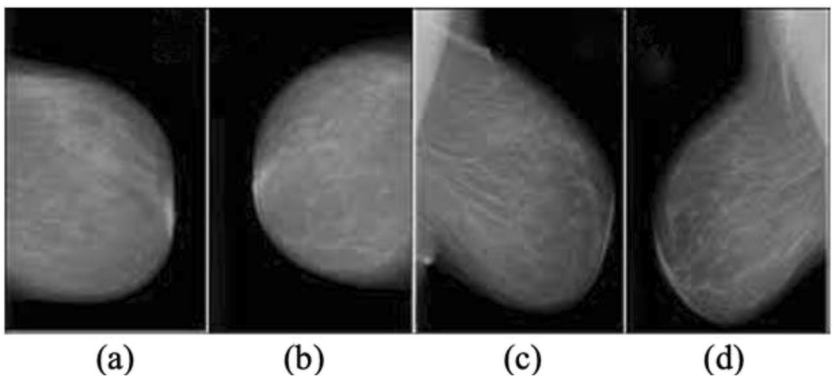


Fig. 2 The left and right CC and MLO views of a single patient from DDSM dataset. (**a**) LCC (**b**) RCC (**c**) LMLO (**d**) RMLO

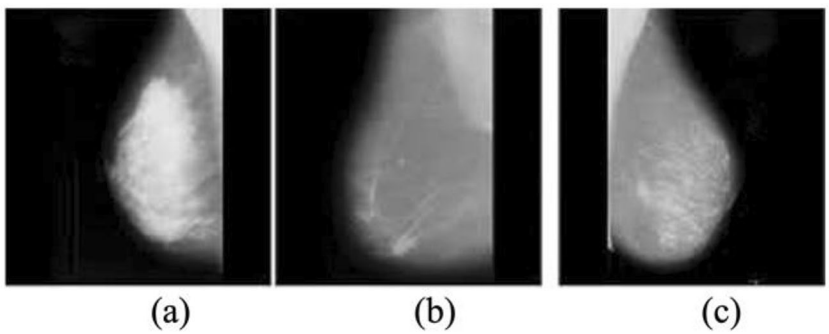


Fig. 3 MIAS mammogram sample **a** benign, b benign, **c** no mass

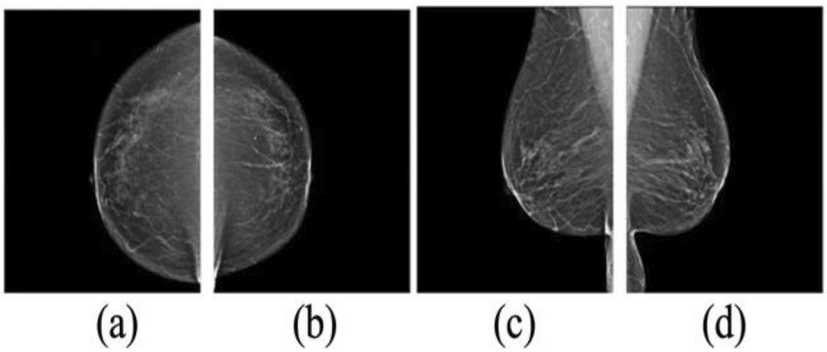


Fig. 4 **a** and **b** represent the left and right CC view, **c** and **d** represent left and right MLO mammogram view of a single patient from INbreast dataset

fundamental concern is the occurrence of noise during image acquisition. The most common noise that impact mammography pictures are salt and pepper noise, Poisson noise, and Gaussian noise [21][22]. Therefore, before further processing, the mammography images are de-noised using the double-density dual-tree complex wavelet transform (DD-DTCWT) [23].

Further, batch normalisation and data augmentation were performed on the obtained datasets.

3.3 Custom CNN model

The proposed CNN is formed using three convolutional layers, three maxpooling (mp) layers, and one fully connected layer. The hyperparameters of this simple network are fine-tuned in such a way that it is suitable for hardware deployment while not compromising the classification performance. The de-noised mammogram images from the INbreast, MIAS, and DDSM datasets are sent into the proposed CNN for classification. For the convolutional layers, 32, 32, and 30 kernels each of size of 3×3 is employed. These kernels extract information pertaining to the texture and intensity of an image which is followed by a 2×2 mp layer (downsampler) with a stride of two and zero padding. ReLU activation was used for all three convolutional layers. The fully connected layer parameters 16,448 are routed to the softmax activation function. The CNN model in total consists of 53,505 learnable parameters.

3.4 Miniaturisation of CNN for hardware development

The proposed CNN is a miniaturised DL model architecture to promote ease of deployment in hardware with reduced number of architectural parameters. The proposed CNN is applied a dropout of 50%. By doing so, the parameters are minimised, thus making it feasible for hardware deployment. Following the neural model training procedure, we obtain a collection of weights for every trainable layer. Such weights are not distributed evenly for a given data format. Most of the weights are centred on 0 or extremely near to it. As a result, their influence on the final activation levels is negligible. Depending on the network implementation constraints, storing weights may necessitate a large amount of memory [24]. The use of such miniaturisation is to eliminate certain weights do have a direct influence on reducing storage needs.

3.5 FPGA acceleration for the proposed CNN

The trained NN is then deployed on PYNQ-Z2 FPGA board. FPGAs have fewer resources than GPU-based devices, yet they can still analyse NNs quite quickly while consuming less power and taking up less space. This makes it an excellent alternative for usage in the field. In recent years, academics have developed many approaches for implementing CNNs on FPGAs. In this study, we have implemented the CNN on FPGA using Vivado tool (Fig. 5). The miniature CNN is then accelerated using Tensil accelerator [25] after which the Register Transfer Level (RTL) block is generated for PYNQ architecture. The generated RTL block is then used to develop the schematic in Vivado tool. Once the developed schematic is synthesised, the bitstream file (.bit) and the hardware wrapper file (.hwh) are extracted. In the final phase, the extracted bitstream file and hardware wrapper are utilised for the final PYNQ overlay phase.

4 Hyperparameter tuning of the proposed convolutional neural network

Hyperparameters have a direct influence on model structure, function, and performance. Tuning hyperparameters facilitate optimising the model performance.

