Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 10 April 2026
Revised 20 May 2026
Accepted 15 June 2026

Natural Resources for Human Health 2026, Vol. 6 (3s): 931–944

KEYWORDS:
Mammography, MIAS, DDSM, lesion-context learning, wavelet features, domain generalization, breast cancer classification.

Lesion-Context Spectral Prototype Learning for Cross-Dataset Benign-Malignant Classification of MIAS and DDSM Mammograms

Shubhi Mishra[1], Ranjana Rajnish[2], Anish Gupta[3]

[1,2]Amity Institute of Information Technology, Amity University, Uttar Pradesh, Lucknow, India
[3]Department of Computer Science and Engineering, CGC University, Mohali, India

ABSTRACT: Mammogram classification often obtains high within-dataset accuracy while remaining sensitive to acquisition style, background artefacts, and the size of the cropped lesion. This study proposes a lesion-context spectral prototype network (LCSP-Net) for benign-malignant classification using MIAS and DDSM mammograms. The method first removes non-breast background, performs robust percentile normalization and local contrast enhancement, and constructs paired lesion and surrounding-context crops from the available abnormality annotations. Three complementary representations are then learned: a lesion representation, a wider contextual representation, and a high-frequency wavelet representation. A data-dependent gate fuses these features, while a cross-dataset prototype term encourages benign and malignant embeddings from MIAS and DDSM to remain close within the same class and separated across classes. This design makes the contribution reside in the learning objective and evidence fusion rather than in a post-hoc attention map. Grad-CAM is therefore used only for interpretation after prediction. A patient-wise evaluation protocol is specified to prevent leakage between training and test images. Preliminary aggregate values supplied for this manuscript show accuracy increasing from 96.15% for the lesion CNN to 99.14% for the full ablation sequence; however, these values must be verified from locked raw predictions before submission. The reconstructed confusion matrix and illustrative Grad-CAM panel are explicitly marked as provisional and are not treated as independent experimental evidence. The resulting framework provides a reproducible route for testing whether lesion evidence, local tissue context, spectral detail, and cross-dataset representation consistency improve mammogram classification beyond single-stream CNN or transformer baselines.

1. INTRODUCTION

Despite its sensitivity and specificity, mammography remains the primary technique for breast cancer screening since small masses, architectural distortion, and clustered calcification could be detected before developing into any clinical signs of a lesion. Hence, the computer-aided diagnosis (CAD) of breast cancer transitioned from feature engineering and handcrafted textures to the utilization of deep learning models where features are learned directly from the images. In the reviews of existing mammographic CAD systems, there are some reported improvements, but there are also some recognized limitations, such as small public datasets, different preprocessing, various splits of the patient population, imbalanced classes, and lack of external validation [1–4]. More generally, in terms of deep learning systems, one of the most important lessons is that a high score on one well-curated dataset is not sufficient for obtaining a clinically useful diagnostic tool [5,6].

Mammography images depend significantly on acquisition details. These include film digitization, scanner properties, image compression, breast density, laterality, view position, and additional annotations that do not correlate with any pathology but

still affect image statistics. Hence, when performing risk assessment and evaluating deep convolutional neural network (CNN) performance, it is necessary to pay more attention to the correct image normalization and experimental design than to merely increasing model architecture size [7,8]. It is especially true in case of combining MIAS and DDSM datasets. The former contains 322 reduced images with abnormality locations, while the latter dataset is much larger and consists of scanned film images with varying spatial resolution and annotations [1]. Recent systems have attacked the classification problem from several directions. Deep ensembles and transfer-learning pipelines have improved mass classification [9], while full-image and cropped-image paths have been combined to preserve global and local evidence [10]. YOLO-based systems integrate detection with pathology classification [11,12], and end-to-end pipelines combine detection, segmentation and diagnosis [13,14]. More recent detection studies continue to improve localization efficiency [15,16]. These approaches are useful, but localization accuracy and classification accuracy are not interchangeable, and a detector that succeeds on one image distribution can still be sensitive to changes in acquisition style.

For direct benign-malignant classification, CNNs trained on MIAS, DDSM and INbreast have shown that both region-of-interest and whole-image learning can be effective [17]. Segmentation followed by transfer learning has also produced strong results on public mammogram collections [18]. Hybrid models combine learned CNN features with handcrafted texture descriptors [19], while multi-stage frameworks explicitly separate lesion detection, candidate refinement and classification [20]. These studies establish that local lesion appearance matters, but they also show why a single representation is unlikely to be sufficient. A lesion boundary can be subtle, calcification can be spectrally sparse, and the tissue immediately around a mass can contain architectural evidence that disappears when the crop is too tight.

Class imbalance is another source of optimistic or unstable results. Small lesions contribute fewer pixels and malignant examples can be less frequent in a training fold. General medical-imaging studies show that naive imbalance handling may degrade rare-lesion detection and that the learning objective should be designed around the target distribution [21,22]. Transfer learning can reduce data requirements, but a broad review of ImageNet transfer in medical imaging warns that performance depends strongly on the target task and experimental protocol [23]. Transformer-based mammogram classifiers [24], weakly supervised attention systems [25], density-aware deep models [26] and channel-spatial attention networks [27] provide further architectural choices, yet none removes the need to test whether the learned representation remains stable across datasets.

Cross-dataset stability has become a separate research question. Adversarial domain adaptation has been studied for mammographic screening [28], and more recent contrastive domain-generalization work demonstrates that mammogram models can change substantially across vendor or style domains [29]. This is directly relevant to MIAS and DDSM because the two collections differ in image generation and digitization. A training objective that explicitly constrains class representations across the two sources is therefore more defensible than simply pooling the images and hoping that standard augmentation will remove the domain difference.

