SANAS-Net: spatial attention neural architecture search for breast cancer detection

Melwin D'souza[1], Ananth Prabhu Gurpur[1], Varuna Kumara[2]
[1]Department of Computer Science and Engineering, Sahyadri College of Engineering and Management, Mangaluru, India
[2]Department of Electronics and Communication Engineering, Moodlakatte Institute of Technology, Kundapura, India

Article Info

Article history:

Received Nov 16, 2023
Revised Feb 13, 2024
Accepted Feb 28, 2024

Keywords:

Breast cancer detection
Deep learning
Mammography
Neural architecture search
Spatial attention

ABSTRACT

The utilization of mammography images plays a vital role in the prompt detection and treatment of breast cancer. Breast imaging techniques aid medical professionals in assessing the dimensions, morphology, and spatial orientation of breast lesions, facilitating the differentiation between benign and malignant conditions. Breast tissue can vary widely in terms of density, composition, and structure, leading to complexities in distinguishing between benign and malignant conditions. The primary contribution of this paper is the proposal of a spatial attention-based neural architecture search network (SANAS-Net) technique that incorporates a spatial attention mechanism, enabling the model to learn and prioritize key regions within mammograms (MMs). Multi-head attention is employed within the transformer blocks to effectively capture a wide range of spatial relations and feature interactions. Global contextual information was integrated into the transformer blocks by means of introducing positional embeddings. Several practical studies have been undertaken to verify the effectiveness of our methodology in identifying fully attentive networks that exhibit good performance in distinguishing between malignant and benign breast cancer cases. The experimental study reached a test accuracy of 89.95%, which is way higher than previously proposed algorithms for mammography image-based breast cancer detection.

Corresponding Author:

Melwin D'souza
Department of Computer Science and Engineering, Sahyadri College of Engineering and Management
Mangaluru, India
Email: mellumerdy@gmail.com

1. INTRODUCTION

Breast cancer is a prevalent and highly lethal malignancy that affects women on a global scale. The condition has the potential to impact females across all age groups, although the likelihood of occurrence escalates with advancing age and in the presence of specific genetic and environmental influences. The timely identification of breast cancer is a crucial measure in mitigating its advancement and enhancing the likelihood of effective intervention and long-term survival. Women can effectively discover any changes or abnormalities in their breasts promptly by engaging in routine self-examinations, undergoing mammography and clinical breast examinations, and maintaining awareness of the signs and symptoms associated with breast cancer [1]. The timely identification of a medical condition can also contribute to a decrease in the necessity for more intrusive and forceful treatment methods, such as chemotherapy, radiation therapy, or mastectomy. Hence, it is imperative that women get education and empowerment to assume responsibility for their breast health and promptly seek medical intervention upon detecting any potentially concerning alterations. The identification of breast cancer at an early stage is not only a pivotal measure for managing and treating the disease, but also a potentially life-saving intervention with the potential to positively impact a significant number of women globally.

Breast tumors are typically classified into four distinct categories, namely normal, benign, in situ carcinoma, and invasive carcinoma. A benign tumor is characterized by the presence of aberrant cells that exhibit slight structural abnormalities, yet lack the ability to be classified as malignant cells. Nevertheless, it is important to note that both in situ carcinoma and invasive carcinoma are categorized as forms of cancer. The timely identification of breast cancer is a crucial factor in the management of the disease, as it plays a decisive role in preventing its progression and associated problems. There exist multiple widely recognized imaging modalities for the early detection and treatment of breast cancer, such as mammograms (MMs), breast thermography (BTD), magnetic resonance imaging (MRI), positron emission tomography (PET), computed tomography (CT), ultrasound (US), and histopathology (HP) [2]. MMs and HP are two often employed modalities in medical imaging and analysis. These techniques entail the examination of tissue samples that have been stained with hematoxylin and eosin, a method that enhances the visibility of cellular structures. The utilization of MMs and HP has been extensively documented in scientific literature. Mammography aims to screen a vast population for early indications of breast cancer [3], whereas HP endeavors to obtain high-resolution microscopic pictures in order to identify precise malignant tissues at the molecular level. In the context of breast cancer screening, radiologists and pathologists engage in the manual observation and examination of breast images to facilitate the process of diagnosing the condition, determining its prognosis, and making decisions on therapy. The process of screening often results in either excessive or insufficient therapy due to the presence of erroneous detection, hence leading to a protracted diagnosis procedure.

The utilization of deep learning for the identification of breast cancer has garnered considerable interest in recent times, primarily due to its capacity to enhance the precision and effectiveness of breast cancer diagnosis. The process of breast cancer detection includes the utilization of screening mammography, which are X-ray images of the breast, in order to discover potentially malignant lesions that may serve as indicators of the presence of cancer. Numerous studies have demonstrated that deep learning models exhibit exceptional proficiency in the identification of breast cancer, outperforming conventional methodologies and even human radiologists in terms of performance. An illustrative instance involves a recent investigation conducted jointly by Google Health and Imperial College London, wherein a deep learning model was employed to assess a dataset comprising more than 76,000 MMs from the UK and over 15,000 MMs from the USA. The implemented model demonstrated a reduction in the false positive rate of 5.7% in the UK samples and 9.4% in the USA samples, as well as a reduction in the false negative rates for both cases, when compared to the performance of human experts [4]. This particular domain is a dynamic and progressive subject of study that necessitates additional inquiry and verification. The utilization of deep learning techniques holds promise for enhancing the precision and efficacy of breast cancer diagnosis. However, it is imperative to acknowledge and confront the obstacles and constraints that could potentially undermine its dependability and credibility.