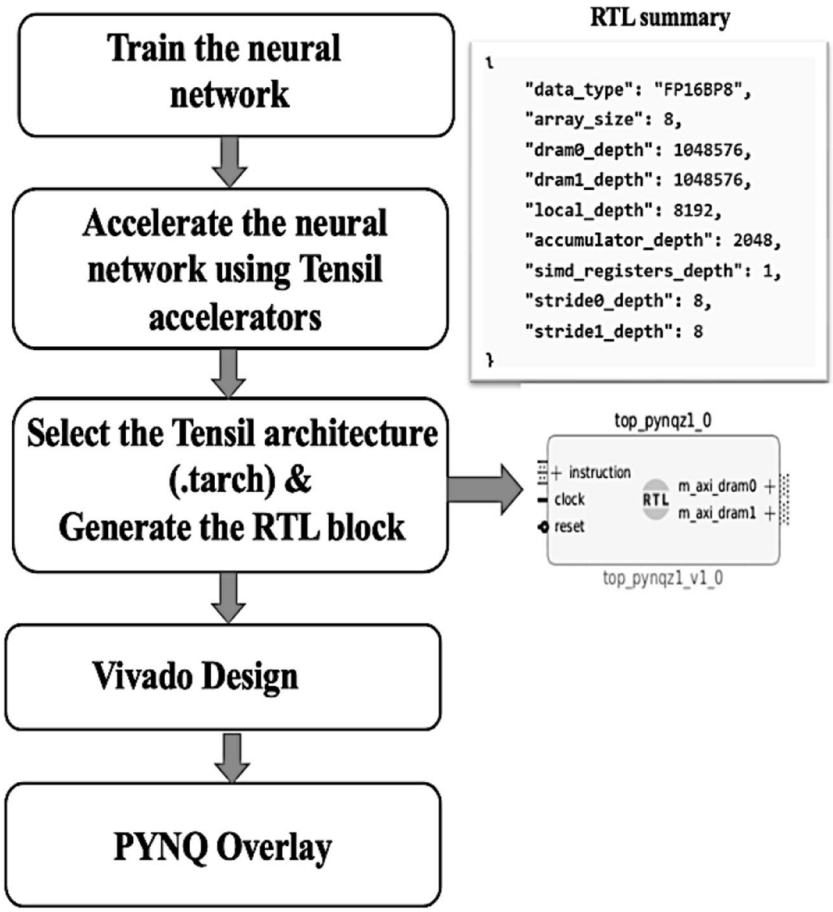
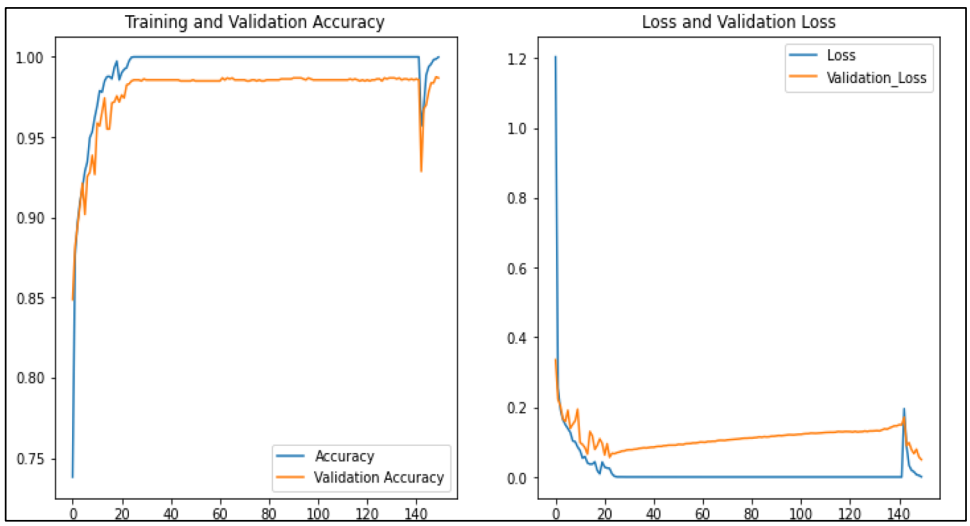


Fig. 5 Flowchart of PYNQ board deployment

4.1 Learning rate (lr)

The lr of a NN is the most important hyperparameter that has a substantial influence on its performance. The efficacy of NNs changed when trained with different lrs.

A lr that is very large or very less will have an adverse effect on the training of the NN. And the key aspect is determining how to acquire the best performance for a NN [26]. A number of learning rate techniques have recently been developed to obtain a higher learning rate value.



(a) (b)