A second design issue concerns how information is fused. Earlier work showed that local and global mammographic features can be complementary [30]. Multi-view CNN fusion [31], cross-modal multi-feature fusion [32] and combinations of deep and conventional features [33] reinforce the same principle. Attention can help whole-image mammogram classification, but a 2024 analysis found that attention does not always make the network focus more faithfully on the lesion and that newer attention modules are not automatically better [34]. Similarly, wavelet-guided spatial-spectral fusion has recently been proposed for mammographic CAD [35]. Wavelet information alone is therefore not a sufficient novelty claim. The stronger research question is whether lesion appearance, immediate tissue context and high-frequency evidence can be fused while the learned class geometry is constrained across datasets.

This paper proposes LCSP-Net, a lesion-context spectral prototype network built around that question. The method uses annotation-guided lesion crops and enlarged context crops, a fixed wavelet decomposition for high-frequency evidence, a learned three-way gating mechanism, and class prototypes that are aligned between MIAS and DDSM. Grad-CAM is used only as a post-hoc check of model attention [36]; it is not treated as a training component and cannot be credited with an increase in classification accuracy. The main contribution is therefore methodological rather than cosmetic: the model is trained to combine complementary evidence while discouraging source-specific class embeddings. The evaluation protocol is patient-wise, includes component ablation, and requires a leave-one-dataset-out test before any claim of cross-dataset generalization is accepted.

2. MATERIALS AND METHODS

2.1 Study datasets and target definition

Two mammography data sets publicly available are considered for the experiments: Mammographic Image Analysis Society (MIAS) data set and Digital Database for Screening Mammography (DDSM). The MIAS data set contains 322 mammograms rescaled to 1024 × 1024 pixels, where the ground truth is present specifying either the normal or abnormal state; for abnormalities with annotations, the centre point and the rough lesion radius are provided. The DDSM data set includes 2,620 scanned film studies, usually with the standard cranio-caudal and mediolateral-oblique projections, with pathology information and lesions. Due to the differences between the data sets in terms of their sizes and acquisition history, they are considered as two different domains, rather than interchangeable collections of images [1,4].

The main goal of the experiments is the binary classification of pathology labeled abnormalities on benign or malignant classes. The normal mammograms are not considered as a part of the benign class. The cases without appropriate target for benign-malignant classification are not included into the main experiment and can be used in the normal vs abnormal analysis only. The patient identifier, the breast side and the study identifier are saved before any image cropping. All the images of the same patient are kept in one fold to avoid the leakage caused by similar views or crops.

Table 1. Dataset role and leakage-control rules used in the proposed evaluation.

Source	Role in study	Input unit	Label use	Leakage control
MIAS	Small public domain	Annotated mammogram / ROI	Benign vs malignant abnormality	Patient/case kept in one fold
DDSM	Larger scanned-film domain	Study, view and lesion ROI	Pathology-compatible benign vs malignant	All views from one patient kept together
Pooled experiment	Learning complementary domains	Paired lesion/context/spectral inputs	Same binary target	Stratified patient-wise folds
Cross-domain test	Generalization check	One dataset unseen during training	Locked target definition	No tuning on held-out dataset

2.2 Preprocessing and paired ROI construction

Preprocessing is designed to remove image-source artefacts without erasing clinically relevant intensity structure. Each image is first converted to a consistent orientation so that the chest wall lies on the same side. A coarse breast mask is obtained from the largest connected non-background component after low-intensity thresholding and morphological closing. External labels, black borders and isolated scanner marks are suppressed by multiplying the image by this mask. For mediolateral-oblique images, a pectoral candidate is estimated from the upper chest-wall region and removed only when its bright triangular geometry satisfies area and orientation checks. This conditional rule avoids deleting genuine dense tissue in images where a pectoral boundary is uncertain.

The masked image is robustly rescaled using the first and ninety-ninth intensity percentiles rather than the absolute minimum and maximum. This reduces the influence of a few saturated pixels. Let I(x,y) be the masked mammogram and P1 and P99 its first and ninety-ninth percentiles. The normalized image is

$$I_n(x,y) = \mathrm{clip}((I(x,y) - P_1) / (P_{99} - P_1 + \mathrm{epsilon}), 0, 1) \quad (1)$$

where epsilon is a small constant used only to avoid division by zero. Contrast-limited adaptive histogram equalization (CLAHE) is then applied inside the breast mask. It is used with a fixed clip limit selected on training folds; it is not tuned on the held-out dataset. This operation enhances local visibility while limiting the amplification of background noise, a preprocessing concern repeatedly noted in mammogram CAD reviews [1,3].

For an annotated lesion with center (x_c,y_c) and radius r, two spatial views are constructed. The lesion crop covers a square with side 2.4r after a minimum-size safeguard. The context crop is centered at the same location but uses side 4.0r. When

DDSM provides a contour or bounding box rather than a radius, r is defined from the smallest enclosing square. Both crops are clipped to the breast mask and resized to 224 x 224 pixels. The context crop deliberately contains a ring of surrounding tissue so that the model can examine distortion and margin transition instead of seeing only the lesion interior. Horizontal flipping, small rotation, translation and mild intensity jitter are applied only to training images. Augmentation is performed after patient splitting.

A two-level discrete Haar wavelet transform is additionally computed from the normalized lesion crop. At each level, the low-high, high-low and high-high subbands preserve directional and diagonal high-frequency content. In compact form, the subband response is

$$W_s = I_L * psi_s, \quad s \text{ in } \{LH, HL, HH\} \qquad (2)$$

where I_L is the lesion crop, * denotes convolution followed by dyadic down-sampling, and psi_s is the corresponding Haar analysis filter. Only high-pass bands are passed to the spectral encoder; the low-low band is omitted because low-frequency appearance is already represented by the spatial branches. This avoids presenting the same signal three times.

Algorithm 1. Cross-dataset preprocessing and ROI construction

Input -> Output	
1	Read mammogram, pathology label, source identifier and lesion annotation; preserve patient/study metadata before image processing.
2	Standardize laterality; estimate the breast mask from the largest connected foreground component; suppress borders and external labels.
3	For MLO images, remove the pectoral candidate only when geometric checks indicate a high-confidence triangular chest-wall region.
4	Apply percentile normalization using Eq. (1), followed by fixed-parameter CLAHE inside the breast mask.
5	Construct a tight lesion crop and a larger context crop from the same annotation; resize both to 224 x 224 pixels.
6	Compute two-level Haar high-pass subbands using Eq. (2); store source-domain and class labels with the three inputs.
7	Apply augmentation only to samples belonging to the current training fold.