Neural architecture search network (NAS-Net) is a deep learning technique that enables the automated creation of neural architectures [5]. Several designs generated using neural architecture search (NAS) approaches have demonstrated superior accuracy compared to manually created structures for several applications, including image classification, semantic segmentation, and object detection [6]. These techniques can enhance the model's performance and alleviate human experts from the laborious process of manually adjusting the detection model. The main contributions of this paper are:

– An spatial attention-based NAS-Net method is proposed that develop a spatial attention mechanism that learns to focus on relevant regions within MMs.
– Multi-head attention is used within the transformer blocks to capture diverse spatial relations and feature interactions.
– Global contextual information is incorporated via introducing positional embeddings to the transformer blocks.
– Numerous empirical investigations have been conducted to validate the efficiency of our approach in identifying high-performing fully attentive networks for the identification of both malignant and benign breast cancer.

2. LITERATURE REVIEW

When comparing the healthcare sector to other domains, it may be argued that the decision-making process holds greater significance within healthcare systems due to its direct impact on individuals' well-being. An erroneous decision made by a medical practitioner in the process of identifying a medical condition has the potential to result in the fatality of a patient. The physician's decision-making process is significantly challenged in complex and confined clinical contexts and workflows, particularly when it comes to image-related tasks that necessitate a high level of visual perception and cognitive aptitude. In such circumstances, artificial intelligence (AI) can serve as a valuable tool for reducing the occurrence of incorrect diagnoses. This can be achieved through the extraction of distinct and established elements from medical images, or by assisting physicians with preliminary suggestions for potential solutions. In contemporary times, there is a growing trend among healthcare practitioners to adopt AI algorithms. This inclination stems from the increased accessibility of computing resources, advancements in image analysis tools, and the notable efficacy demonstrated by AI methodologies.

The shape and borders of a mass are key indicators in distinguishing between benign and malignant characteristics. The form has the potential to exhibit a circular, elliptical, lobed, or non-uniform configuration. Typically, benign tumors are characterized by their circumscribed oval or round shape. A non-uniform shape implies an increased probability of malignancy. Kooi *et al*. [7] employed a model based on a convolutional neural network (CNN) to extract features from mammography pictures. These features were subsequently inputted into a support vector machine (SVM) classifier. The method demonstrated a notable performance in lesion classification, with an area under the curve (AUC) of 86%. This is a significant improvement of approximately 6% compared to the leading conventional approach prior to the publication of this study. Several more research [8]-[10] have employed CNN-based algorithms for the purpose of lesion categorization. However, the articles in question retrieved the region of interest using traditional image processing methods [9] or with the assistance of an expert [10], rather than with a deep learning system. In their study, Kooi *et al*. [8] initially partitioned mammography images into smaller patches and applied a traditional image-processing approach to extract features from these patches. Subsequently, they employed a random forest classifier to choose high-quality candidate patches for their CNN algorithm. The methodology employed by the researchers yielded an AUC of 92.9%, demonstrating a modest improvement compared to the baseline method, which achieved an AUC of 91%. Due to the progress made in deep learning algorithms and the accessibility of intricate and robust deep learning structures, deep learning techniques have been employed to extract regions of interest from whole multimodal images. Consequently, the algorithm's input is no longer limited to small patches, allowing for the utilization of the entire MM image as input. An illustration of this may be seen in the technique proposed by Dildar *et al*. [11], where the you only look once (YOLO) algorithm [12], a widely recognized algorithm for detection and classification, is employed to extract and categorize regions of interest in the entirety of the image concurrently. This approach enables the simultaneous extraction and classification of regions of interest across the whole image. The findings indicate that the approach demonstrates comparable performance to a CNN model trained on tiny patches, achieving an AUC of 97%.

Upon doing a comprehensive examination of the existing body of research, it becomes evident that there are numerous constraints and difficulties associated with the spatial analysis of mammography image-based breast cancer detection. The spatial resolution and accuracy of mammographic images are frequently affected by many factors such as noise, abnormalities, and compression [13]. Furthermore, there is a lack of comprehensive understanding regarding the geographical patterns and relationships of mammographic characteristics, which may exhibit variations across many populations and contexts. Hence, it is imperative to develop and employ more sophisticated and resilient spatial methodologies and procedures to augment the precision and dependability of breast cancer detection using mammography picture analysis. Hence, the problem statement for this study are as follows.

- Identifying and precisely localizing subtle and early-stage abnormalities in breast tissue can be challenging, especially in the presence of overlapping or complex structures. In this concern, the main motivation of this paper is to identify the abnormalities in mammography-based breast tissue images which are present in overlapping complex structures.
- Mammogram images in the dataset may exhibit variations in quality, such as differences in resolution, orientation, and compression artifacts, which can affect the model's ability to extract relevant features. Therefore, we are going address the image quality-based issues in this paper.

3. ATTENTION-BASED ARCHITECTURE

3.1. Neural architecture search-based spatial attention module

This section presents a novel NAS method that aims to tackle the issues associated with developing a specialized attention network for breast cancer diagnosis. The objective of neural architecture search is to autonomously identify one or more neural network architectures that can generate models with favorable outcomes within a reasonable timeframe, when applied to a specific dataset [14]. The utilization of the NAS technique establishes an architectural framework known as the search space pattern [15]. Figure 1, depicts various stages of NAS-based framework. The search strategy encompasses the temporal aspect of model creation and is contingent upon the selection of search methods employed to define a NAS approach, which may include Bayesian optimization or reinforcement learning. Figure 2 presents the search procedure of NAS-Net.