Fig. 6 **a** Accuracy at lr -0.001 and **b** loss at lr-0.01

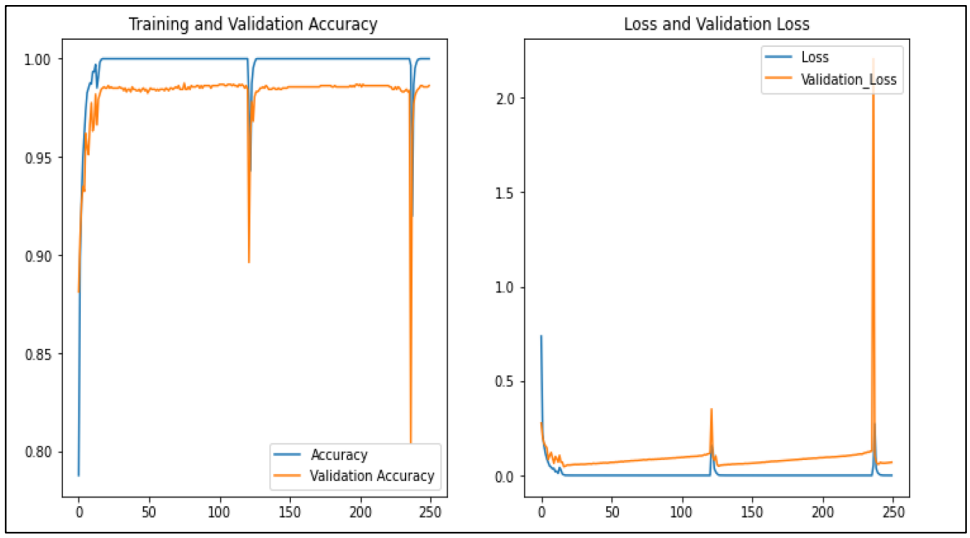


Fig. 7 **a** Accuracy at lr -0.005 and **b** loss at lr-0.001

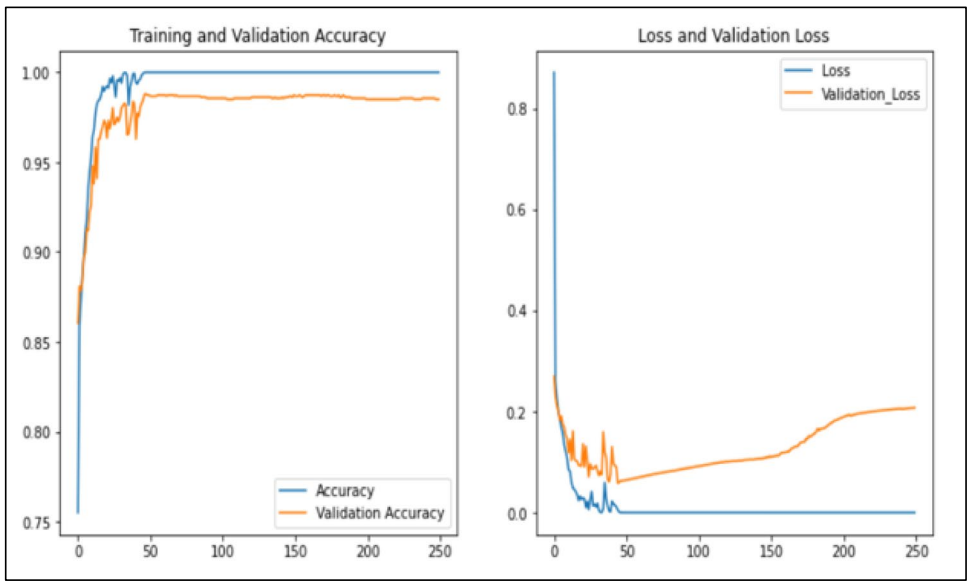


Fig. 8 **a** Accuracy at lr -0.0001 and **b** loss at lr-0.0001

These solutions can boost network performance, increase accuracy, and lower the loss function value, therefore making hyperparameter setting more challenging. Owing to this concern, various learning rates were considered for training and the observations are listed below.

From Figs. 6, 7, and 8, it is understood that the model training was not ideal and none of the learning rate values were appropriate for mammogram images. Therefore, combinations of dropouts were applied.

4.2 Dropouts

CNN has a lot of parameters, specifically in FC layer; hence having very large number of parameters might lead to overfitting. To avoid overfitting, many distinct networks are often trained as the model combination. However, training the different models independently takes time and is challenging to implement. Dropping out is one solution to these issues. The main principle behind dropout is to randomly remove a portion of units having probability 1-p at each training step, with p decided by experiment. In other words, after incorporating dropout, the networks diverge from one another and grow thinner than a typical neural net, increasing the model's susceptibility to overfitting and speeding up training [27]. Dropouts of 30%, 50%, and 70% were chosen at random and applied on the NN. From the learning rate trials, lr of 0.0001 was selected and combined with different combinations of dropouts as shown in Figs. 9, 10 to Fig. 11. From Fig. 11, it

