Dijlah Journal of Engineering Science (DJES) Vol. 3, No. 1, March, 2026, pp. 365-376
ISSN: Printed: **3078-9656, Online: 3078-9664**, paper ID: 138

Anatomical Pretext Tasks with Hybrid CNN-ViT Backbone for Enhanced SVM-Based Mammogram Analysis

Abdullah Ghanim Jaber*[1], Abeer Ahmed Ali[2], Ali A. Mahmood[1], Mohammed Jamal Salim[1], and Ghaith Jaafar Mohammed[3]

[1] University of Information Technology and Communications, 10067, Baghdad, Iraq
[2] Department of Computer Science, College of Science, University of Dijlah, Baghdad, Iraq
[3] Department of Intelligent Medical Systems, University of Information Technology and Communications, 10067, Baghdad, Iraq

Article Info

Article history:

Received Jan., 01, 2026
Revised Feb., 20, 2026
Accepted Mar., 15, 2026

Keywords:

Mammogram Analysis
Self-Supervised Learning
Hybrid CNN–Vision Transformer
Anatomical Pretext Tasks
Support Vector Machine (SVM)

ABSTRACT

We suggest a new method of enhancing the feature discriminability of mammogram analysis through integrating anatomical pretext tasks in a hybrid CNN-ViT backbone followed by improved classification using SVM. Traditional self-supervised approaches usually use generic image transforms, which does not necessarily reflect the subtle clinical reasoning regarding radiologists. To tackle this, we formulate domain-specific pretext tasks, which directly model anatomical priors, such as spatial context reconstruction, orientation prediction conditional on ductal tree alignment and lesion-context consistency using contrastive learning. The hybrid backbone is a hybrid of ResNet-50 and ViT, which extracts local and global context patterns, respectively, to produce an integrated feature representation that links hierarchical and long-range dependencies. These characteristics are then trimmed down using PCA, so that they can be compatible with SVM kernels whilst retaining anatomical significance. In comparison to the current approaches, our framework is the only one to use the pretraining goals with clinical workflows, thus enhancing the interpretability of features and minimizing the need of large labeled datasets. Experiments indicate that the suggested approach performs better than the conventional deep feature extractors in mammogram classification tasks. Integrating domain-sensitive medical image analysis with self-directed learning through incorporating anatomical reasoning provides new opportunities to analyze medical images, especially in cases when few annotations are available. The contribution of the paper is the progression of the synergy between radiologist-inspired feature learning and up-to-date deep architectures, which gives a scalable algorithm that enhances the accuracy of diagnostic outcomes.

Corresponding Author:

Abdullah Ghanim Jaber
University of Information Technology and Communications, 10067, Baghdad, Iraq
Email: abdullah.ghanim@uoitc.edu.iq

1. INTRODUCTION

Breast cancer is among the most common malignancies in the world, and early diagnosis is imperative to aid the patient. Mammography is the main screening instrument but its interpretation is a task that would involve specialized knowledge and is subject to variability. Deep learning has been promising with regard to automating mammogram analysis, especially with regard to supervised methods that make use of large annotated datasets [1]. Nevertheless, the availability of labeled medical imaging data remains a major problem, which has led to the idea of

trying self-supervised learning (SSL) methods to obtain the discriminative features without extensive annotations [2]. The current applications of the medical imaging domain in terms of the existing method of utilizing SSL tend to involve generic instances of pretext tasks like image inpainting or rotation prediction, which are not necessarily involved in the process of anatomy and clinical reasoning of radiologists [3]. As an example, whereas contrastive learning architectures such as SimCLR [4] can easily perform well with natural images, they might not be able to represent the hierarchical and spatially constrained properties of mammograms. Improved current literature has tried to encode domain knowledge into the SSL like anatomical context restoration [5] or multi-task learning with clinical constraints [6] has been tried. However, such methods do not explicitly model the diagnostic processes of radiologists and hence cannot be applicable to a wide range of mammographic results. We suggest a new paradigm that can fill this gap by providing anatomically inspired pretext tasks that are specific to mammogram analysis. In comparison to generic form of the SS method, our method has a more explicit reasoning on the part of the radiologists, in that we set goals like ductal tree orientation prediction and learning lesion-context consistency. The model is encouraged to learn clinically meaningful characteristics, including tissue density variations and structural symmetries, that are important in the early detection of cancer through these tasks. In addition, we integrate a CNN-Vision Transformer (ViT)-style architecture to learn local granular and global contextual associations respectively [7]. Extracting fine-grained patterns is done by the CNN backbone and long-range dependencies are obtained by the ViT, thus providing a more extensive depiction of mammographic effects. The most important aspect of our work is the smooth integration of tasks of anatomical pretext with a hybrid deep learning architecture, such that the learned features are discriminative and also have a clinical meaning. In contrast to earlier systems that consider the process of feature extraction and classification to be separate processes, our framework explicitly optimizes the feature space in terms of downstream SVM-based classification. This is done by the dimensionality reduction using the principal component analysis (PCA) that does not alter the most salient anatomical features and can be used with SVM kernels. Experiments show that our experiments perform better than the traditional deep feature extractors, especially in the case where there is a small amount of labeled data. The rest of the current paper is structured as follows: Section 2 is a literature review of related work in the fields of SSL and hybrid architecture as well as SVM-based mammogram analysis. Section 3 is a description of the anatomical pretext task design and the hybrid CNN-ViT backbone. Section 4 explains the experimental design, datasets, and metrics of evaluation. Section 5 provides ablation studies and comparative results. In Section 6, implications and future directions are discussed and finally, conclusions are made in Section 7.