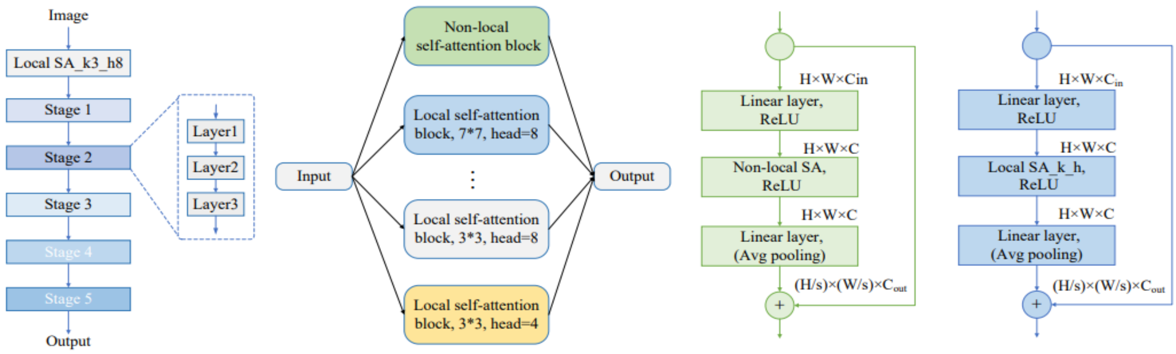


Figure 1. Various stages of NAS-based framework

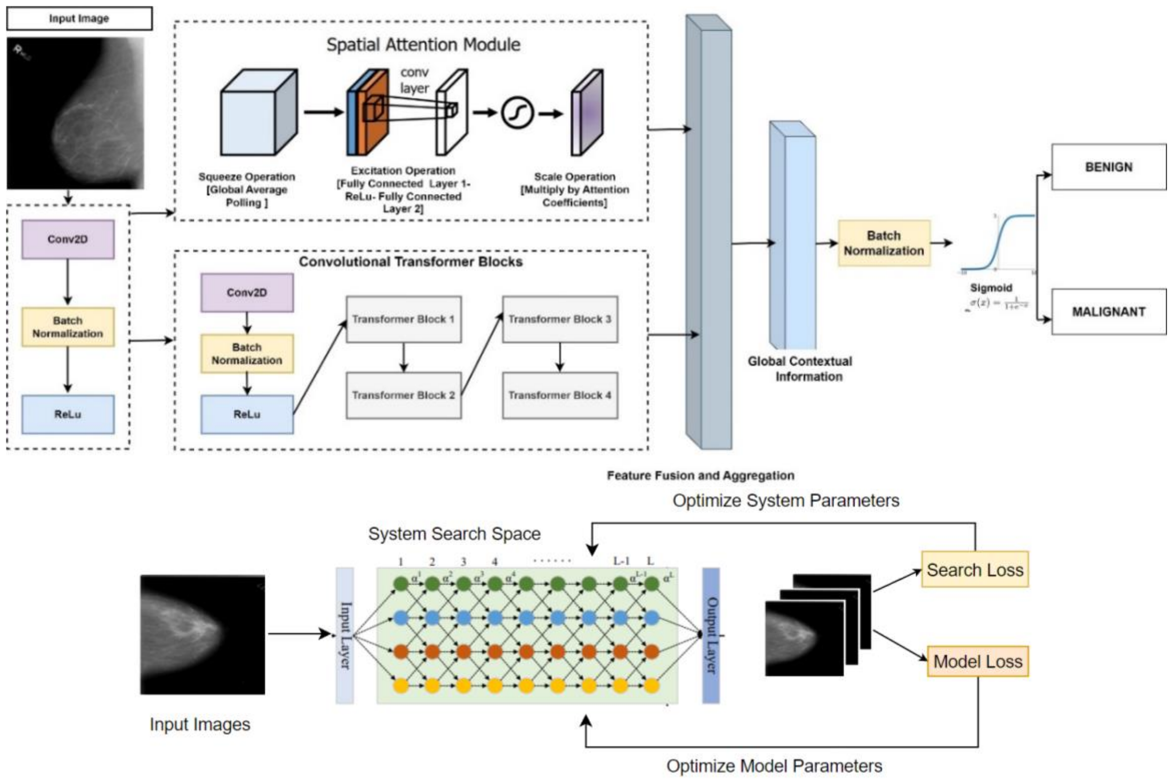


Figure 2. Proposed spatial attention-based neural architecture search network (SANAS-Net)

This paper introduces a sequential approach for exploring a comprehensive search space in order to identify a complete attention network. The model being investigated is divided into a series of predetermined steps that systematically decrease the spatial resolutions of the feature maps. The macro-architecture encompasses the determination of the layer count, input dimensions, and output channel number for each individual layer. Additionally, it delineates the specific layers inside the network that are to be explored. The initial and concluding layers of the network are characterized by predetermined operators. The remaining portion of the macro-architecture is comprised of five sequential phases, with each level containing three intermediate layers that can be searched. The technique of local multi-head self-attention is employed to extract spatial information from the immediate context [16]. Several heads are utilized to learn several unique representations of the input, similar to the concept of group convolutions.