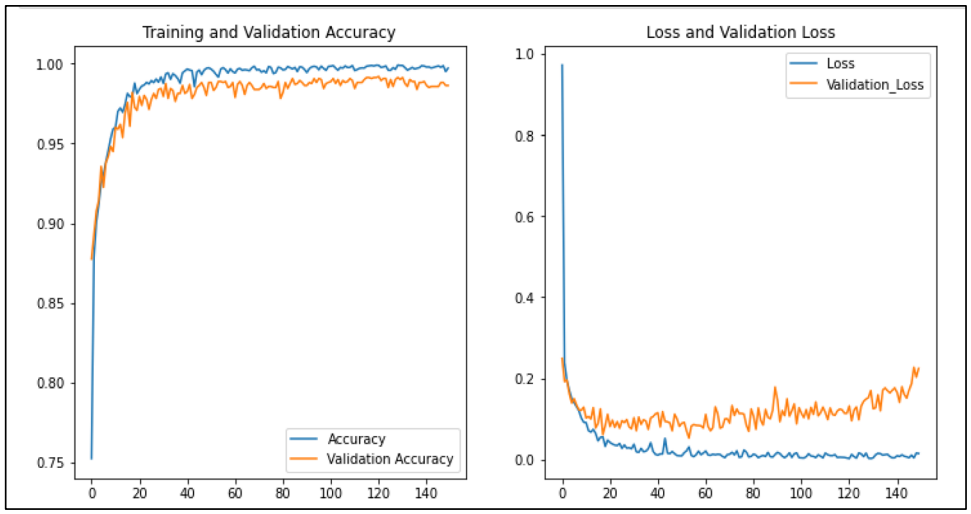


Fig. 9 Dropout 30% and lr-0.0001 **a** accuracy and **b** loss

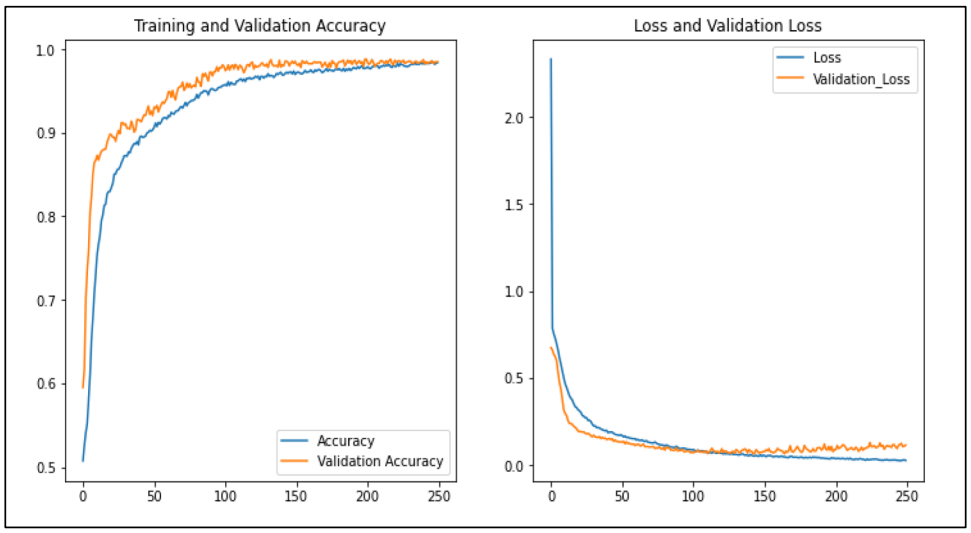


Fig. 10 Dropout 50% and lr-0.0001 **a** accuracy and **b** loss

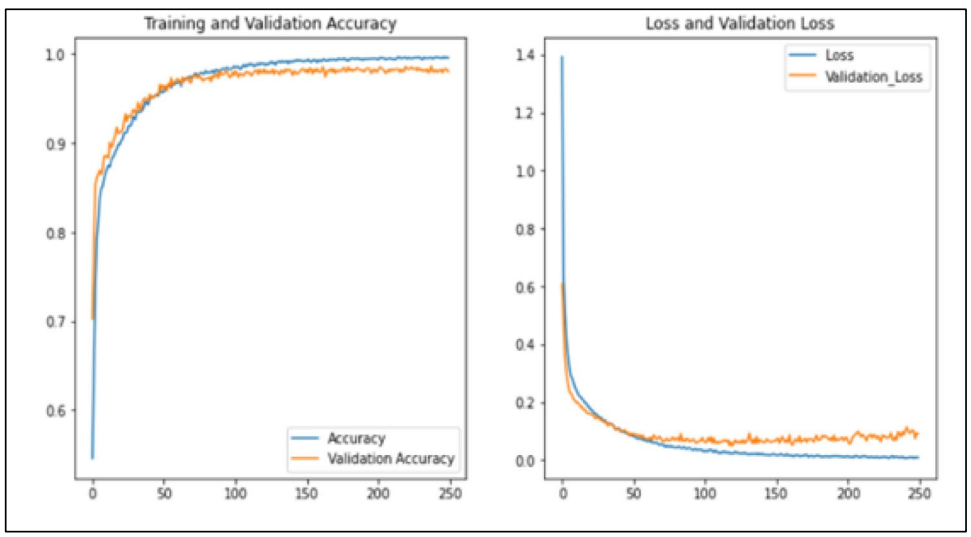


Fig. 11 Dropout 70% and lr-0.0001 **a** accuracy and **b** loss

is evident that 70% dropout with a learning rate of 0.0001 has yielded better results comparatively.

4.3 Regulariser

Regularisation in NNs alters learning algorithms to minimise generalisation (testing) error but not training error. It is primarily meant for linear learning models such as linear regression, which provide for easy, straightforward, and compelling regularisation approaches. The preferred regularisation mechanism for DL is based on regularising estimators. The most successful regulariser is one that provides a lucrative transaction while lowering error significantly by not raising the bias substantially. Regularisers are of two types, namely, L1 and L2. L1 regularisation has the intriguing trait of leading to sparse weight vectors during optimisation. It is an effective strategy for feature selection throughout the learning process. With L1 regularisation, input layer neurons propagate a sparse subset of the most significant information and become almost invariant to the noisy features. Because of its capacity to simplify solutions, L2 regularisation is the most popular squared magnitude of weight parameters to the loss function [28]. L2 regularisation, on the other hand, prevents overfitting by penalising the regularisation, which is referred to as elastic penalisation. A combination of L1 and L2 is a choice in machine learning. Similarly, in this study both l1 and l2 regularisers are used.

5 Experimental outcome and discussion

The practical analysis of the suggested system is shown in this section. For the implementation, Python platform built on top of TensorFlow was employed. Three separate