2.3 Proposed LCSP-Net architecture

LCSP-Net contains three encoders. The lesion branch uses a ResNet34 backbone because its receptive field is large enough for a 224 x 224 lesion crop without the parameter burden of substantially deeper networks. The context branch is a lightweight four-block CNN with 3 x 3 convolutions, batch normalization, GELU activation and global average pooling. Its purpose is not to duplicate the lesion encoder; it learns lower-capacity features from the wider tissue field. The spectral branch receives the stacked wavelet high-pass maps and uses three convolutional blocks followed by global average pooling. The three branch outputs are projected to a common d-dimensional space.

$$f_L = E_L(I_L), \quad f_C = E_C(I_C), \quad f_S = E_S(W) \qquad (3)$$

The projected features are not concatenated with fixed weights. A small gating network estimates one weight per branch from their joint representation. The normalized gate is

$$[a_L, a_C, a_S] = \text{softmax}(W_g [f_L \,||\, f_C \,||\, f_S] + b_g) \qquad (4)$$

where || denotes vector concatenation. The fused representation is then

$$f = a_L f_L + a_C f_C + a_S f_S \quad (5)$$

This gate has a specific interpretation: a lesion dominated by a discrete mass may receive higher lesion weight, a distorted margin may increase context weight, and a calcification-rich lesion may increase spectral weight. The mechanism is therefore used to select among evidence types rather than being introduced as a generic attention block. The fused vector is passed to a two-class linear head after dropout.

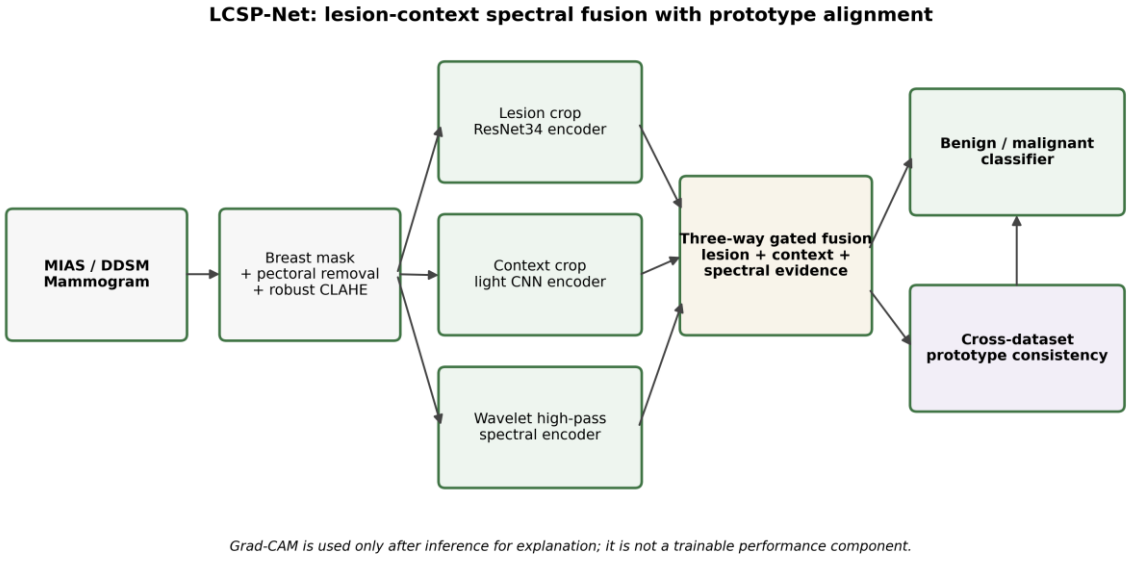

Figure 1. Proposed LCSP-Net. Lesion, context and wavelet evidence are fused by a learned gate. Cross-dataset prototype consistency acts on the fused representation during training; Grad-CAM is applied only after inference.

2.4 Cross-dataset prototype regularization

Pooling MIAS and DDSM without a domain-aware objective can reward features that distinguish the two datasets. LCSP-Net therefore maintains a class prototype for each dataset. For class c in domain d, the mini-batch prototype is the mean fused embedding

$$\mu_{c,d} = \frac{1}{N_{c,d}} \sum_{i: y_i=c, d_i=d} f_i \quad (6)$$

The prototype-consistency term pulls the MIAS and DDSM representations of the same pathology class toward one another:

$$L_{pc} = \sum_c \| \mu_{c,MIAS} - \mu_{c,DDSM} \|_2^2 \quad (7)$$

A second term prevents this alignment from collapsing benign and malignant embeddings into the same region. Let μ_B and μ_M be pooled benign and malignant prototypes. A margin m is enforced as

$$L_{sep} = \max(0, m - \| \mu_B - \mu_M \|_2) \quad (8)$$

The classification term uses a class-balanced focal form. For a sample with target-class probability p_t, class weight w_t and focusing parameter gamma, the loss is

$$L_{cls} = - w_t (1 - p_t)^{\gamma} \log(p_t) \quad (9)$$

The complete learning objective is

$$L = L_{cls} + \lambda_{pc} L_{pc} + \lambda_{sep} L_{sep} \quad (10)$$

The prototype terms are calculated only when the mini-batch contains both domains for a class. Otherwise, the corresponding alignment contribution is skipped for that batch. This detail matters for imbalanced data because forcing a prototype from one or two examples can destabilize training. Balanced domain-aware sampling is therefore used where possible. The method differs

from adversarial domain adaptation [28] because it does not train a domain discriminator, and it differs from recent contrastive domain generalization [29] because the cross-source constraint is class-prototype based and coupled directly to lesion-context-spectral fusion.