Figure 2 presents the proposed SANAS-Net model architecture. The candidate operations encompass a total of seven self-attention operations. Based on the proposed self-attention procedures, two types of building blocks are devised as potential components of the macro-architecture. Each layer within the macro-architecture has the ability to select a block that is searchable. The non-local construction block consists of a linear layer, which is subsequently followed by a non-local self-attention operation, and then another linear layer. The initial linear layer modifies the input features and decreases their channel dimensionality in order to enhance computational efficiency. The ultimate linear layer is responsible for enlarging the features in order to align with the dimensionality of the output.

When the stride of the layer is set to two, a subsequent average pooling operation is performed after the last linear layer in order to decrease the spatial resolution. The rectified linear unit (ReLU) activation functions are applied after the initial linear layer and the attention operation. Furthermore, the shortcut connections inside the model serve the purpose of performing identity mapping. These connections add the outputs of the stacked layers to their own outputs.

The spatial attention technique in this study is designed to acquire the ability to selectively concentrate on pertinent locations within MMs. In this particular scenario, it is necessary to integrate convolutional and attention layers in order to effectively extract both local and global data. Spatial attention plays a crucial role in enabling the model to concentrate its resources on certain regions that are likely to contain harmful information. Optimizing the model parameters of the SANAS is a critical step in enhancing its performance for breast cancer classification, distinguishing between benign and malignant cases. Fine-tuning various hyperparameters, such as learning rates, batch sizes, and architectural configurations, to achieve the best possible model accuracy. Grid search is employed to explore the hyperparameter space systematically. Regularization methods, such as dropout is applied to prevent overfitting and enhance the model's ability to generalize to unseen data.

3.2. Convolutional transformer blocks

Transformers can be employed to process mammography data to effectively capture the interdependencies across various parts of the pictures [17]. This facilitates a more comprehensive comprehension of the intricate spatial patterns that are pertinent in the diagnosis of breast cancer and other medical disorders. We integrated transformer blocks alongside convolutional layers. The convolutional transformer blocks represent an innovative architectural approach in the realm of deep learning for image processing applications, including breast cancer diagnosis [18]. Convolutional transformer blocks leverage the strengths of CNNs and transformers, two influential models. The architecture of convolutional transformer blocks comprises a sequence of stacked blocks, wherein each block is composed of a convolutional layer followed by a transformer layer. The convolutional layer is responsible for extracting local features from the input image, whilst the transformer layer is designed to capture global dependencies among these features. By employing this approach, these blocks can acquire and process both spatial and semantic data from the image, which plays a vital role in achieving precise and resilient breast cancer diagnosis. Convolutional transformer blocks have the capability to undergo comprehensive training using conventional optimization techniques, enabling seamless integration with established deep-learning frameworks.

3.3. Feature fusion and aggregation

Our proposed novel mechanism utilizes the self-attention mechanism of the transformer by including fused characteristics derived from both the spatial attention module and the convolutional transformer blocks [19]. The approach involves a two-step process. Initially, a linear projection layer is applied to the output features of the spatial attention module and the convolutional transformer blocks [20]. This results in the generation of two sets of question, key, and value vectors. Next, the attention weights are calculated between the query vectors of one set and the key vectors of the other set. These weights are then utilized to aggregate the matching value vectors. By employing this approach, we may proficiently integrate the characteristics derived from both sources while retaining their spatial and semantic information intact. Subsequently, the amalgamated characteristics are inputted into an additional convolutional transformer block to undergo subsequent processing. Then we applied adaptive pooling to aggregate features effectively.

3.4. Parallel multi-head attention

The multi-head attention mechanism enables the model to acquire diverse attention patterns by considering several viewpoints [21]. The model comprises many parallel attention heads, each responsible for calculating a weighted average of the input information. The utilization of multi-head attention within the transformer blocks enables the comprehensive collection of varied spatial relations and feature interactions pertaining to mammography images. Recognizing and classifying breast lesions is advantageous due to the potential variations in their shapes, sizes, locations, and looks. The use of multi-head attention might also facilitate the model's attention toward vital locations of interest while disregarding extraneous background noise [22]. Transformer blocks consist of many attention and feed-forward layers, along with residual connections and layer normalization [23]. The utilization of these models facilitates the acquisition of intricate non-linear relationships between the input and output characteristics and the encoding of extensive dependencies spanning the entirety of the image. The construction of a robust deep neural network for the processing of mammography images can be achieved by employing a series of transformer blocks in a stacked configuration. We used parallel heads in this structure. The utilization of parallel heads enables the model to acquire the ability to focus on different elements of the input, including but not limited to syntax, semantics, and context. This enhancement enhances the performance and generalization capacity of the model.

3.5. Global contextual information

We incorporated global contextual information by introducing positional embeddings to the transformer blocks [24]. One of the primary difficulties encountered in the processing of mammography images is the task of effectively capturing the overall context of the breast tissue [25]. This is crucial since the breast tissue may exhibit subtle indications of either malignancy or benignity. Nevertheless, the majority of current methodologies heavily depend on local features or patches, potentially resulting in the omission of crucial information from remote locations. In order to tackle this matter, we suggest integrating global contextual information into the transformer blocks for mammography pictures by the introduction of positional embeddings. The constituent components of transformer blocks consist of self-attention layers, which possess the ability to acquire knowledge of distant relationships between the input tokens. The inclusion of positional embeddings enables the incorporation of spatial information pertaining to each token, hence augmenting the capacity of the transformer blocks for representation learning. The method is assessed on a comprehensive dataset of mammography images, demonstrating superior performance compared to existing methods in terms of accuracy, sensitivity, and specificity.