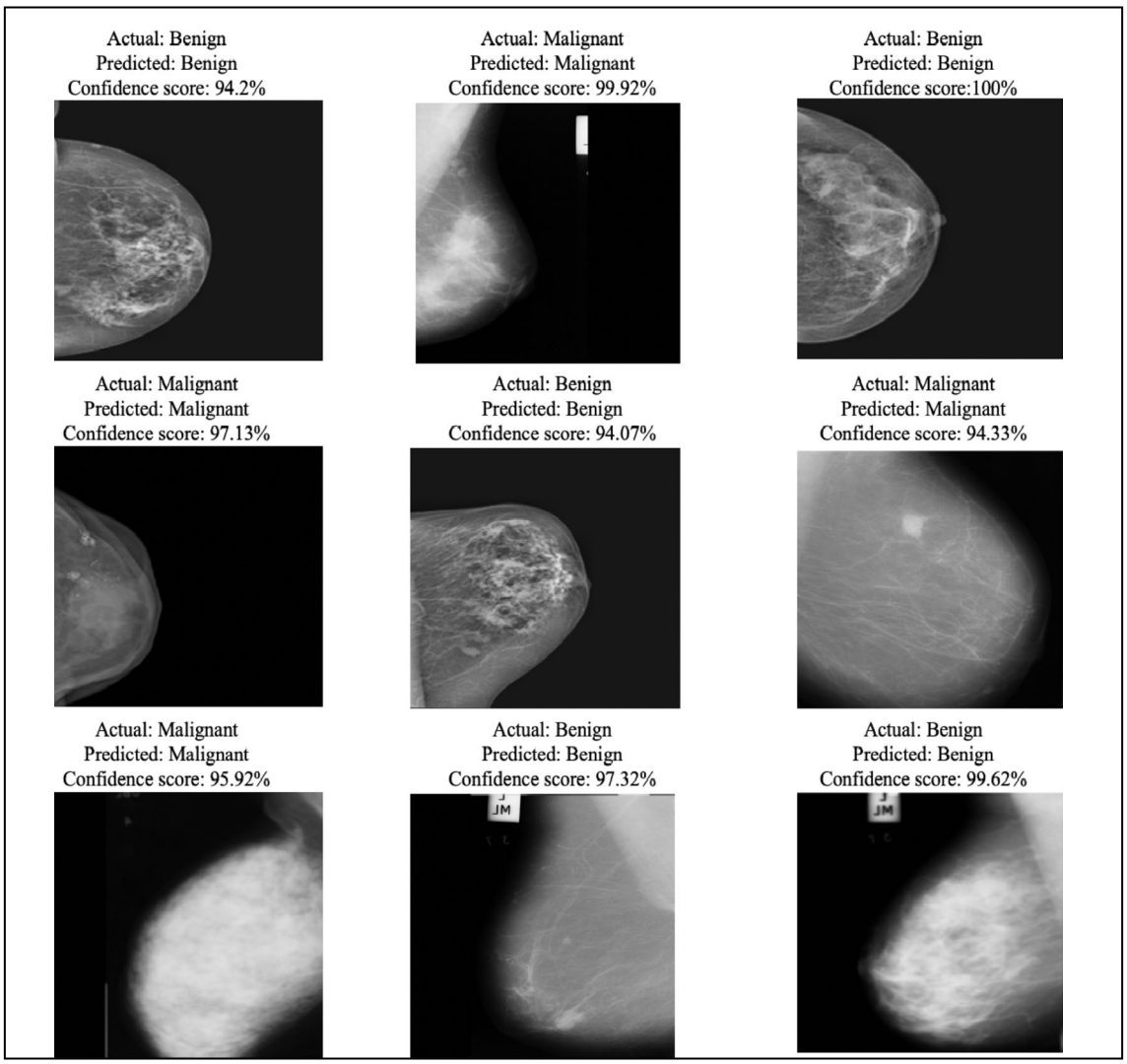


Fig. 12 Selective results of software simulation

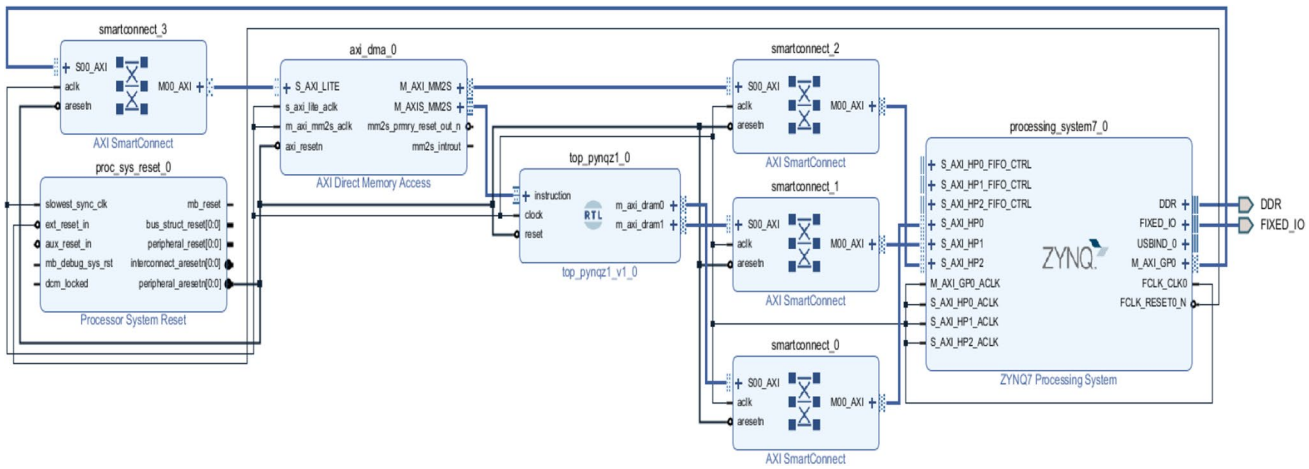


Fig. 13 Schematic of PYNQ generated using the Vivado tool

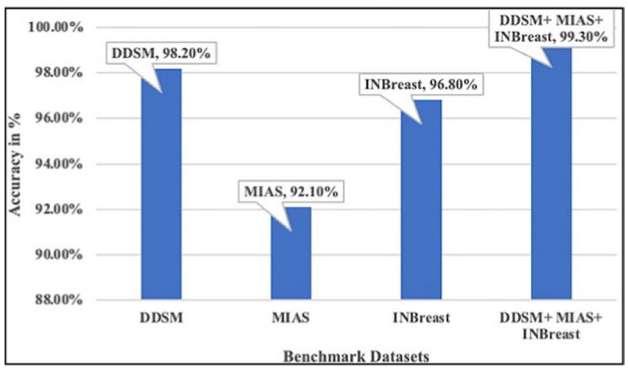


Fig. 14 Performance accuracy observed during testing

standard mammography imaging datasets were used: the MIAS, INbreast, and DDSM. The suggested approach was evaluated by classifying the digital mammograms into malignant or non-malignant utilising these databases and their combinations. Figure 12 represents the simulation results and confidence score of few mammogram samples.

According to the results of the trials, the suggested model has a high classification rate (Fig. 13). To compare the suggested model to the conventional models, it was compared to the evaluated papers that employed the same databases, which are MIAS, DDSM, and INbreast. From Fig. 14, it is noted that, for the MIAS dataset, it has been discovered that the suggested model results outperform [29]. Using the DDSM dataset, it was discovered that the suggested model outperformed [30]. It has been discovered that the suggested model outperforms the results of [31] for the INbreast dataset.

From the confusion matrix displayed in Fig. 15 the true positives, true negative, false positive, and false negative values are derived and were used to compute the parameters discussed in Table 1. Figure 15a represents CNN model performance for DDSM, Fig. 15b for MIAS, Fig. 15c for INbreast dataset images, and Fig. 15d for combined images,

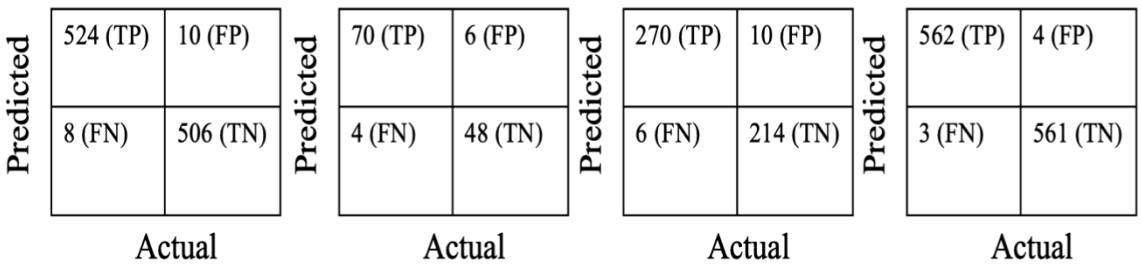


Fig. 15 Confusion matrix for datasets **a** DDSM, **b** MIAS, **c**, INbreast, **d** combined DDSM + MIAS + INbreast

Table 1 Performance observations on various datasets

Dataset	Precision	Recall	F1
DDSM	0.98	0.98	0.97
MIAS	0.92	0.94	0.92
INbreast	0.96	0.97	0.96
DDSM + MIAS + INbreast	0.99	0.99	0.98

namely, DDSM + MIAS + INbreast which the authors have created and pre-processed. On simulation, we found that the confusion matrix of Fig. 15d performs well. The performance of the proposed method in Fig. 15d shows that the false positive is 60% less when compared to the confusion matrix in Fig. 15a and c and 33% less when compared to the confusion matrix in Fig. 15c. This shows that when the model was trained with a higher and wider range of datasets, the generalisation of the model has increased. In the combined dataset, 1130 images were evaluated in proposed CNN, with 562 malignant images predicted as malignant and 561 benign images projected as benign. Additionally, we have taken 15 patches of fatty tissue images and 15 patches of tumours and computed the co-occurrence matrix from which the contrast and number of occurrences were found based on which the CNN identifies the tumour region precisely. Precision, recall, and F1 score are classification algorithm performance indicators obtained from the confusion matrix and listed in Table 1. In Table 2 a brief comparative