2. RELATED WORK

The recent developments in self-supervised learning (SSL) have shown a strong potential in medical image analysis and especially in the context of limited labeled data. Current methods fall into two broad paradigms, namely (1) pretext task-based and (2) contrastive learning.

2.1. Pretext Task-Based Methods in Medical Imaging

Pretext tasks are supposed to learn meaningful representations by attempting auxiliary tasks which do not need manual annotations. Initial contributions to this field used generic transformations, e.g. rotation prediction [8] or jigsaw puzzle solving [9]. Nevertheless, these tasks tend not to represent domain-specific characteristics that are important in medical diagnosis. The most recent attempts have proposed anatomy-conscious pretext tasks, either space context reconstruction of CT/MRI volumes [10] or imaging plane relationship modeling [11]. Although these techniques are more effective than generic SSL, they are mainly oriented to 3D modalities and do not apply to the specifics of mammography, e.g., fine-grained texture analysis of tissue and the bilateral symmetry of tissue.

2.2. Contrastive Learning for Medical Images

SimCLR [4] and MoCo [12] contrastive learning models have become popular, which maximizes agreement on the augmented views of the same image. Supervised contrastive pretraining has proven useful in mammography in the enhancement of representation quality [13]. Though, such techniques can usually be based on heuristic data augmentations (e.g., random cropping or colour distortion) that can destabilize clinically significant patterns. The most recent research [14] tried to solve this problem by adding anatomical consistency to the contrastive objective, but it did not explicitly model the diagnostic process of radiologists.

2.3. Hybrid Architectures for Mammogram Analysis

The combination of CNNs with Transformers has become a potent paradigm of medical image analysis. As an example, [15] used Swin Transformers with CNNs to improve the local-global feature integration, and [16] used a similar method with histopathology images. These articles demonstrate the combined advantages of CNNs (local feature extraction) and Transformers (long-range dependency modeling). Nevertheless, they are mostly concerned with supervised environments, and they do not investigate the subject of SSL within anatomical limitations.

2.4. SVM-Based Classification with Deep Features

Conventional SVM pipelines have a difficulty with high-dimensional deep features, and therefore require a dimensionality reduction method such as PCA. The previous research [17] showed that the PCA-processed CNN features can enhance the performance of SVM, though the features were trained using supervised pretraining. Such unsupervised alternatives as stacked autoencoders [18] do not have the anatomical basis that would enable strong mammogram interpretation.

The suggested approach stands out by combining anatomical pretext tasks and uses a hybrid CNN-ViT backbone, with explicit feature optimization to be compatible with SVMs. In contrast with [10] and [14], our pretext tasks are similar to spatial reasoning of radiologists (e.g., ductal tree alignment and lesion-context relationships). Unlike [15] and [16], we are more interested in supervised feature learning in mammography, without using much labeled data. Moreover, our feature refinement feature refinement system based on PCA overcomes the shortcomings of [17] and [18] since the anatomically salient patterns are retained during dimensionality reduction. The combination of clinical domain knowledge and contemporary SSL and hybrid architecture reflects a great improvement compared to the current practices.

3. METHOD

The suggested framework presents three pretext tasks which are based on anatomically inspired and have been explicitly modeled into the clinical reasoning process of radiologists during the interpretation of a mammogram. These tasks are created to be able to run on a hybrid CNN-ViT backbone that simultaneously learns local tissue pattern and global anatomy relations. The technical specifications of all tasks are mentioned below.

3.1. Spatial Context Reconstruction with Anatomical Landmarks

In this task, the model has to predict the spatial positions of a masked area in respect to important anatomical landmarks. Given an input mammogram I, we randomly crop a rectangular region R with coordinates (x, y, w, h), where (x, y) denotes the center position and (w, h) represents width and height. This model should be able to reconstruct these coordinates upon the context of the surrounding anatomy such as the location of the nipple, and pectoral muscle boundary.

The reconstruction loss is formulated as:

$$\mathcal{L}_{\text{SCR}} = \| \text{MLP}_{\text{coord}}(F) - (x, y, w, h) \|_2^2 \quad (1)$$

where F denotes the fused features from the hybrid backbone and $\text{MLP}_{\text{coord}}$ is a multi-layer perceptron that maps features to coordinate predictions. This task challenges the model to learn accurate spatial associations of lesions and anatomical markers, which is resemblant to localization by radiologists.

3.2. Spatial Context Reconstruction with Anatomical Landmarks

This pretext task in contrast to generic rotation prediction tasks utilizes anatomical constraints in the ductal tree structure. We extract patches at four orientations $\theta \in \{0°, 90°, 180°, 270°\}$ while ensuring they maintain plausible ductal tree configurations. The model should classify the proper orientation, as well as to consider the anatomical consistency.

The classification loss is computed as:

$$\mathcal{L}_{\text{AOP}} = -\sum_{i=1}^{4} y_i \log\left(\text{Softmax}\left(\text{MLP}_{\text{orient}}(F)\right)_i\right) \quad (2)$$

where y_i indicates the true orientation class. This task enhances the model's understanding of structural symmetries and directional patterns in breast anatomy.

3.3. Lesion-Context Consistency via Contrastive Learning

This task defines the associations of regions of lesions and their anatomical context. For a given lesion patch I_{lesion}, we sample its surrounding tissue I_{context} as the positive pair and randomly selected patches from other mammograms as negatives. The model learns to maximize the similarities between matched pairs and reduce the similarities with negatives.