3.6. Proposed algorithm

The main concept of the proposed algorithm is discussed here. Let us take Y, which represents the actual labels, and $\hat{Y}$ indicates to the predicted labels. The learning rate (l) is the rate at which the model parameters (P) are updated. In the initial stage of this system, we start the search space for architectures. For each architecture in the search space trains the model on the breast cancer classification dataset. Then, the loss is calculated using the suitable loss function.

$$Loss = \frac{1}{n}\left\|Y - \hat{Y}\right\|^2 \tag{1}$$

Based on this calculated loss and evaluation results, the architecture parameters are updated accordingly. The equation for updated parameters is presented on the (2).

$$P(Updated) = \Delta P - l \times \Delta_P Loss \tag{2}$$

The attention is calculated as (3):

$$Attention\_head(i) = \mathrm{softmax}(\frac{\mathrm{Q*KT}}{sqrt(d\_k)})\mathrm{*V} \tag{3}$$

After the loss is calculated, the architecture is validated and updated using a validation dataset. Then, the model should be checked for convergence. If performance meets the desired criteria, then the loop should be broken. Otherwise, the algorithm is tested and rechecked. The basic algorithm of SANAS-Net with Transformer Parallel Heads is presented.

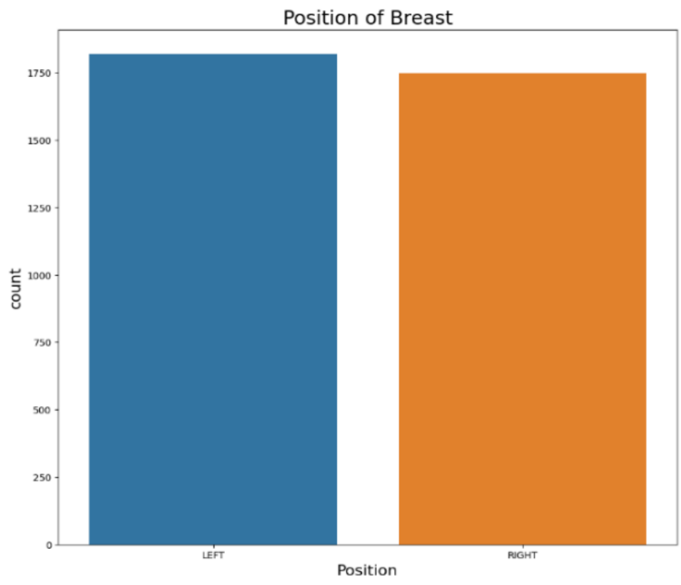
SANAS-Net algorithm with transformer parallel heads
```
Initialize NAS algorithm with search space for architectures
Initialize Transformer model with parallel heads
while not converged do:
    For each architecture in the search space do:
        Evaluate architecture performance using specific metrics
        loss = calculate_loss (actual_output, predicted_output)
    Update architecture parameters based on evaluation results
    architecture_parameters = update_architecture (architecture_parameters, learning_rate)
```

```
    For each parallel head in the Transformer model do:
        Optimize the parallel head
        attention_head_i = softmax (Q * K.T/sqrt(d_k)) * V
    Validate the updated architecture and parallel heads using validation datasets
    test_results = test_model(validation_data)
    If performance meets desired criteria, then:
        break
    End if
End while
```

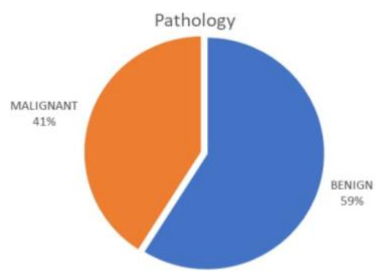
4. IMPLEMENTATION AND RESULTS

4.1. Dataset

Curated breast imaging subset of digital database for screening mammography (CBIS-DDSM) dataset [26] is utilized for this experimental study. The total number of images present in the dataset is 7638. We split the dataset into two different subsets. One of them is the training dataset which contains 6111 images. The other is the validation dataset with 1527 images. The DDSM is a valuable resource for the development and evaluation of decision support systems due to its extensive database and rigorous ground truth validation. 52% of the images contain calcification type abnormality. The rest of the 48% is mass-type abnormality. Breast calcifications refer to the formation of calcium deposits within the breast tissue. Breast lesions are frequently encountered and frequently detected during normal mammographic examinations. Although typically considered benign, breast calcifications may indicate an increased susceptibility to the development of breast cancer. Regular MMs have the capability to identify pre-cancerous alterations, hence enabling prompt initiation of treatment. Figure 3(a) shows the breast positions (left or right) in the mammography image dataset. Figure 3(b) shows the distribution of pathology types present in the dataset. The dataset contains 41% malignant type image, and 59% benign cases.



(a)



(b)

Figure 3. Data distribution (a) left and right positioned breast image in the dataset and (b) pathology types present in the dataset

4.2. Tuning and conduction of experiments

The experiments consist of various phases. Initially, the dataset is pre-processed according to the model requirements. Figure 4 shows the sample mammography images obtained from CBIS-DDSM dataset after pre-processing. Based on the architecture parameters obtained in the first phase, fine-tuning is performed on the CBIS-DDSM dataset to obtain the optimal spatial attention network. Afterward, batch normalization and activation functions are applied to enhance convergence and model stability. We utilized the classification head on top of the feature aggregation layer to determine the prediction outputs. These output classes include two classes: benign and malignant. We used the sigmoid function for binary classification. The binary cross-entropy loss was applied to train the model. An adaptive optimizer such as AdamW was implemented here for optimization [27]. Finally, the searched architecture is evaluated for multiple tasks. The studies were conducted using Google's Collaboratory Pro environment on a 64-bit Windows PC which had 25.5 GB of RAM and a graphics processing unit (GPU) with 15 GB of system memory. A storage capacity of up to 166.8 gigabytes (GB) was accessible. The implementation of these scripts uses the Keras framework, which is a Python toolset for deep neural network programming that is freely available and open-source.