Table 2 Comparative analysis with existing breast cancer classification technique

Author	Task	Dataset	Type of network	Input Size	Evaluation results	Hardware implementation/ acceleration
Salma et al. [32]	Classification of cancerous and non-cancerous anomalies	302(MIAS) 534(DDSM)	InceptionV3, DenseNet121, ResNet50, VGG16 and MobileNetV2 for Classification	32 × 32	Accuracy: 0.988	No
Ragab et al. [33]	Classification of cancerous and non-cancerous anomalies	DDSM 5257 images	DCNN: (AlexNet[5 Conv. Layers + 3 Pooling Layers + 2 Fully Connected Layers] + SVM Classifier	227 × 227	Accuracy: 0.872	No
Ting et al. [29]	Classification of cancerous and non-cancerous anomalies	MIAS 221 (21 benign, 17 malignant, 183 normal) 78	CNN: 5 Conv. Layers + 2 Fully Connected Layers + Softmax Classifier	128 × 128	Accuracy: 0.905	No
Dhungel et al. [12]	Classification of cancerous and non-cancerous anomalies	INBreast 410	CNN: 4 Conv. Layers + 1 Fully Connected Layer + Softmax Classifier	40 × 40	Accuracy: 0.850	No
Proposed model	Classification of cancerous and non-cancerous anomalies	MIAS(128) DDSM(1048) INbreast(500) Combined(1110)	CNN: 3 convolutional layers + 2 fully connected layers	64 × 64	Accuracy: 0.993	YES

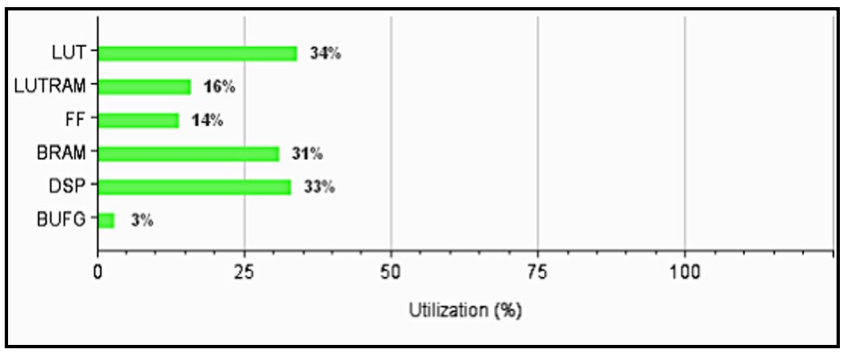


Fig. 16 Resource utilisation report for PYNQ hardware

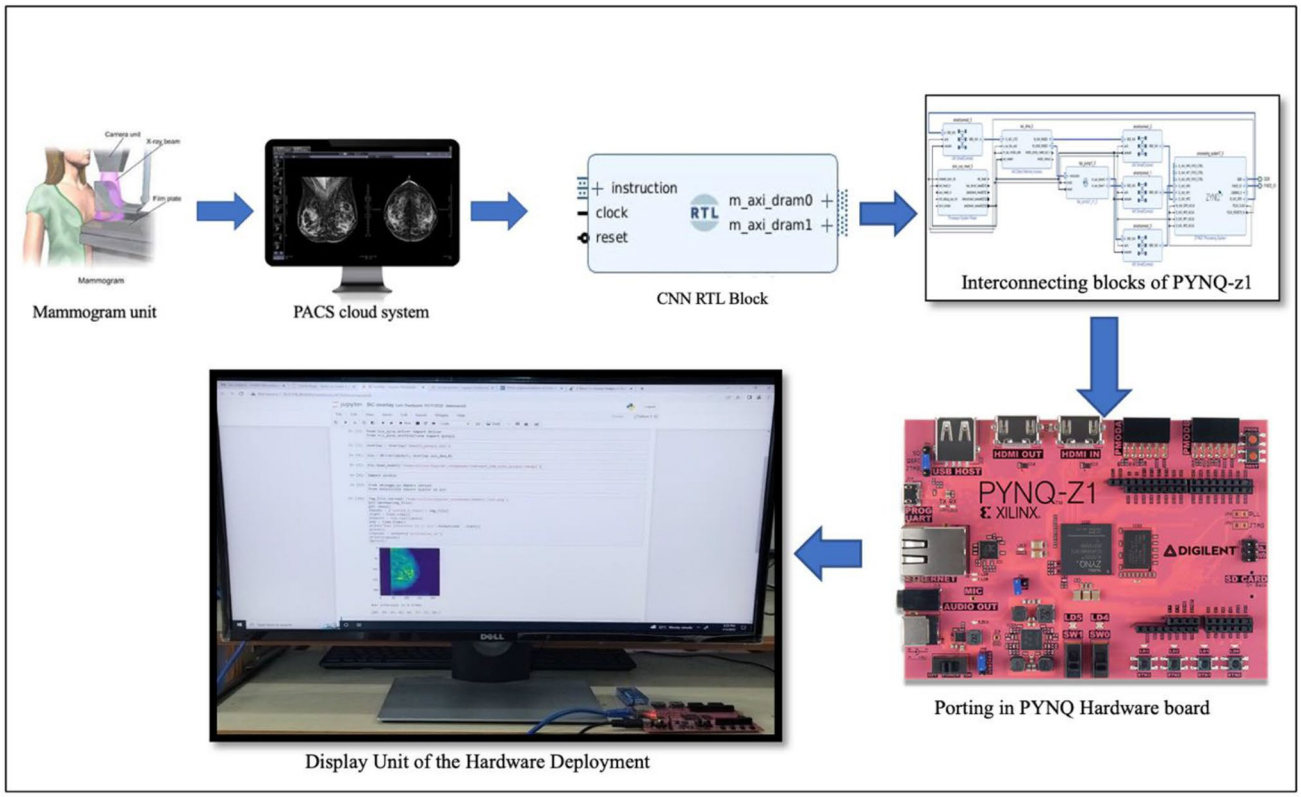


Fig. 17 Process flow of real-time testing

study of the proposed approach with the state-of-the-art techniques [12, 29, 32, 33] is provided.