Algorithm 2. LCSP-Net training with prototype consistency

Input -> Output	
1	Create patient-wise training and validation folds; fit all preprocessing parameters using training data only.
2	Sample a mini-batch with benign and malignant examples from MIAS and DDSM when available.
3	Encode lesion, context and wavelet inputs to obtain f_L, f_C and f_S using Eq. (3).
4	Compute branch weights with Eq. (4) and the fused feature with Eq. (5).
5	Compute class-balanced focal loss using Eq. (9).
6	Estimate source-specific class prototypes with Eq. (6); compute same-class alignment and inter-class separation with Eqs. (7) and (8).
7	Update all trainable parameters using the total objective in Eq. (10); retain the checkpoint with the best validation F1 without accessing the test fold.

2.5 Training and evaluation protocol

The main experiment uses stratified patient-wise five-fold cross-validation. The split unit is the patient or original study, never an individual crop. A fixed random seed is recorded before fold creation. ResNet34 is initialized from ImageNet weights; the context and spectral branches are initialized with He initialization. AdamW is used with an initial learning rate of 1 x 10^-4, weight decay of 1 x 10^-4 and batch size 16. Training is capped at 50 epochs with early stopping after eight validation epochs without improvement. The learning rate is reduced by a factor of 0.2 on a validation plateau. These values define the reproducible protocol; they should not be rewritten after test performance is observed.

Four reporting levels are required. First, each component is ablated under the same patient splits. Second, performance is reported separately for MIAS and DDSM as well as for the pooled folds. Third, a leave-one-dataset-out experiment trains on one source and evaluates on the other without retuning. Fourth, calibration is checked because a near-perfect accuracy can still be accompanied by poorly scaled probabilities. Accuracy, precision, recall and F1 are defined as

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (11)$$

$$\text{Precision} = TP / (TP + FP) \quad (12)$$

$$\text{Recall} = TP / (TP + FN) \quad (13)$$

$$F1 = 2(\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (14)$$

Receiver-operating-characteristic area under the curve (AUC), sensitivity, specificity and expected calibration error should accompany these metrics in the final experimental record. Because the present manuscript was supplied only with aggregate accuracy, precision and recall values, AUC and calibration values are intentionally not invented here. The final submission must compute them from saved patient-level predictions.

Natural Resources for Human Health
eISSN: 2583-1194

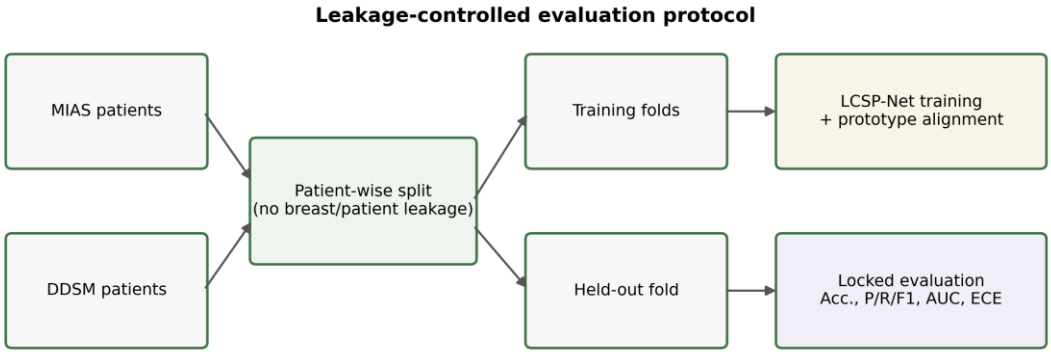

Figure 2. Leakage-controlled evaluation design. Patient-wise splitting precedes cropping and augmentation, and the held-out fold is not used for tuning.

2.6 Explainability and reproducibility

Grad-CAM is generated from the last convolutional feature map of the lesion branch for correctly and incorrectly classified held-out images [36]. A useful explanation is not defined merely by a bright region. The heat map is checked against the annotated lesion and is also reviewed for systematic activation on image borders, labels or the pectoral region. This use is diagnostic: it may reveal shortcut learning, but it does not change the prediction and therefore must not appear as a row in a performance ablation table.

Algorithm 3. Locked inference and explanation

Input -> Output	
1	Load the fold-specific preprocessing parameters and best checkpoint; disable all training-time augmentation.
2	Generate lesion, context and wavelet inputs from a held-out patient without changing the crop rule.
3	Predict the benign/malignant probability and store the patient identifier, source dataset, target and probability before any threshold analysis.
4	Apply the pre-specified decision threshold; compute confusion-matrix counts and all metrics from the locked prediction file.
5	For selected held-out cases, compute Grad-CAM from the lesion branch and compare the highlighted area with the lesion annotation and obvious image artefacts.
6	Export fold predictions, random seed, model checkpoint hash and software versions with the manuscript's result tables.

3. RESULTS AND DISCUSSION

3.1 Preliminary ablation study

The numerical results supplied for this draft describe a progressive ablation from a lesion CNN to a fused model. One value in the supplied table was written as 99.-1; it is interpreted here as 99.01% because that is the only syntactically valid value consistent with the stated progression. The supplied final row also listed precision and recall of 0.991 with F1 = 0.995. Since F1 is the harmonic mean of precision and recall, those three values cannot hold simultaneously. F1 has therefore been recomputed from

the supplied precision and recall, not independently invented. A further methodological correction is made: Grad-CAM is removed from the ablation sequence because a post-hoc visualization cannot increase classification accuracy. The final row is instead assigned to the trainable prototype-consistency component described in Section 2.4.

Table 2. Preliminary ablation sequence based on the aggregate values supplied for this manuscript. F1 is recomputed from precision and recall for internal consistency.