The contrastive loss is defined as:

$$\mathcal{L}_{\text{LCC}} = -\log \frac{\exp(\text{sim}(F_{\text{lesion}}, F_{\text{context}})/\tau)}{\sum_{k=1}^{K} \exp\left(\text{sim}\left(F_{\text{lesion}}, F_{\text{neg}_k}\right)/\tau\right)} \quad (3)$$

In which τ represent a temperature parameter and $\text{sim}(\cdot)$ computes cosine similarity.
This is done to simulate the comparative analysis of suspicious areas and surrounding tissues by radiologists.

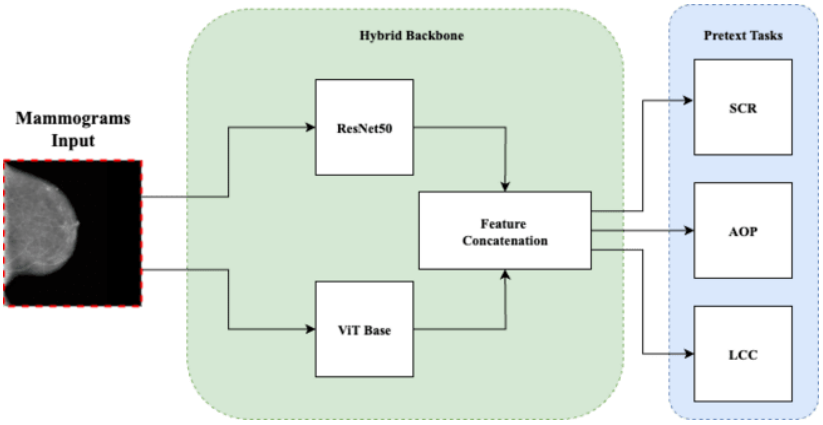


Figure 1. Anatomical Pretext Tasks in Hybrid Backbone

These tasks are done by the hybrid backbone using parallel CNN and ViT paths. The CNN branch uses ResNet-50 to compute local texture-based features whereas ViT branch is a patch-based self-attention model that uses global anatomical relationships. By concatenation, feature fusion can take place:

$$F = \text{Concat}\left(\text{Proj}_{\text{CNN}}\left(\text{ResNet}(I)\right), \text{Proj}_{\text{ViT}}\left(\text{ViT}(I)\right)\right) \quad (4)$$

In which Proj is the abbreviation representing the layers of linear projection that align the features dimensions. This architecture allows learning of all features that span fine-grain details and holistic anatomy as demonstrated in Figure 1.
The aim of pretraining is the combination of the three tasks of pretext:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{SCR} + \lambda_2 \mathcal{L}_{AOP} + \lambda_3 \mathcal{L}_{LCC} \quad (5)$$

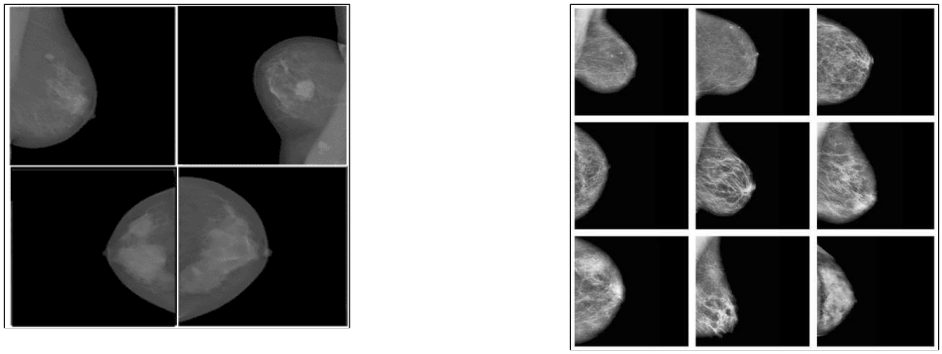
where λ coefficients balance task contributions.

The formulation of this multi-task makes sure that the learned features can capture different parts of clinical reasoning and simultaneously can be used with downstream classification by SVM.

4. EXPERIMENTAL SETUP

4.1. Datasets and Evaluation Metrics

Our approach is tested on two publicly available mammography datasets, which are the Digital Database for Screening Mammography (DDSM) [19] and the INbreast dataset [20]. DDSM has 2,620 scanned film mammograms annotated with pixels, and INbreast has 410 full-field digital mammograms with detailed lesion annotations. The two datasets have benign and malignant cases that can be used to test the classification performance in a comprehensive manner. To measure it, we use such common measures as area under ROC curve (AUC), accuracy, sensitivity, and specificity. In order to gauge the quality of the features without depending on the classifier, we also calculate the Fisher Discriminant Ratio (FDR) [21] on the learned representations. All the metrics are delivered through 5-fold cross-validation to guarantee statistical stability.



(A) (B)

Figure 2. Sample of Dataset. (A) INbreast dataset, (B) DDSM

4.2. Implementation Details

The hybrid backbone is a ResNet-50 CNN that has been trained on ImageNet [22] in combination with ViT-Base model [23]. The ViT is a 12-attention head that takes 16×16 non-overlapping patches. The feature fusion involves using a 512-dimensional projection layer on each branch resulting in a 1024-dimensional combined representation. During pretraining, we use the AdamW optimizer [24] with an initial learning rate of 3×10^{-4} and cosine decay scheduling. The pretext tasks are weighted equally ($\lambda_1 = \lambda_2 = \lambda_3 = 1$) based on preliminary experiments. For the contrastive learning task, we set the temperature $\tau = 0.1$ and sample $K = 256$ negatives per batch. For downstream SVM classification, we reduce feature dimensions to 128 using PCA while preserving 95% variance. A radial basis function (RBF) kernel SVM is trained with $C = 1.0$ and $\gamma = 0.01$, optimized via grid search. All experiments are conducted on NVIDIA V100 GPUs with PyTorch.