The CNN model trained on the combined dataset was then deployed on PYNQ board for resulting in a power consumption as less as 3 Watts and an inference rate of 0.8 s per digital mammogram approximately. Figure 13 shows the schematic representation of PYNQ designed using Vivado tool suite. The resource utilisation report for Fig. 13 is given in Fig. 16 from which it is understood that resource utilisation is minimal. The direct links between Tensil block Advanced eXtensible Interface (AXI) masters and ZYNQ FPGA AXI slaves are shown in Fig. 13. The AXI Direct Memory Access (DMA) block connects the command AXI stream interface.

The AXI4-Stream is used for high-speed streaming data, with unlimited burst uploads. There is no address system; this bus type is best adapted to direct data flow between source and target. AXI interfaces include FIFO buffers to handle "bursty" read and write behaviour and enable high rate communications between the program logic and memory elements in the program system.

Further, for validating the robustness of the proposed model, 15 real-time cases were obtained from SRM Medical College and Hospital in dicom format, Chennai along with the required ethical clearance. Figure 17 shows the flow of real-time testing. The mammograms taken by the mammogram unit are transferred to picture archiving and communication system (PACS), which is a cloud system adopted at hospitals globally. From PACS, the image is downloaded and imported to the proposed neural network which is up and running on the hardware. Once the testing is complete, the diagnostic inference is displayed on the monitor connected to the hardware. Few test results of the real time cases are displayed in Fig. 18.

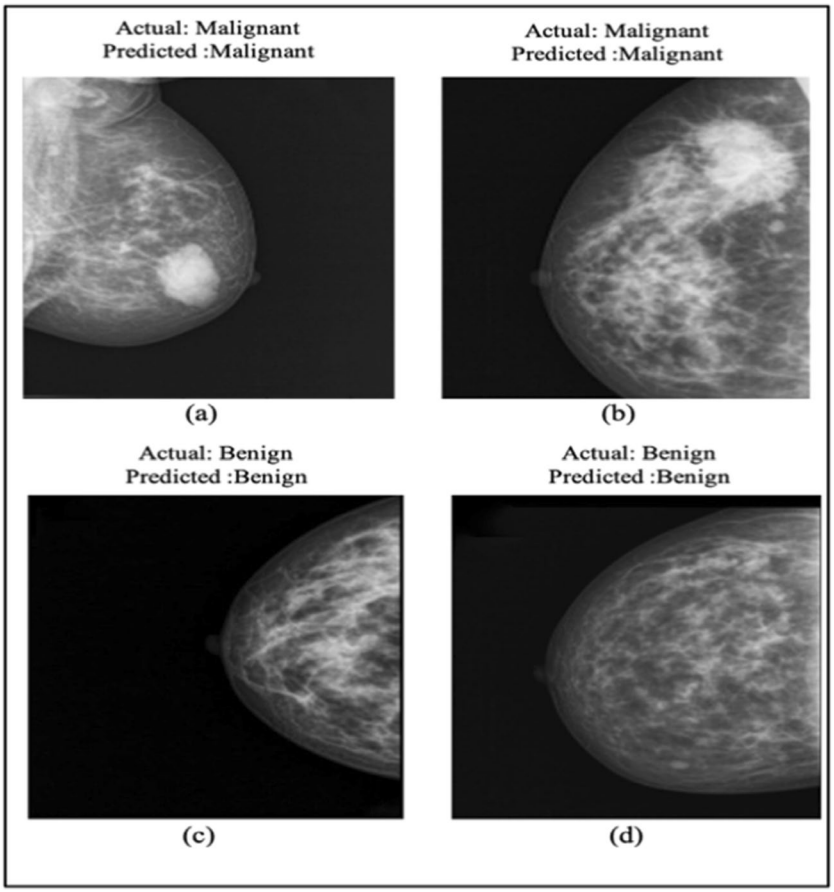


Fig. 18 Simulation results of real-time testing

6 Conclusion

Breast cancer is still an issue that has to be treated at its most fundamental level. Given the repercussions of breast malignancy in women, a preventative measure is implemented for the benefit of society. Also, breast cancer detection utilising digital mammography pictures has sparked new research opportunities since machine learning and DL techniques and resources can identify numerous previously unknown domains. We improve performance by changing the network design and specifications. The proposed approach comprises the CNN model paired with its hardware deployment, which can improve classification efficiency. The testing accuracy of the proposed technique on the MIAS repository is 92.1%, 96.8% on INbreast, and 98.2% on the DDSM repository. In comparison to prior research, the suggested strategy has improved classification accuracy. This study may be expanded to evaluate 3D mammograms as well as classify and categorise other medical scans of other prevalent carcinoma.

Funding This work was funded by Xilinx Women In Technology Fall Grant 2021.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Hassan RO, Mostafa H (2021) Implementation of deep neural networks on FPGA-CPU platform using Xilinx SDSOC. Analog Integr Circuits Signal Process 106:399–408. https://doi.org/10.1007/s10470-020-01638-5
2. Sze V, Chen YH, Emer J et al (2018) (2018) Hardware for machine learning: challenges and opportunities. IEEE Cust Integr Circuits Conf CICC 2018:1–8. https://doi.org/10.1109/CICC.2018.8357072
3. Sze V, Chen YH, Yang TJ, Emer JS (2017) Efficient processing of deep neural networks: a tutorial and survey. Proc IEEE 105:2295–2329. https://doi.org/10.1109/JPROC.2017.2761740
4. Chen YH, Krishna T, Emer JS, Sze V (2017) Eyeriss: an energy-efficient reconfigurable accelerator for deep convolutional neural networks. IEEE J Solid-State Circuits 52:127–138. https://doi.org/10.1109/JSSC.2016.2616357
5. Hassan SA, Sayed MS, Abdalla MI, Rashwan MA (2020) Breast cancer masses classification using deep convolutional neural networks and transfer learning. Multimed Tools Appl 79:30735–30768. https://doi.org/10.1007/s11042-020-09518-w
6. Singh L, Alam A (2022) An efficient hybrid methodology for an early detection of breast cancer in digital mammograms.