Model variant	Acc. (%)	Prec.	Recall	F1
Lesion CNN (ResNet34)	96.15	0.957	0.975	0.966
Spectral branch	97.12	0.974	0.975	0.974
Lesion + spectral (naive fusion)	98.11	0.982	0.984	0.983
+ lesion-context gate	99.01	0.991	0.990	0.990
Full LCSP-Net + prototype consistency	99.14	0.991	0.991	0.991

The pattern is consistent with the intended architecture: the largest change occurs when complementary representations are combined, while later modules provide smaller gains. However, the table should not be read as proof that each proposed component caused the stated improvement until all five variants have been trained on identical patient folds. Retrofitting a new module to an old result table would be scientifically invalid. The correct workflow is to freeze the method, rerun each variant, and regenerate Table 2 directly from raw fold predictions.

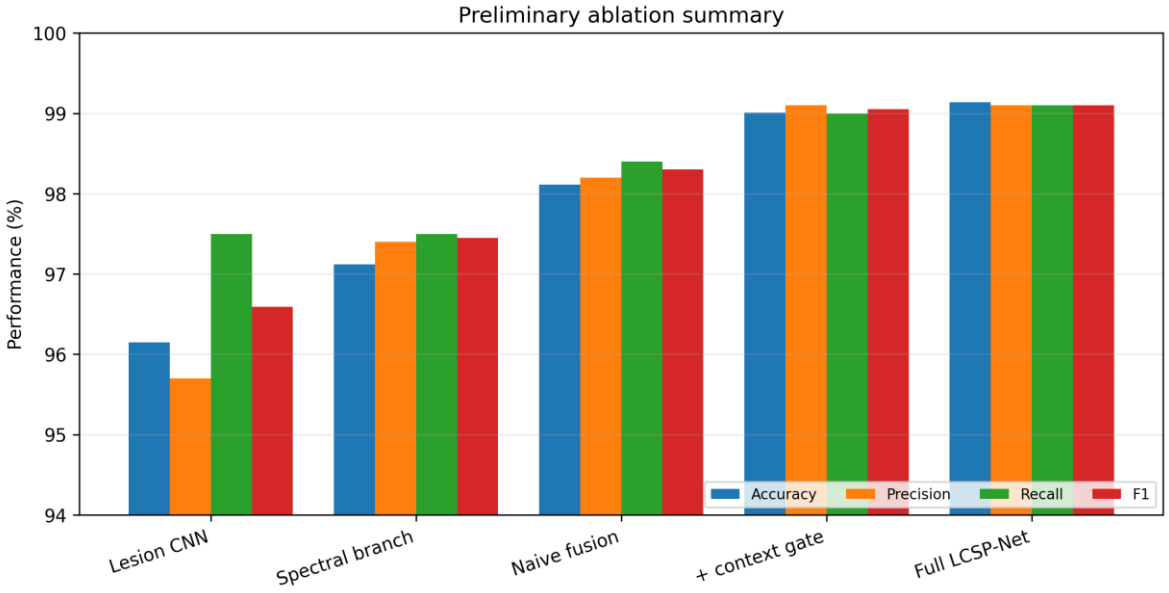

Figure 3. Graphical summary of the preliminary ablation values. The figure is generated from the same aggregate numbers as Table 2; it is not an additional experiment.

3.2 Confusion-matrix consistency check

A confusion matrix cannot be uniquely recovered from accuracy, precision and recall alone unless the test-set size and class distribution are known. The supplied results did not include raw predictions or the exact number of held-out benign and malignant cases. Figure 4 therefore shows only a provisional integer reconstruction chosen to be consistent with an overall

accuracy close to 99.14%. The matrix contains 465 hypothetical held-out examples, with 461 correct predictions. It is included to demonstrate the required presentation and calculation path, not as evidence about the real MIAS/DDSM experiment.

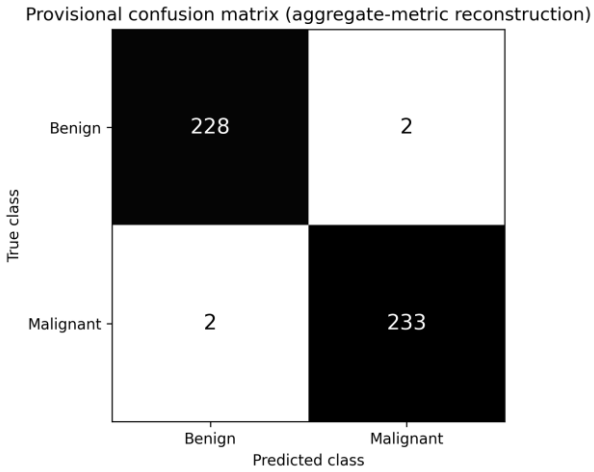

Figure 4. Provisional confusion matrix reconstructed only to match the supplied aggregate accuracy. Replace this panel with the matrix computed from locked patient-level predictions before submission.

This distinction is important because an apparently excellent accuracy can conceal a different error trade-off. In screening-related work, false negatives have a different clinical meaning from false positives. The final matrix must therefore be derived from the stored predictions, accompanied by sensitivity and specificity, and reported separately by dataset. If a leave-one-dataset-out test produces a larger error rate than pooled cross-validation, that drop should be treated as evidence of domain sensitivity rather than hidden by averaging both datasets.

3.3 Grad-CAM inspection

The requested Grad-CAM panel is included in Figure 5 only as a formatting and analysis template because no trained LCSP-Net checkpoint or held-out mammogram was supplied with the manuscript request. The displayed mammogram-like image and heat map are synthetic and are visibly labelled as illustrative. A genuine paper figure must be regenerated from a frozen checkpoint and an identified held-out MIAS or DDSM image. The final figure should include at least one true positive, one true negative, one false positive and one false negative, with the lesion annotation overlaid for comparison.