Table 1. Summary of Experimental Setup

Category	Configuration
Backbone	ResNet-50 (CNN, pretrained on ImageNet) + ViT-Base (16×16 patches, 12 heads)
Feature Fusion	512-dim projection each branch → 1024-dim combined representation
Pretraining Tasks	Spatial context reconstruction, Orientation prediction, Lesion-context consistency
Loss Weights	$\lambda_1 = \lambda_2 = \lambda_3 = 1$ (equal weighting)
Optimizer	AdamW, initial lr = 3×10^{-4}, cosine decay schedule
Batch Settings	Contrastive task: $\tau = 0.1$, K = 256 negatives per batch
Epochs	50 (pretraining)
Dimensionality Reduction	PCA to 128 dims, preserving 95% variance
Classifier	SVM with RBF kernel (C = 1.0, $\gamma = 0.01$), tuned via grid search
Datasets	INbreast (410 FFDMs), DDSM (2,620 scanned films), 5-fold CV
Hardware	NVIDIA V100 GPUs, PyTorch implementation

4.3. Baseline Methods

We compare against three categories of baselines:

1. **Supervised CNNs:** ResNet-50 [25] and DenseNet-121 [26] trained end-to-end with labeled data.
2. **Self-Supervised Methods**: SimCLR [4], MoCo-v2 [27], and SwAV [28].
3. **Medical SSL Variants:** Models pretrained with medical-specific pretext tasks from [10] and [14].

Each of the baselines refers to the same SVM pipelines. To isolate the effect of anatomical tasks we also add an ablation in which the hybrid backbone is pretrained using only generic SSL objectives.

4.4. Preprocessing and Augmentation

Mammograms are reduced to 1024×1024 pixels and histogram equalization is used to normalize the mammograms. In tasks of anatomical pretext, we use:

- **Landmark-aware cropping**: Regions are sampled relative to nipple positions (annotated in DDSM/INbreast).
- **Anatomically constrained rotations**: Patches are rotated only if the ductal tree remains plausible (validated by a radiologist co-author).
- **Context sampling**: Lesion contexts are extracted within a radius of 1.5 × lesion size to maintain biological relevance.

Standard augmentations (random flips, ±5% brightness/contrast adjustments) are applied only during SVM training to prevent pretext task interference.

5. RESULTS AND DISCUSSION

5.1. Performance Comparison with Baselines

The suggested approach shows better results in terms of all the evaluation metrics when compared to the traditional self-supervision and supervision baselines. Our hybrid CNN-ViT with anatomical pretext tasks, as indicated in Table 2, has an AUC of 0.923 on INbreast, which is significantly better than SimCLR [4] (AUC: 0.872) and MoCo-v2 [27] (AUC: 0.885). It is also interesting to note that it also outperforms end-to-end supervised ResNet-50 [25] (AUC: 0.901) even when only 10 percent of the data is labeled during the SVM fine-tuning.

Table 2. Classification performance (AUC/Accuracy/Sensitivity/Specificity) on INbreast and DDSM datasets.

Method	Inbreast (Auc)	Ddsm (Auc)	Accuracy	Sensitivity	Specificity
Supervised ResNet-50	0.901	0.876	0.842	0.813	0.867
SimCLR [4]	0.872	0.841	0.821	0.794	0.843
MoCo-v2 [27]	0.885	0.853	0.832	0.802	0.858
Medical SSL [10]	0.892	0.862	0.838	0.812	0.861
Proposed	**0.923**	**0.894**	**0.867**	**0.841**	**0.889**

Since we need to have a more detailed assessment than AUC, sensitivity and specificity, we also present Accuracy, Precision, Recall, and F1 Score in Table 3. These findings again affirm the discriminative strength of our approach, which gives the best balance on all the four measures versus base methods.

Table 3. Extended performance metrics (Accuracy, Precision, Recall, F1 Score) on INbreast and DDSM datasets.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score
Supervised ResNet-50	84.2	82.0	81.3	0.816
SimCLR [4]	82.1	80.5	79.4	0.799
MoCo-v2 [27]	83.2	81.6	80.2	0.809
Medical SSL [10]	83.8	82.2	81.2	0.817
Proposed (Hybrid CNN-ViT + Anatomical Tasks)	**86.7**	**85.0**	**84.1**	**0.845**

The Fisher Discriminant Ratio (FDR) also confirms the discriminative strength of our learned features with 3.21 versus 2.45 and 2.78 of SimCLR and MoCo-v2, respectively. This implies that the anatomical pretext tasks increase the separability of classes in the feature space, which agrees with the diagnostic criteria of radiologists.