J Ambient Intell Humaniz Comput. https://doi.org/10.1007/s12652-022-03895-w
7. Mughal B, Muhammad N, Sharif M (2019) Adaptive hysteresis thresholding segmentation technique for localizing the breast masses in the curve stitching domain. Int J Med Inform 126:26–34. https://doi.org/10.1016/j.ijmedinf.2019.02.001
8. Ding S, Zhao H, Zhang Y et al (2015) Extreme learning machine: algorithm, theory and applications. Artif Intell Rev 44:103–115. https://doi.org/10.1007/s10462-013-9405-z
9. Shah SM, Khan RA, Arif S, Sajid U (2022) Artificial intelligence for breast cancer analysis: trends & directions. Comput Biol Med 142:105221. https://doi.org/10.1016/j.compbiomed.2022.105221
10. Weiss K, Khoshgoftaar TM, Wang D (2016) A survey of transfer learning. J Big Data 3:9. https://doi.org/10.1186/s40537-016-0043-6
11. Hassan NM, Hamad S, Mahar K (2022) Mammogram breast cancer CAD systems for mass detection and classification: a review. Multimed Tools Appl 81:20043–20075. https://doi.org/10.1007/s11042-022-12332-1
12. Dhungel N, Carneiro G, Bradley AP (2017) A deep learning approach for the analysis of masses in mammograms with minimal user intervention. Med Image Anal 37:114–128. https://doi.org/10.1016/j.media.2017.01.009
13. Shen R, Yao J, Yan K et al (2020) Unsupervised domain adaptation with adversarial learning for mass detection in mammogram. Neurocomputing 393:27–37. https://doi.org/10.1016/j.neucom.2020.01.099
14. Ribli D, Horváth A, Unger Z et al (2018) Detecting and classifying lesions in mammograms with deep learning. Sci Rep 8:1–7. https://doi.org/10.1038/s41598-018-22437-z
15. Ren S, He K, Girshick R, Sun J (2017) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39:1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031
16. Shawahna A, Sait SM, El-Maleh A (2019) FPGA-based accelerators of deep learning networks for learning and classification: a review. IEEE Access 7:7823–7859. https://doi.org/10.1109/ACCESS.2018.2890150
17. Liu B, Zou D, Feng L, et al (2019) An FPGA-based CNN accelerator integrating depthwise separable convolution. Electron 8. https://doi.org/10.3390/electronics8030281
18. Liu Z, Chow P, Xu J, et al (2019) A uniform architecture design for accelerating 2d and 3d cnns on fpgas. Electron 8. https://doi.org/10.3390/electronics8010065
19. Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, Ricketts I et al (2015) Mammographic Image Analysis Society (MIAS) database v1.21. https://www.repository.cam.ac.uk/handle/1810/250394. Accessed Mar 2022
20. Moreira IC, Amaral I, Domingues I et al (2012) INbreast: toward a full-field digital mammographic database. Acad Radiol 19:236–248. https://doi.org/10.1016/j.acra.2011.09.014
21. Joseph AM, John MG, Dhas AS (2017) Mammogram image denoising filters: a comparative study. 2017 Conference on Emerging Devices and Smart Systems (ICEDSS), Mallasamudram, India, pp 184–1891. https://doi.org/10.1109/ICEDSS.2017.8073679
22. Ramachandran V, Kishorebabu V (2019) A tri-state filter for the removal of salt and pepper noise in mammogram images. J Med Syst 43. https://doi.org/10.1007/s10916-018-1133-0
23. Maria HH, Jossy AM, Malarvizhi G, Jenitta A (2021) Analysis of lifting scheme based double density dual-tree complex wavelet transform for de-noising medical images. Optik 241:2–3. https://doi.org/10.1016/j.ijleo.2021.166883
24. Jang S, Liu W, Cho Y (2022) Convolutional neural network model compression method for software—hardware co-design. Information 13(10):451. https://doi.org/10.3390/info13100451
25. https://www.tensil.ai/. Accessed Mar 2022
26. Hu X, Wen S, Lam HK (2022) Dynamic random distribution learning rate for neural networks training. Appl Soft Comput 124:109058. https://doi.org/10.1016/j.asoc.2022.109058
27. Wang SH, Lv YD, Sui Y, et al (2018) Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling. J Med Syst 42. https://doi.org/10.1007/s10916-017-0845-x
28. Rahangdale A, Raut S (2019) Deep neural network regularization for feature selection in learning-to-rank. IEEE Access 7:53988–54006. https://doi.org/10.1109/ACCESS.2019.2902640
29. Ting FF, Tan YJ, Sim KS (2019) Convolutional neural network improvement for breast cancer classification. Expert Syst Appl 120:103–115. https://doi.org/10.1016/j.eswa.2018.11.008
30. Jiao Z, Gao X, Wang Y, Li J (2016) A deep feature based framework for breast masses classification. Neurocomputing 197:221–231. https://doi.org/10.1016/j.neucom.2016.02.060
31. Al-antari MA, Al-masni MA, Choi MT et al (2018) A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. Int J Med Inform 117:44–54. https://doi.org/10.1016/j.ijmedinf.2018.06.003
32. Salama WM, Aly MH (2021) Deep learning in mammography images segmentation and classification: automated CNN approach. Alexandria Eng J 60:4701–4709. https://doi.org/10.1016/j.aej.2021.03.048
33. Ragab DA, Sharkas M, Marshall S, Ren J (2019) Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ 2019:1–23. https://doi.org/10.7717/peerj.6201

Kayalvizhi R is a full-time research scholar pursuing her research in VLSI domain at SRM Institute of Science and Technology, Kattankulathur, Chennai. She is also deputed as Research Assistant for the Xilinx Women In Technology project.

Heartlin Maria H is a full-time research scholar pursuing her research in Deep Learning domain at SRM Institute of Science and Technology, Kattankulathur, Chennai. She is also deputed as Research Assistant for the Xilinx Women In Technology project.

Malarvizhi S is an awardee of the Xilinx Women In Technology Grant. She is a senior professor in the Department of Electronics and Communication Engineering at SRM Institute of Science and Technology, Kattankulathur, Chennai. She is an IEEE senior member. She is also the editor and reviewer of various reputed journals. She holds 4 patents and more than 90 publications to her name.

Revathi Venkatraman is an awardee of the Xilinx Women In Technology Grant. She is a Professor and Chairperson of the School of Computing, SRM Institute of Science and Technology, Kattankulathur. She has 6 patents and a vast range of publications to her name.

Shantanu Patil is Professor and Head of Department, Department of Translational Medicine and Research, Kattankulathur Campus, SRM Institute of Science and Technology. He heads the Innovation and Incubation Cell of SRM. He has numerous patents and publications to his name.