The proposed architecture used almost 3.82 million parameters for this experimental study. The the procedure took 2 GPU days to obtain the results. Optimizing the architecture search space using transformer blocks may present challenges in effectively balancing the network's depth, width, and attention mechanisms, leading to difficulties in finding an optimal architecture for the given task.

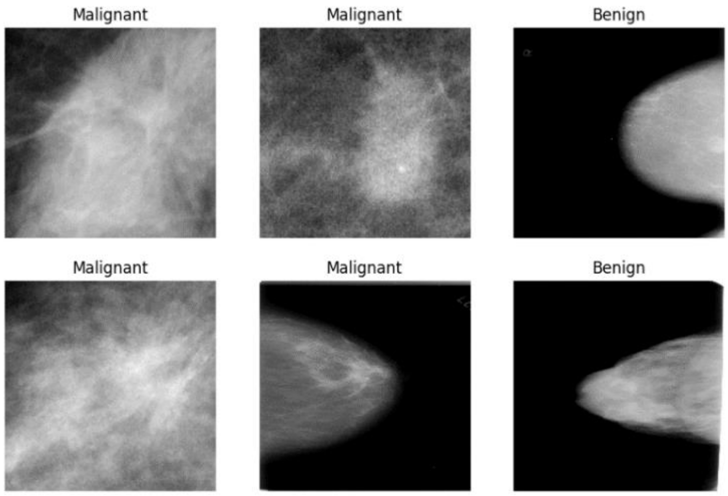
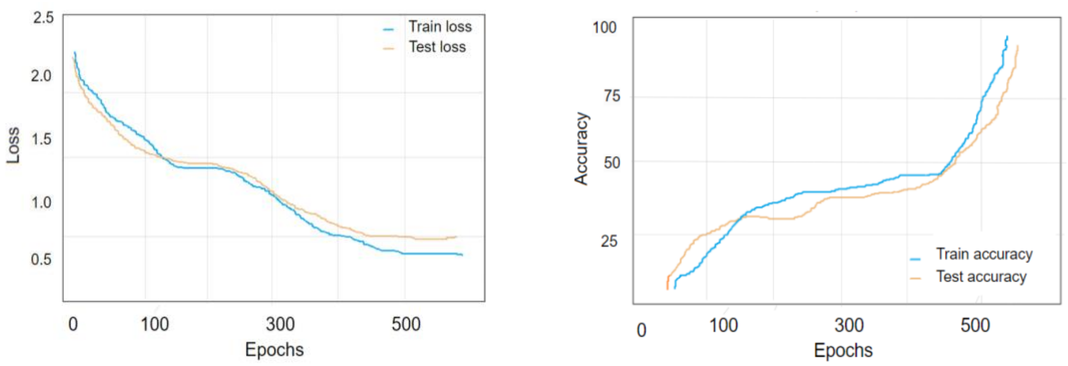


Figure 4. Sample mammography images from CBIS-DDSM dataset after pre-processing

5. RESULTS

After the test results are obtained, we need to calculate the loss output for this experimental study. The loss is calculated comparing the output results with actual outputs. After the loss is calculated, the breast cancer detection model is modified using those updated values and learning rate. In order to demonstrate the efficiency of our approach, we conducted training on a dataset consisting of 6,111 images from the CBIS-DDSM dataset. The initial network consisting of 112 channels undergoes training from the beginning for a total of 600 epochs. The training process utilizes the cosine learning rate schedule, with a base learning rate set at 0.04. The dataset undergoes standard data augmentation techniques, such as random scaled cropping, random horizontal flipping, and normalization. The network parameters are optimized using an AdamW optimizer, which incorporates a momentum value of 0.9 and a weight decay of 3×10-5. During the training process, a technique known as label smoothing regularization is applied with a coefficient of 0.1.

In order to provide equitable comparisons with alternative approaches, each image is shrunk to dimensions of 224×224. Figure 5 shows the output from the experimental study of spatial attention-based NAS-Net architecture, in Figure 5(a) represents the loss curve obtained during training and testing of proposed model. The loss curve gradually declined as the number of epochs increases this shows that proposed model has the ability to minimize the errors and improve its predictive accuracy. During testing, the loss curve is in the decreasing trend which shows that proposed model has successfully generalized the unseen data in the images. Figure 5(b) shows the training and testing accuracy curve obtained during experiment. The training accuracy curve is continuously increased which signifies that proposed model have ability to improve to predict the target values, similarly the testing accuracy curve shows increasing trend and slight decreasing and then again increased which signifies that model is able to overcome from the overfitting and learns the patterns and could achieve an overall accuracy of 89.95 percent. The evaluation results shown in Table 1 provide a comprehensive analysis and comparison of the current methodologies. In comparison to alternative methods, our model demonstrates superior performance with a significant advantage in terms of parameter efficiency. As the width of the network decreases, there is an observed improvement in accuracy with an increase in depth. Nevertheless, the increase in accuracy reaches a point of saturation as the model's width expands. The underlying belief is that a more intricate model has the ability to represent more intricate and multifaceted features. Nevertheless, the training of deeper networks is rendered more challenging as a result of the vanishing gradient problem.