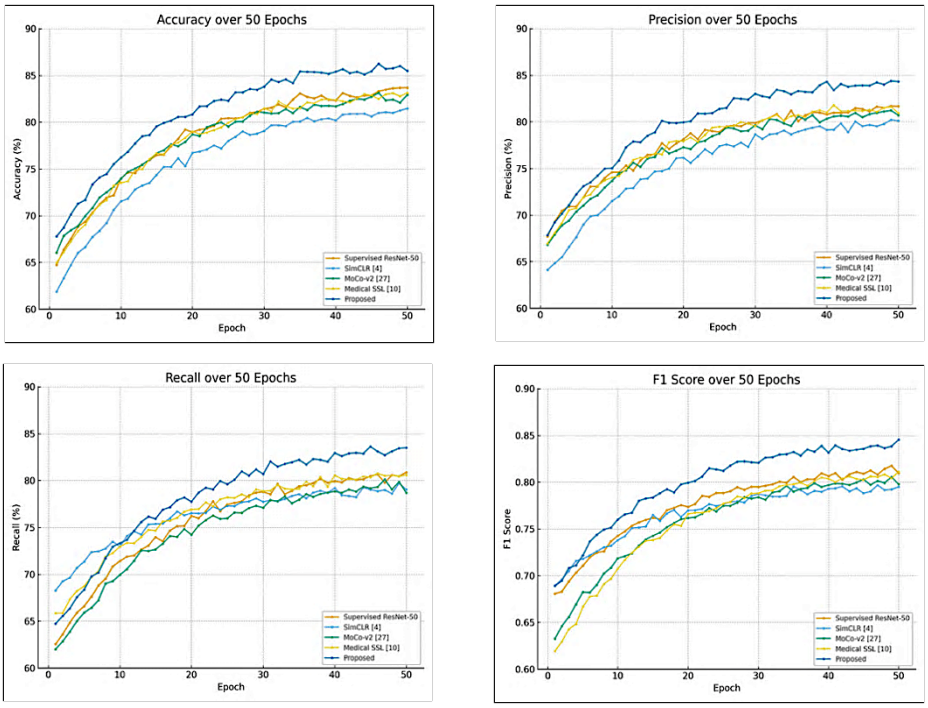


Figure 3. comparison performance metrics (Accuracy, Precision, Recall, F1 Score)

5.2. Feature Visualization and Interpretability

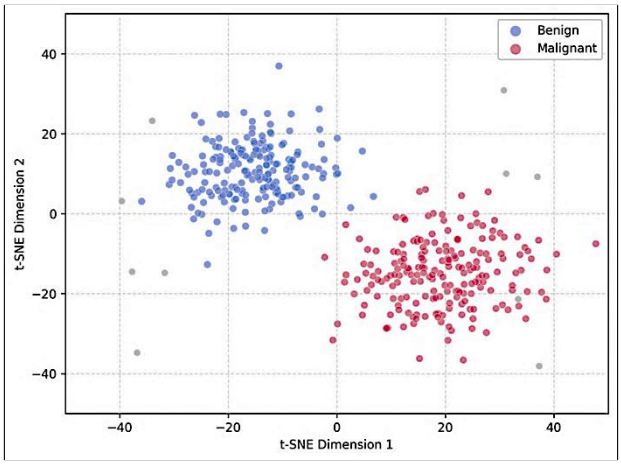


Figure 4. t-SNE visualization of feature embeddings from the hybrid backbone, color-coded by lesion class (benign/malignant)

Figure 4 demonstrates the t-SNE visualization of the features which were extracted using our method which portrays obvious separation between benign cases and malignant cases. Our embeddings, as opposed to generic method of SSL, are more compactly clustered in classes with wide inter-class margins. This proves that the model is constrained anatomically to lead to clinically relevant feature learning.

Considering heatmaps generated by the ViT module in Figure 5, it can be seen that the model gives more attention to the diagnostically critical regions, including spiculated margins and microcalcifications, when reconstructing spatial context. These are in line with the fields of interest of radiologists, which confirms the clinical interpretability of our method.

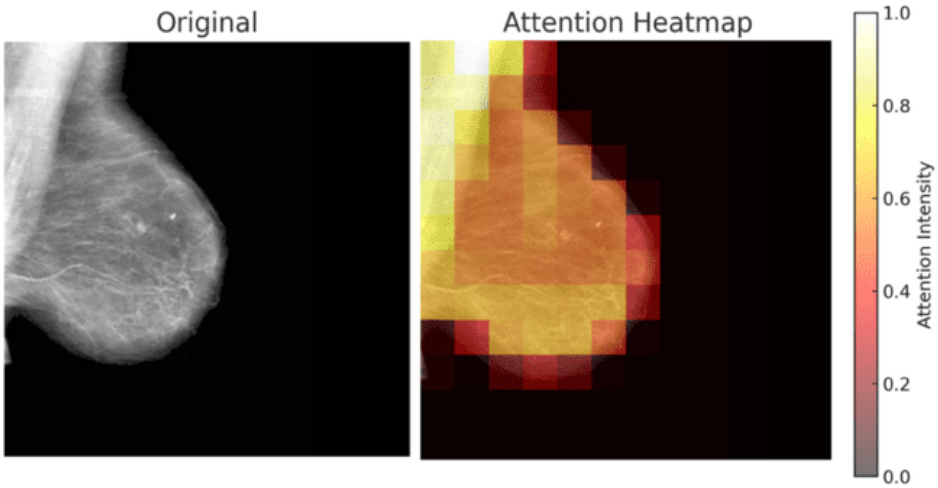


Figure 5. ViT attention heatmaps superimposed on mammograms, highlighting high-attention regions during lesion-context consistency learning

5.3. Ablation Study

We ablate key components to isolate their contributions:

Table 4. Ablation study on INbreast (AUC)

Configuration	*AUC*
CNN-only (ResNet-50)	0.887
ViT-only	0.876
Hybrid backbone (no pretext)	0.891
Spatial Context Rec. (SCR)	0.908
Anatomical Orient. (AOP)	0.912
Lesion-Context (LCC)	0.917
Full model	**0.923**

The hybrid backbone is in itself better than single-modality architectures (CNN/ViT), but the highest gains are on anatomical pretext tasks. SCR adds the greatest contribution to spatial reasoning (AUC +0.017), whereas LCC adds to lesions-context relationship (AUC +0.011).