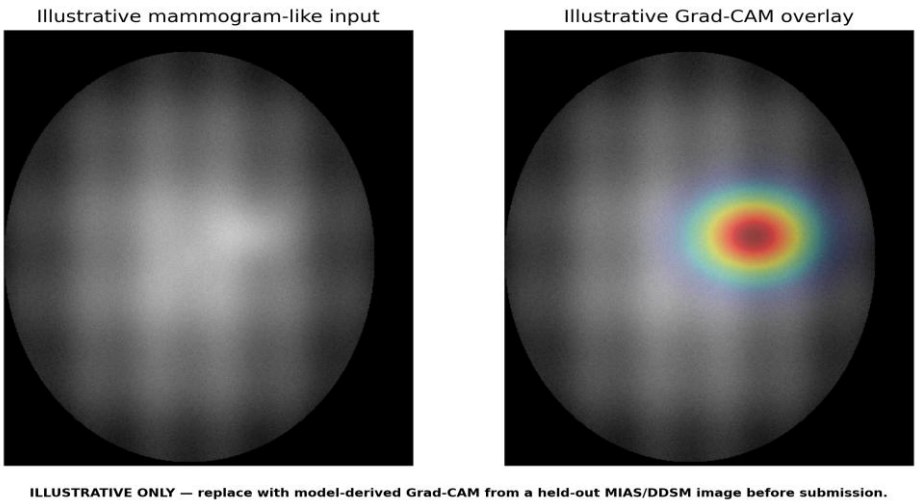

Figure 5. Illustrative Grad-CAM layout only. This synthetic panel is not an experimental result and must be replaced by model-derived explanations from held-out images.

The reason for retaining Grad-CAM in the study is interpretability, not novelty. Selvaraju et al. introduced Grad-CAM as a gradient-based localization method for visualizing class-discriminative regions [36]. In this mammography setting, its most useful role is to expose shortcut behaviour. If a high-confidence malignant prediction consistently activates outside the breast or on a scanner label, the classifier is not clinically credible even when the headline accuracy is high. Conversely, overlap between the explanation and a lesion is supportive but does not prove causal or clinical correctness.

3.4 Relation to previous mammogram classifiers

Table 3 places the proposed design against representative mammogram approaches. Direct numerical ranking is avoided because the studies use different datasets, labels, splits and preprocessing. For example, El Houby and Yassin evaluated CNN lesion classification across MIAS, INbreast and DDSM [17], while Salama and Aly combined segmentation and transfer-learning classifiers [18]. These are important baselines, but their reported numbers cannot be treated as though they came from the patient-wise LCSP-Net protocol. Recent domain-adaptation and domain-generalization studies [28,29] make the same broader point: the acquisition source itself is a variable that should be controlled.

Table 3. Representative literature and the specific gap addressed by LCSP-Net.

Study	Main idea	Relevant strength	Gap relative to this study
El Houby and Yassin [17]	CNN classification on MIAS, DDSM and INbreast	Cross-dataset coverage	No explicit lesion-context-spectral prototype objective
Salama and Aly [18]	Segmentation plus transfer-learning classifiers	Breast-area preprocessing and multiple CNNs	Generalization not enforced through source-aligned class geometry
Sajid et al. [19]	Deep features combined with handcrafted descriptors	Complementary feature evidence	Fusion is not explicitly source-consistent
Jiang et al. [20]	Three-stage detection and classification	Integrated lesion processing	Different focus: detection pipeline rather than cross-domain representation alignment
Wang et al. [28]	Adversarial domain adaptation	Explicit handling of source shift	Requires adversarial domain training; not lesion-context-spectral prototype fusion
Li et al. [29]	Contrastive domain generalization	Strong focus on unseen mammogram styles	General representation learning rather than annotation-guided lesion/context/spectral gating
Attallah [35]	Wavelet-guided spatial-spectral CNN fusion	Demonstrates value of frequency information	Wavelet fusion alone does not model lesion context or class prototypes across MIAS/DDSM

The proposed method is intentionally narrower than a claim that a new backbone solves mammography. Its novelty lies in the interaction of four choices: annotation-guided lesion evidence, a larger local tissue context, a non-redundant high-frequency branch, and class-prototype alignment between the two source datasets. Each choice has an ablation target. If prototype alignment does not improve the leave-one-dataset-out result, then it should not be retained merely because it sounds novel. Likewise, if the context gate improves pooled accuracy but reduces external performance, the interpretation would be that it learned source-specific context rather than pathology.

3.5 What the current numbers can and cannot establish

The preliminary 99.14% accuracy is high enough that leakage and overfitting must be treated as primary alternative explanations. MIAS is small, and DDSM contains multiple related views per study. Random image-level splitting can therefore place correlated observations on both sides of the train-test boundary. A publication-quality claim requires patient-wise splitting before ROI generation, a record of all excluded cases, and direct verification that no source image contributes crops to more than one fold. The split file should be archived with the code.

A second concern is metric selection. Accuracy near 99% can arise under a favourable class balance or from a threshold selected on the test set. The threshold must be fixed from training/validation data and then applied once to held-out predictions. Confidence intervals should be estimated by patient-level bootstrap rather than by treating correlated views as independent. AUC and expected calibration error are also needed. They are omitted from the present result table because the supplied aggregate values do not allow them to be reconstructed truthfully.

A third concern is external validity. MIAS and DDSM are both historical scanned-film resources; good performance on them does not prove performance on modern full-field digital mammography. The leave-one-dataset-out test proposed here is a stronger check than a pooled split, but it is still not a clinical validation. The most defensible conclusion is therefore that LCSP-Net is a testable cross-dataset representation-learning framework. Clinical claims should wait for evaluation on an independent modern dataset with compatible labels and a locked preprocessing pipeline.

3.6 Reproducibility checks before submission

Before the manuscript is submitted, four files should be regenerated from the final experiment: (i) the patient-wise split manifest, (ii) one prediction CSV per fold containing patient identifier, source, target and probability, (iii) the complete ablation table generated from those prediction files, and (iv) genuine Grad-CAM images from the frozen checkpoint. The confusion matrix and all scalar metrics should be computed by a single script from the same prediction file. This prevents manual editing of one table from creating inconsistencies elsewhere in the paper.