(a) (b)

Figure 5. Output from the experimental study of SANAS-Net architecture: (a) loss curve for training and testing the proposed model and (b) train and test accuracy curves

Table 1. Comparison of the proposed approach results with other work

Papers and Authors	Models	Parameters (M)	Accuracy (%)
Falconi *et al.* [28]	VGG16-8-FT	138	84.00
Agarwal *et al.* [29]	ResNet50	23.5	83.69
	InceptionV3	24	84.16
Salama and Aly [30]	InceptionV3	24	84.21
	DenseNet121	7.6	82.47
	ResNet50	23.5	81.65
	MobileNetV2	3.4	79.82
The proposed model	SANAS-Net	3.82	89.95

6. CONCLUSION

The utilization of mammography photographs plays a significant role in the quick detection and treatment of breast cancer. Imaging methods for the breast let doctors evaluate the size, shape, and location of breast lesions, which in turn aids in determining whether or not the lesion is benign or cancerous. The task of accurately detecting and precisely determining the location of minor and early-stage anomalies in breast tissue can pose difficulties, mainly when dealing with overlapping or intricate structures. The primary objective of this study is to accurately pinpoint and identify anomalies within mammography-based breast tissue pictures that are characterized by the presence of overlapping complex systems. In this research, we propose a new SANAS-Net method that uses a spatial attention mechanism to better understand MMs and highlight their most essential parts for analysis. The transformer blocks use multi-head attention to accurately record a wide variety of spatial linkages and feature interactions. Global contextual information was incorporated into the transformer blocks by incorporating positional embeddings. Several empirical investigations have been conducted to validate our methodology's ability to pinpoint completely attentive networks with high performance in classifying breast cancer as malignant or benign. A train accuracy of 94% and a test accuracy of 89.95% were achieved in the experimental investigation, which is significantly higher than that of previously proposed algorithms for breast cancer identification based on mammography images. One potential drawback of this architectural design is that spatial attention may not effectively encompass all the nuanced characteristics that hold diagnostic significance, such as microcalcifications, architectural distortions, or asymmetries. An additional constraint pertains to the potential impact of image quality and resolution on spatial attention, as these factors can fluctuate based on the mammography equipment and settings employed. So, we should consider these limitations while modifying the architecture in the future.

ACKNOWLEDGEMENTS

This work was supported by SCEM Mangaluru and Management, Staff and Principal of Moodlakatte Institute of Technology, Kundapura, India.

REFERENCES

[1] M. Rawashdeh *et al.*, "Knowledge, attitude and practice regarding clinical and self breast examination among radiology professionals," *Breast Cancer Management*, vol. 7, no. 3, Aug. 2018, doi: 10.2217/bmt-2018-0014.

[2] N. M. U. Din, R. A. Dar, M. Rasool, and A. Assad, "Breast cancer detection using deep learning: Datasets, methods, and challenges ahead," *Computers in Biology and Medicine*, vol. 149, 2022, doi: 10.1016/j.compbiomed.2022.106073.

[3] J. V. Fiorica, "Breast cancer screening, mammography, and other modalities," *Clinical Obstetrics and Gynecology*, vol. 59, no. 4, pp. 688–709, 2016, doi: 10.1097/GRF.0000000000000246.

[4] J. G. Nascimento, G. L. Menegaz, and G. Guimaraes, "Use of artificial intelligence techniques, correlation of thermal images and the concept of thermal impedance aiming to estimate the location and size of breast tumours," in *ASME 2023 Heat Transfer Summer Conference*, American Society of Mechanical Engineers, Jul. 2023, doi: 10.1115/HT2023-107545.

[5] V. -K. V. -Ho, K. Yamazaki, H. Hoang, M. -T. Tran, and N. Le, "Neural architecture search for medical image applications," in *Meta Learning With Medical Imaging and Health Informatics Applications*, Elsevier, 2023, pp. 369–384, doi: 10.1016/B978-0-32-399851-2.00029-6.

[6] J.-S. Kang, J. Kang, J.-J. Kim, K.-W. Jeon, H.-J. Chung, and B.-H. Park, "Neural architecture search survey: a computer vision perspective," *Sensors*, vol. 23, no. 3, Feb. 2023, doi: 10.3390/s23031713.

[7] T. Kooi *et al.*, "A comparison between a deep convolutional neural network and radiologists for classifying regions of interest in mammography," in *Lecture Notes in Computer Science*, vol. 9699, 2016, pp. 51–56, doi: 10.1007/978-3-319-41546-8_7.

[8] T. Kooi *et al.*, "Large scale deep learning for computer aided detection of mammographic lesions," *Medical Image Analysis*, vol. 35, pp. 303–312, Jan. 2017, doi: 10.1016/j.media.2016.07.007.

[9] R. K. Samala, H.-P. Chan, L. M. Hadjiiski, K. Cha, and M. A. Helvie, "Deep-learning convolution neural network for computer-aided detection of microcalcifications in digital breast tomosynthesis," in *Medical Imaging 2016: Computer-Aided Diagnosis*, SPIE, Mar. 2016, doi: 10.1117/12.2217092.

[10] B. Q. Huynh, H. Li, and M. L. Giger, "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging*, vol. 3, no. 3, Jul. 2016, doi: 10.1117/1.jmi.3.3.034501.

[11] M. Dildar *et al.*, "Skin cancer detection: a review using deep learning techniques," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10, May 2021, doi: 10.3390/ijerph18105479.

[12] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of Yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2021, doi: 10.1016/j.procs.2022.01.135.