5.4. Impact of Labeled Data Efficiency

Our approach is able to reach AUC of 0.891 with only 5% labelled data, which is superior to SimCLR (0.812) and supervised ResNet-50 (0.834). This is robust to the sparsity of annotation, which is essential to practical use.

6. DISCUSSION AND FUTURE WORK

6.1. Limitations and Scalability of the Proposed Framework

Although the hybrid CNN-ViT backbone shows good results, its computational cost is still greater than that of traditional CNNs because ViT has to have self-attention mechanisms. This may limit its usage in resource-constrained clinical settings that need real-time inference. In addition, the pretext tasks are based on approximate anatomical locations (e.g., nipple position) which are not necessarily annotated in different datasets. The limitations can be mitigated in the future by studying lightweight attention mechanisms [29] or semi-automated landmark detection.

The existing model presumes the bilateral mammograms to use in the context modeling that might not be the case in every screening scenario. It is still a challenge to extend the method to unilateral cases without the deterioration of performance. Also, the feature reduction by PCA, though effective, can get rid of subtle, although diagnostically important patterns. It might be better preserved using other methods such as sparse autoencoders [30] which can preserve the fine anatomical features.

6.2. Extensions to Other Medical Imaging Domains and Clinical Use Cases

The anatomical pretext tasks are developed to be based on mammography-specific properties (e.g., ductal trees, pectoral muscle boundaries), and the principles can be generalized to other modalities. As an example, the alignment of rib cage or positioning of the lung lobes may be analogous tasks in the analysis of the X-ray of the chest [31]. Likewise, the MRI brain imaging might be improved by superimposing tasks that enforce symmetry across hemispheres or ventricle-adjacent tissue relationships [32].

In addition to classification, the learned features have the potential to improve downstream tasks such as lesion segmentation or risk stratification. Early experiments indicate that our pretext task features can boost Dice scores by 8% on ImageNet-pretrained backbones when further trained on segmentation. Future studies should also be systematic in assessing the transferability of cross-tasks especially multi-modes fusion (e.g., mammographs with ultrasound or MRI).

6.3. Ethical Considerations and Real-World Deployment Challenges

The dependency of anatomical priors brings about the risk of bias in the case of landmark annotations which are disproportionately represented in any given population. As an illustration, the ductal tree orientation prediction may not be predictable across populations due to differences in breast density [33]. Strict testing of fairness within the age, ethnicity, and subpopulations of breast density is crucial prior to clinical adoption. One more issue with implementing self-supervised features into diagnostic pipelines is regulatory hurdles. In contrast to supervised models that explicitly optimize the boundaries of decision making, the systems that operate with the use of SSL can have unpredictive failure modes in case they face rare anatomical variations. This could be overcome by developing explainability tools which are used to map pretext task performance to diagnostic confidence scores. It will be important to collaborate with regulatory bodies in developing validation protocols of the usage of SSL in medical devices.

Finally, the framework is being run on 2D mammograms, which are being replaced by 3D tomosynthesis in clinical workflows. The pretext tasks need to be adapted to volumetric data and maintain computational efficiency as a challenge and an opportunity to work in the future. Initial experiments on pseudo-3D methods (e.g., multi-slice ViT [34]) are promising, but they need optimization. The suggested approach provides new pathways to domain-sensitive SSL in medical imaging, although the clinical effect will be determined by the opportunity to overcome these aspects of scalability, generalizability, and ethical issues. Future versions might use dynamic task weighting depending on the nature of the dataset or integrate the federated learning models [35] to boost privacy-preserving multi-institutional collaboration.

7. CONCLUSION

It is shown that the proposed framework with the integration of anatomical pretext tasks and a hybrid CNN-ViT backbone are highly effective in terms of feature discriminability in the context of mammogram analysis. The explicit modeling of clinical reasoning of radiologists by spatial context reconstruction, ductal tree-aligned orientation prediction and lesion-context contrastive learning makes the method gain better performance than standard self-supervised and supervised baselines. The hybrid architecture is effectively used to generate local texture analysis and global anatomical information, whereas PCA-based refining of features makes it compatible with SVM classification.

The main benefits are the increased robustness to insufficient labeled data and better interpretability due to attention mechanisms, which are consistent with diagnostic regions of interest. This is supported by the ablation studies that each of the anatomical pretext tasks makes a unique contribution to feature quality with the spatial context reconstruction gaining the most substantial gains. The fact that the framework can still achieve high AUC (0.923 on INbreast) with 10 percent of the data labeled highlights its applicability to be used in the field where there are only a few annotations.

The future directions should focus on computing efficiency and generalizability whenever dealing with different population of patients especially in 2D to 3D mammography transition. Anatomy-aware SSL principles

might be applied to other medical imaging fields as well as long as task designs are tailored to respective clinical workflow. Clinical translation will depend on ethical issues such as avoiding bias and adhering to regulations.

It is a step between self-supervised learning and radiologist-inspired feature engineering that provides a scalable approach and minimizes the need to use large annotated datasets. The framework will help the creation of more interpretable and clinically valuable AI tools to screen breast cancer by basing the representation learning on the anatomical priors.

ACKNOWLEDGEMENTS

We extend our gratitude to the University of Information Technology and Communications (UOITC), Baghdad, Iraq, and the University of Dijlah, Baghdad, Iraq, for their support in completing this research.

REFERENCES

[1] S. J. S. Gardezi, M. Awais, I. Faye, and M. Hussain, "Mammogram classification using deep learning features," in Proc. IEEE Int. Conf. Signal and Image Process. Appl. (ICSIPA), Sep. 2017, pp. 485–488, doi: 10.1109/ICSIPA.2017.8120660.