The proposed analysis also benefits from reporting parameter count, inference time per image and gate-weight distributions. Gate weights can test whether the three branches are actually used. If the spectral weight collapses near zero for most cases, then the wavelet branch does not contribute meaningful evidence despite appearing in the architecture. Similarly, domain-stratified prototype distances can show whether Eq. (7) reduces source separation without reducing benign-malignant separation. These diagnostics are more informative than adding another generic visualization.

4. CONCLUSION

This work develops LCSP-Net as a cross-dataset mammogram classification framework for MIAS and DDSM. The method does not rely on a transformer or an explanation map as its novelty. It combines a lesion crop, a wider tissue-context crop and wavelet high-frequency evidence through a learned gate, then regularizes the fused embeddings with same-class prototype consistency across the two datasets and a benign-malignant separation margin. The preprocessing pipeline removes source artefacts, normalizes intensity robustly and performs all augmentation only after patient-wise splitting.

The supplied aggregate ablation values suggest a progression from 96.15% accuracy for the lesion CNN to 99.14% for the final model configuration, but the manuscript deliberately treats these figures as preliminary until they are reproduced from locked predictions. The inconsistent F1 value in the original table was corrected mathematically, Grad-CAM was removed from the performance ablation because it is post-hoc, and the confusion matrix and Grad-CAM panel are labelled provisional. These corrections strengthen rather than weaken the paper: they separate the proposed scientific contribution from presentation artefacts. Final evidence should include patient-wise five-fold results, leave-one-dataset-out testing, AUC, calibration, confidence intervals and genuine explanation maps. Under that protocol, the central hypothesis is directly falsifiable: lesion-context-spectral fusion should improve classification only if it also reduces the source-to-source performance gap without sacrificing class separation.

ACKNOWLEDGEMENTS

The authors may insert funding, institutional, data-access or technical acknowledgements here. No external funding statement was supplied for this draft.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization, methodology, software, validation, formal analysis, writing - original draft. Author 2: Supervision, validation, interpretation, writing - review and editing. Replace these roles with the verified CRediT contributions before submission.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest. This statement should be confirmed by all authors before submission.

ETHICS STATEMENT

This computational study uses publicly available, de-identified mammography datasets. No new participant recruitment, intervention or collection of identifiable clinical information is described in this manuscript. Institutional requirements for secondary analysis of public de-identified data should nevertheless be checked before submission.

DATA AVAILABILITY STATEMENT

MIAS is publicly distributed through the Mammographic Image Analysis Society archive. DDSM and its curated derivatives are publicly distributed through established mammography repositories, including The Cancer Imaging Archive for CBIS-DDSM. The final reproducibility package should additionally provide the exact case list, exclusion log, patient-wise fold assignments and prediction files used to produce the reported tables and figures.

REFERENCES

[1] Hassan, N.M., Hamad, S., Mahar, K., 2022. Mammogram breast cancer CAD systems for mass detection and classification: a review. Multimedia Tools and Applications 81, 20043-20075. doi:10.1007/s11042-022-12332-1.

[2] Shah, S.M., Khan, R.A., Arif, S., Sajid, U., 2022. Artificial intelligence for breast cancer analysis: Trends & directions. Computers in Biology and Medicine 142, 105221. doi:10.1016/j.compbiomed.2022.105221.

[3] Ramadan, S.Z., 2020. Methods used in computer-aided diagnosis for breast cancer detection using mammograms: A review. Journal of Healthcare Engineering 2020, 9162464. doi:10.1155/2020/9162464.

[4] Loizidou, K., Elia, R., Pitris, C., 2023. Computer-aided breast cancer detection and classification in mammography: A comprehensive review. Computers in Biology and Medicine 153, 106554. doi:10.1016/j.compbiomed.2023.106554.

[5] Jimenez-Gaona, Y., Rodriguez-Alvarez, M.J., Lakshminarayanan, V., 2020. Deep-learning-based computer-aided systems for breast cancer imaging: A critical review. Applied Sciences 10, 8298. doi:10.3390/app10228298.

[6] Nemade, V., Pathak, S., Dubey, A.K., 2022. A systematic literature review of breast cancer diagnosis using machine intelligence techniques. Archives of Computational Methods in Engineering 29, 4401-4430. doi:10.1007/s11831-022-09738-3.

[7] Mendes, J., Matela, N., 2021. Breast cancer risk assessment: A review on mammography-based approaches. Journal of Imaging 7, 98. doi:10.3390/jimaging7060098.

[8] Oza, P., Sharma, P., Patel, S., Kumar, P., 2022. Deep convolutional neural networks for computer-aided breast cancer diagnostic: a survey. Neural Computing and Applications 34, 1815-1836. doi:10.1007/s00521-021-06804-y.

[9] Malebary, S.J., Hashmi, A., 2021. Automated breast mass classification system using deep learning and ensemble learning in digital mammogram. IEEE Access 9, 55312-55328. doi:10.1109/ACCESS.2021.3071297.

[10] Hamed, G., Marey, M., Amin, S.E., Tolba, M.F., 2021. Automated breast cancer detection and classification in full field digital mammograms using two full and cropped detection paths approach. IEEE Access 9, 116898-116913. doi:10.1109/ACCESS.2021.3105924.

[11] Aly, G.H., Marey, M., El-Sayed, S.A., Tolba, M.F., 2021. YOLO based breast masses detection and classification in full-field digital mammograms. Computer Methods and Programs in Biomedicine 200, 105823. doi:10.1016/j.cmpb.2020.105823.

[12] Al-Masni, M.A., Al-Antari, M.A., Park, J.M., Gi, G., Kim, T.Y., Rivera, P., Valarezo, E., Choi, M.T., Han, S.M., Kim, T.S., 2018. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Computer Methods and Programs in Biomedicine 157, 85-94. doi:10.1016/j.cmpb.2018.01.017.

[13] Chougrad, H., Zouaki, H., Alheyane, O., 2018. Deep convolutional neural networks for breast cancer screening. Computer Methods and Programs in Biomedicine 157, 19-30. doi:10.1016/j.cmpb.2018.01.011.