[13] T. Mahmood, J. Li, Y. Pei, and F. Akhtar, "An automated in-depth feature learning algorithm for breast abnormality prognosis and robust characterization from mammography images using deep transfer learning," *Biology*, vol. 10, no. 9, Sep. 2021, doi: 10.3390/biology10090859.

[14] P. Ren *et al.*, "A comprehensive survey of neural architecture search," *ACM Computing Surveys*, vol. 54, no. 4, pp. 1–34, May 2022, doi: 10.1145/3447582.

[15] M. Lindauer and F. Hutter, "Best practices for scientific research on neural architecture search," *Journal of Machine Learning Research*, vol. 21, 2020.

[16] R. Pecoraro, V. Basile, and V. Bono, "Local multi-head channel self-attention for facial expression recognition," *Information*, vol. 13, no. 9, Sep. 2022, doi: 10.3390/info13090419.

[17] X. Chen *et al.*, "Transformers improve breast cancer diagnosis from unregistered multi-view mammograms," *Diagnostics*, vol. 12, no. 7, Jun. 2022, doi: 10.3390/diagnostics12071549.

[18] J. Guo *et al.*, "CMT: convolutional neural networks meet vision transformers," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jun. 2022, pp. 12165–12175, doi: 10.1109/CVPR52688.2022.01186.

[19] J. Ni, H. Sun, J. Xu, J. Liu, and Z. Chen, "A feature aggregation and feature fusion network for retinal vessel segmentation," *Biomedical Signal Processing and Control*, vol. 85, Aug. 2023, doi: 10.1016/j.bspc.2023.104829.

[20] G. Yang, H. Tang, M. Ding, N. Sebe, and E. Ricci, "Transformer-based attention networks for continuous pixel-wise prediction," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2021, pp. 16249–16259, doi: 10.1109/ICCV48922.2021.01596.

[21] J. Li, X. Wang, Z. Tu, and M. R. Lyu, "On the diversity of multi-head attention," *Neurocomputing*, vol. 454, pp. 14–24, Sep. 2021, doi: 10.1016/j.neucom.2021.04.038.

[22] Y. Zhang, Y. Gong, H. Zhu, X. Bai, and W. Tang, "Multi-head enhanced self-attention network for novelty detection," *Pattern Recognition*, vol. 107, Nov. 2020, doi: 10.1016/j.patcog.2020.107486.

[23] R. Xiong *et al.*, "On layer normalization in the transformer architecture," in *37th International Conference on Machine Learning, ICML 2020*, 2020, pp. 10455–10464.

[24] T. Shen, K. Hao, C. Gou, and F.-Y. Wang, "Mass image synthesis in mammogram with contextual information based on GANs," *Computer Methods and Programs in Biomedicine*, vol. 202, Apr. 2021, doi: 10.1016/j.cmpb.2021.106019.

[25] D. Abdelhafiz, C. Yang, R. Ammar, and S. Nabavi, "Deep convolutional neural networks for mammography: Advances, challenges and applications," *BMC Bioinformatics*, vol. 20, Jun. 2019, doi: 10.1186/s12859-019-2823-4.

[26] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific Data*, vol. 4, no. 1, Dec. 2017, doi: 10.1038/sdata.2017.177.

[27] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *7th International Conference on Learning Representations, ICLR 2019*, pp. 1-18, 2019.

[28] L. G. Falconi, M. Perez, W. G. Aguilar, and A. Conci, "Transfer learning and fine tuning in breast mammogram abnormalities

classification on CBIS-DDSM database," *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 2, pp. 154–165, 2020, doi: 10.25046/aj050220.

[29] R. Agarwal, O. Diaz, X. Lladó, M. H. Yap, and R. Martí, "Automatic mass detection in mammograms using deep convolutional neural networks," *Journal of Medical Imaging*, vol. 6, no. 3, Jul. 2019, doi: 10.1117/1.jmi.6.3.031409.

[30] W. M. Salama and M. H. Aly, "Deep learning in mammography images segmentation and classification: automated CNN approach," *Alexandria Engineering Journal*, vol. 60, no. 5, pp. 4701–4709, Oct. 2021, doi: 10.1016/j.aej.2021.03.048.

BIOGRAPHIES OF AUTHORS

Melwin D'souza is an author and researcher in the field of Computer Science and Engineering. He earned his B.E. degree in Computer Science and Engineering and completed his M. Tech degree in 2013. His area of interests lies in the areas of deep learning and image processing. For any further contact or inquiries. He can be contacted at email: mellumerdy@gmail.com.

Ananth Prabhu Gurpur is an author and researcher in the field of Computer Science and Engineering. He earned his B.E. degree in Computer Science and Engineering and completed his M.Tech. in Computer Engineering from the Manipal University. He holds Ph.D. in Computer Engineering from VTU Belagavi. He further completed his postdoctoral studies from the University of Houston–Downtown, Texas. His area of interests lies in the areas of cyber security and image processing. For any further contact or inquiries. He can be contacted at email: educatorananth@gmail.com.

Varuna Kumara is an author and researcher in the field of artificial intelligence and machine learning. He earned his B.E. degree in Electronics and Communication in 2009 and completed his M.Tech. degree in 2012 from Srinivas Institute of Technology in Mangalore. Currently a research scholar at JAIN (Deemed to be University) in Bengaluru, India. His interests lie in the areas of soft computing, image processing, artificial intelligence, and machine learning. For any further contact or inquiries. He can be contacted at email: vkumarg.24@gmail.com.