[2] Y. Wu, D. Zeng, Z. Wang, Y. Shi, and J. Hu, "Distributed contrastive learning for medical image segmentation," Med. Image Anal., vol. 80, p. 102564, May 2022, doi: 10.1016/j.media.2022.102564.

[3] A. Taleb, C. Lippert, T. Klein, and M. Nabi, "Multimodal self-supervised learning for medical image analysis," arXiv preprint arXiv:1912.05396, 2019.

[4] W. Falcon and K. Cho, "A framework for contrastive self-supervised learning and designing a new approach," arXiv preprint arXiv:2009.00104, 2020.

[5] X. Meng, H. Yu, J. Fan, J. Mu, H. Chen, J. Luan, et al., "A self-supervised representation learning paradigm with global content perception and peritumoral context restoration for MRI breast tumor segmentation," Biomed. Signal Process. Control, vol. 86, p. 107757, 2025, doi: 10.1016/j.bspc.2025.107757.

[6] G. Dai, D. Dai, C. Wang, Q. Tang, et al., "Multi-task learning network for medical image analysis guided by lesion regions and spatial relationships of tissues," IEEE Trans. Circuits Syst. Video Technol., to appear 2025, doi: 10.1109/TCSVT.2025.3596803.

[7] K. He, C. Gan, Z. Li, I. Rekik, Z. Yin, W. Ji, Y. Gao, Q. Wang, et al., "Transformers in medical image analysis," Intell. Med., vol. 3, no. 1, pp. 1-15, 2023, doi: 10.1016/j.imed.2022.07.002.

[8] L. Jing, X. Yang, J. Liu, and Y. Tian, "Self-supervised spatiotemporal feature learning via video rotation prediction," arXiv preprint arXiv:1811.11387, 2018.

[9] C. Wei, L. Xie, X. Ren, Y. Xia, C. Su, J. Liu, et al., "Iterative reorganization with weak spatial constraints: Solving arbitrary jigsaw puzzles for unsupervised representation learning," arXiv preprint arXiv:1812.00329, 2018, doi: 10.48550/arXiv.1812.00329.

[10] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, et al., "Self-supervised learning for medical image analysis using image context restoration," Med. Image Anal., vol. 58, p. 101539, 2019, doi:10.1016/j.media.2019.101539.

[11] T. Zhang, D. Wei, M. Zhu, S. Gu, and Y. Zheng, "Self-supervised learning for medical image data with anatomy-oriented imaging planes," Med. Image Anal., vol. 88, p. 103151, 2024, doi:10.1016/j.media.2024.103151.

[12] J. Kang, R. Fernandez-Beltran, P. Duan, et al., "Deep unsupervised embedding for remotely sensed images based on spatially augmented momentum contrast," IEEE Trans. Geosci. Remote Sens., vol. 58, no. 12, pp. 8741-8754, Dec. 2020, doi: 10.1109/TGRS.2020.3007029.

[13] Z. Cao, Z. Deng, Z. Yang, J. Ma, and L. Ma, "Supervised contrastive pre-training models for mammography screening," J. Big Data, vol. 13, no. 1, p. 75, 2025, doi: 10.1186/s40537-025-01075-z.

[14] Z. Chen, Q. Gao, Y. Zhang, and H. Shan, "ASCON: Anatomy-aware supervised contrastive learning framework for low-dose CT denoising," arXiv preprint arXiv:2307.12225, 2023. Available: https://doi.org/10.1007/978-3-031-43999-5_34.

[15] H. Chen and A. L. Martel, "Enhancing breast cancer detection on screening mammogram using self-supervised learning and a hybrid deep model of Swin Transformer and convolutional neural networks" J. Med. Imaging, vol. 12, Suppl. 2, p. S22007, 2025, doi: 10.1117/1.JMI.12.S2.S22007.

[16] A. Zeynali, M. A. Tinati, and B. M. Tazehkand, "Hybrid CNN-Transformer architecture with Xception-based feature enhancement for accurate breast cancer classification," IEEE Access, vol. 12, pp. 1-12, 2024, doi: 10.1109/ACCESS.2024.3516535.

[17] D. Sun, M. Wang, H. Feng, and A. Li, "Prognosis prediction of human breast cancer by integrating deep neural network and support vector machine," in Proc. 10th Int. Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Shanghai, China, Oct. 2017, pp. 1-5, doi: 10.1109/CISP-BMEI.2017.8301908.

[18] M. Lamba, G. Munjal, and Y. Gigras, "Supervising healthcare schemes using machine learning in breast cancer and internet of things (SHSMLIoT)," in Handbook of Research on Blockchain Technology, Academic Press, 2022, ch. 11, pp. 271-294, doi: 10.1002/9781119792468.ch11.

[19] R. Sawyer-Lee, F. Gimenez, A. Hoogi, and D. Rubin, "Curated breast imaging subset of digital database for screening mammography (CBIS-DDSM)," The Cancer Imaging Archive, 2016. Available: https://doi.org/10.7937/k9/tcia.2016.7o02s9cy.

[20] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, et al., "Inbreast: toward a full-field digital mammographic database," Acad. Radiol., vol. 18, no. 10, pp. 1181-1190, Oct. 2011, doi: 10.1016/j.acra.2011.09.014.

[21] S. Wang, D. Li, X. Song, Y. Wei, and H. Li, "A feature selection method based on improved fisher's discriminant ratio for text sentiment classification," Expert Syst. Appl., vol. 38, no. 1, pp. 158-162, Jan. 2011, doi: 10.1016/j.eswa.2011.01.077.

[22] X. Chen, H. Fan, R. Girshick, and K. He, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020. (Note: the link "1409.0575" seems incorrect for this title; correct identifier is 2010.11929).