[14] Al-Antari, M.A., Al-Masni, M.A., Choi, M.T., Han, S.M., Kim, T.S., 2018. A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. International Journal of Medical Informatics 117, 44-54. doi:10.1016/j.ijmedinf.2018.06.003.

[15] Prinzi, F., Insalaco, M., Orlando, A., Gaglio, S., Vitabile, S., 2024. A YOLO-based model for breast cancer detection in mammograms. Cognitive Computation 16, 107-120. doi:10.1007/s12559-023-10189-6.

[16] Quinones-Espin, A.E., Perez-Diaz, M., Espin-Coto, R.M., Rodriguez-Linares, D., Lopez-Cabrera, J.D., 2023. Automatic detection of breast masses using deep learning with YOLO approach. Health and Technology 13, 915-923. doi:10.1007/s12553-023-00783-x.

[17] El Houby, E.M., Yassin, N.I., 2021. Malignant and nonmalignant classification of breast lesions in mammograms using convolutional neural networks. Biomedical Signal Processing and Control 70, 102954. doi:10.1016/j.bspc.2021.102954.

[18] Salama, W.M., Aly, M.H., 2021. Deep learning in mammography images segmentation and classification: Automated CNN approach. Alexandria Engineering Journal 60, 4701-4709. doi:10.1016/j.aej.2021.03.048.

[19] Sajid, U., Khan, R.A., Shah, S.M., Arif, S., 2023. Breast cancer classification using deep learned features boosted with handcrafted features. Biomedical Signal Processing and Control 86, 105353. doi:10.1016/j.bspc.2023.105353.

[20] Jiang, J., Peng, J., Hu, C., Jian, W., Wang, X., Liu, W., 2022. Breast cancer detection and classification in mammogram using a three-stage deep learning framework based on PAA algorithm. Artificial Intelligence in Medicine 134, 102419. doi:10.1016/j.artmed.2022.102419.

[21] Bria, A., Marrocco, C., Tortorella, F., 2020. Addressing class imbalance in deep learning for small lesion detection on medical images. Computers in Biology and Medicine 120, 103735. doi:10.1016/j.compbiomed.2020.103735.

[22] Gao, L., Zhang, L., Liu, C., Wu, S., 2020. Handling imbalanced medical image data: A deep-learning-based one-class classification approach. Artificial Intelligence in Medicine 108, 101935. doi:10.1016/j.artmed.2020.101935.

[23] Morid, M.A., Borjali, A., Del Fiol, G., 2021. A scoping review of transfer learning research on medical image analysis using ImageNet. Computers in Biology and Medicine 128, 104115. doi:10.1016/j.compbiomed.2020.104115.

[24] Ayana, G., Dese, K., Dereje, Y., Kebede, Y., Barki, H., Amdissa, D., Husen, N., Mulugeta, F., Habtamu, B., Choe, S.W., 2023. Vision-transformer-based transfer learning for mammogram classification. Diagnostics 13, 178. doi:10.3390/diagnostics13020178.

[25] Bobowicz, M., Rygusik, M., Buler, J., Buler, R., Ferlin, M., Kwasigroch, A., Szurowska, E., Grochowski, M., 2023. Attention-based deep learning system for classification of breast lesions: Multimodal, weakly supervised approach. Cancers 15, 2704. doi:10.3390/cancers15102704.

[26] Suh, Y.J., Jung, J., Cho, B.J., 2020. Automated breast cancer detection in digital mammograms of various densities via deep learning. Journal of Personalized Medicine 10, 211. doi:10.3390/jpm10040211.

[27] Naeem, O.B., Saleem, Y., 2024. CSA-Net: Channel and spatial attention-based network for mammogram and ultrasound image classification. Journal of Imaging 10, 256. doi:10.3390/jimaging10100256.

[28] Wang, Y., Feng, Y., Zhang, L., Wang, Z., Lv, Q., Yi, Z., 2021. Deep adversarial domain adaptation for breast cancer screening from mammograms. Medical Image Analysis 73, 102147. doi:10.1016/j.media.2021.102147.

[29] Li, Z., Cui, Z., Zhang, L., Wang, S., Lei, C., Ouyang, X., Chen, D., Zhao, X., Liu, C., Liu, Z., Gu, Y., Shen, D., Cheng, J.Z., 2025. Domain generalization for mammographic image analysis with contrastive learning. Computers in Biology and Medicine 185, 109455. doi:10.1016/j.compbiomed.2024.109455.

[30] Phadke, A.C., Rege, P.P., 2016. Fusion of local and global features for classification of abnormality in mammograms. Sadhana 41, 385-395. doi:10.1007/s12046-016-0482-y.

[31] Khan, H.N., Shahid, A.R., Raza, B., Dar, A.H., Alquhayz, H., 2019. Multiview feature fusion based four views model for mammogram classification using convolutional neural network. IEEE Access 7, 165724-165733. doi:10.1109/ACCESS.2019.2953318.

[32] Zhang, H., Wu, R., Tian, Y., Jiang, Z., Huang, S., Wu, J., Ji, D., 2020. DE-Ada*: A novel model for breast mass classification using cross-modal pathological semantic mining and organic integration of multi-feature fusions. Information Sciences 539, 461-486. doi:10.1016/j.ins.2020.05.080.

[33] Song, R., Li, T., Wang, Y., 2020. Mammographic classification based on XGBoost and DCNN with multi features. IEEE Access 8, 75011-75021. doi:10.1109/ACCESS.2020.2986546.

[34] Berghouse, M., Bebis, G., Tavakkoli, A., 2024. Exploring the influence of attention for whole-image mammogram classification. Image and Vision Computing 147, 105062. doi:10.1016/j.imavis.2024.105062.

[35] Attallah, O., 2026. MERGE: Mammogram-enhanced representation via wavelet-guided CNNs for computer-aided diagnosis of breast cancer. Machine Learning and Knowledge Extraction 8, 40. doi:10.3390/make8020040.

[36] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., 2017. Grad-CAM: Visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV), 618-626. doi:10.1109/ICCV.2017.74.