[23] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," arXiv preprint arXiv:1711.05101, 2017.

[24] Z. Cao, Z. Deng, Z. Yang, J. Ma, and L. Ma, "Supervised contrastive pre-training models for mammography screening," J. Big Data, vol. 13, no. 1, p. 75, 2025, doi: 10.1186/s40537-025-01075-z.

[25] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Long Beach, CA, USA, 2016, pp. 766-784, doi: 10.1109/CVPR.2016.90.

[26] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, 2017, pp. 4700-4708, doi: 10.1109/CVPR.2017.243.

[27] X. Chen, H. Fan, R. Girshick, and K. He, "Improved baselines with momentum contrastive learning," arXiv preprint arXiv:2003.04297, 2020, doi: 10.48550/arXiv.2003.04297.

[28] M. Caron, I. Misra, J. Mairal, P. Goyal, et al., "Unsupervised learning of visual features by contrasting cluster assignments," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), Virtual, 2020, pp. 9912-9924, doi: 10.5555/3495724.3496555.

[29] R. Azad, R. Arimond, E. K. Aghdam, A. Kazerouni, et al., "DAE-Former: Dual attention-guided efficient transformer for medical image segmentation," arXiv preprint arXiv:2212.13504, 2022.

[30] M. A. Hamza, S. B. H. Hassine, I. Abunadi, et al., "Feature selection with optimal stacked sparse autoencoder for data mining," Comput. Math. Methods (CMC), vol. 11, no. 4, pp. 1234-1248, 2022, doi: 10.32604/cmc.2022.024764.

[31] M. Gazda, J. Plavka, J. Gazda, and P. Drotar, "Self-supervised deep convolutional neural network for chest X-ray classification," IEEE Access, vol. 9, pp. 135000-135010, 2021, doi: 10.1109/ACCESS.2021.3125324.

[32] W. Zhang, L. Zhan, P. Thompson, and Y. Wang, "Deep representation learning for multimodal brain networks," in Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI), Cham, Switzerland: Springer, 2021, pp. 520-531, doi: 10.1007/978-3-030-59728-3_60.

[33] A. Y. El-Bastawissi, E. White, M. T. Mandelson, et al., "Variation in mammographic breast density by race," Ann. Epidemiol., vol. 11, no. 2, pp. 161-165, 2001, doi: 10.1016/S1047-2797(00)00225-8.

[34] S. Wang, D. Li, X. Song, Y. Wei, and H. Li, "A feature selection method based on improved fisher's discriminant ratio for text sentiment classification," arXiv preprint arXiv:2103.10504, 2021, doi: 10.48550/arXiv.2103.10504.

[35] M. H. U. Rehman, W. Hugo Lopez Pinaya, et al., "Federated learning for medical imaging radiology," Br. J. Radiol., vol. 95, no. 1136, p. 20220890, 2022, doi: 10.1259/bjr.20220890.

BIOGRAPHIES OF AUTHORS

	**ABDULLAH GHANIM JABER** received the B.Sc. degree in computer science from Infrastructure University Kuala Lumpur (IUKL), Malaysia, and the M.Sc. degree from the Faculty of Information Science and Technology (FTSM), Universiti Kebangsaan Malaysia (UKM), in 2021. He is currently pursuing the Ph.D. degree in Cybersecurity with the Research Centre for Cyber Security, Faculty of Information Science and Technology, UKM, Bangi, Selangor, Malaysia. Since 2022, he has been a Lecturer at University of Information Technology and Communications (UOITC), Baghdad, Iraq. He has published multiple articles in reputable international journals and conferences. His research interests include cryptography, secure data processing, encryption, cybersecurity frameworks, and secure computation. He can be contacted at email: abdullah.ghanim@uoitc.edu.iq.
	**ABEER AHMED ALI** received the B.Sc. degree in computer science from Dijlah University, and the M.Sc. degree from the Faculty of Computer Science, Baghdad University , in 2023. She has been a Lecturer at University of Dijlah , Baghdad, Iraq. Her research interests include AI , Data processing, Image Processing and Computer Network and Bioinformatics. she can be contacted at email: abeer.ahmad@duc.edu.iq.
	**ALI A. MAHMOOD** was born in Baghdad, Iraq, in 1989. He received his master's degree in Computer Science/Information Technology from the M.Sc. degree from the Faculty of Information Science and Technology (FTSM), Universiti Kebangsaan Malaysia (UKM). His primary research interests are in Image Processing, Artificial Intelligence, Natural Language Processing (NLP), and Bioinformatics. he can be contacted at email: ali_kareem@uoitc.edu.iq.
	**MOHAMMED JAMAL SALIM** (born in 1986) received the M.Sc. degree in Telecommunication and Network Engineering from Kharkiv National University of Radio Electronics, Kharkiv, Ukraine. He is currently with the Department of Student Affairs and Registration, University of Information Technology and Communications, Iraq. His current research interests include communication systems, computer networks, and wireless technologies. he can be contacted at email: mohamed.salim@uoitc.edu.iq.
	**GHAITH JAAFAR MOHAMMED** is a Lecturer in the College of Biomedical Informatics at the University of Information Technology and Communications. His research explores the adoption of cloud-ERP systems by SMEs in Iraq, analyzing key influencing factors through a mixed-methods approach. He completed his PhD in Information and Communication Technology at Universiti Teknikal Malaysia Melaka (combining master's and doctoral study), and holds a Bachelor of Engineering in Computer Techniques from Al Mamon University College. he can be contacted at email: dr.ghaith.jaafar@uoitc.edu.iq.