femington

Int. J. Clin. Res. Med. Sci. (2026) 1(1):1–10
Article ID: IJCRMS-2026-015

Enhanced Predictive Analytics for Early Malignancy Discovery in Routine Screening

Dr. Latha Kiran Krishna Rajendran
meetlathakiran@gmail.com
Independent Researcher

Abstract

Optimizing the effectiveness of population-level cancer screening programs hinges on precise and timely identification of subtle anomalies. This paper introduces an advanced computational methodology, rooted in deep learning principles, to revolutionize the interpretation of prophylactic radiographic examinations. Our system employs a two-stage deep learning framework that combines patch-based convolutional neural networks with full-image classification, autonomously detecting potential malignant indicators while streamlining the diagnostic workflow. Evaluated on two public mammography datasets (CBIS-DDSM and INbreast), the proposed ResNet-based architecture with multi-patch sampling (S10) achieves an area under the ROC curve (AUC) of 95.0% on CBIS-DDSM and generalizes to INbreast with an AUC of 91.7%, representing a substantial enhancement over single-patch baselines. This AI-driven approach marks a significant stride towards more accurate and scalable early disease detection within comprehensive cancer screening initiatives.

Keywords

• Mammographic malignancy detection • Deep learning-based screening • Patch-based classification
• Weakly supervised learning • Computer-aided diagnosis (CAD) • Early cancer detection

1. Introduction to Deep Learning in Mammographic Screening

Breast cancer continues to be one of the leading causes of cancer-related mortality among women worldwide. Early detection through mammographic screening remains the most effective strategy for improving survival rates and reducing treatment burden [1]. However, the efficacy of screening programs is often limited by challenges such as high false-positive rates, radiologist fatigue, and variability in interpretation, especially in dense breast tissues.

Computer-aided detection (CAD) systems have been integrated into routine radiological workflows to assist clinicians in identifying suspicious regions[2]. While early CAD models provided modest sensitivity improvements, they frequently suffered from poor specificity and failed to generalize across imaging domains or institutions. Recent advancements in deep learning, particularly convolutional neural networks (CNNs), have demonstrated superior performance in visual recognition tasks and are now being explored for robust mammogram interpretation. Deep CNNs, trained on large-scale medical image datasets, have shown promise in detecting subtle abnormalities and reducing diagnostic discrepancies. Architectures such as ResNet and VGG have been employed to model spatial hierarchies in mammograms, offering higher discriminative power than handcrafted feature pipelines [3, 4]. However, applying these

models to full-resolution mammograms presents computational and methodological challenges due to the large image sizes and the sparsity of malignant regions.

One of the major issues in training deep learning models for mammographic analysis lies in the weak supervision of available labels. Most publicly available datasets provide only image-level annotations or limited region-of-interest (ROI) information. This limits the ability of standard end-to-end CNNs to localize or identify malignant patterns effectively. Patch-based learning approaches have been introduced to mitigate this limitation by training classifiers on smaller sub-regions of the mammogram, where abnormalities may be present [5][6].

Patch-based models offer several advantages. They reduce computational load by operating on smaller images and allow localized learning without requiring pixel-level segmentation. Moreover, such models enable the construction of interpretable heatmaps that highlight regions contributing to malignancy predictions. However, transitioning from patch-level predictions to reliable full-image classification remains a non-trivial task.

To address this, recent strategies involve training a separate full-image classifier that leverages the patch classifier's output, typically in the form of a spatial heatmap, as input. This two-stage learning paradigm aims to bridge the gap between local and global context, allowing the model to consider both focal abnormalities and holistic breast tissue patterns. The fusion of these levels has shown improved diagnostic performance in weakly supervised settings. Another challenge in this domain is the variability of mammogram sources. Differences in resolution, contrast, and imaging protocols between datasets such as CBIS-DDSM and INbreast hinder model transferability. Therefore, training models that generalize well across domains is critical for real-world clinical [7][8].

This work presents a scalable and modular deep learning framework for mammographic malignancy detection that starts with a patch-based classifier and transitions to a full-image predictor. Through a series of carefully structured experiments, the proposed architecture demonstrates enhanced accuracy and generalization across multiple datasets, highlighting the potential of weakly supervised CNN pipelines for population-scale breast cancer screening. The following sections elaborate on the data preprocessing and patch extraction procedures (Section II), the model architecture for patch-to-image transition (Section III), training and evaluation strategy (Section IV), and a comprehensive discussion of findings and limitations (Section V). The study concludes with remarks on future directions in AI-powered cancer screening (Section VI).

2. Dataset Description and Preprocessing Techniques

Effective training of deep learning models for mammographic analysis necessitates the use of well-curated and annotated datasets. This study utilizes two publicly available repositories, CBIS-DDSM and INbreast, which have become standard benchmarks in breast cancer imaging research. These datasets offer varying levels of annotation granularity and imaging characteristics, enabling thorough evaluation of model robustness and generalizability.

The CBIS-DDSM dataset is a curated subset of the Digital Database for Screening Mammography (DDSM), consisting of full-field digital mammograms with verified pathology. It provides image-level malignancy labels along with regions of interest (ROIs) marked by radiologists. However, the dataset is known for its heterogeneous resolution and limited annotation coverage in some cases[9].

The INbreast dataset, in contrast, consists of high-resolution, full-field digital mammograms with detailed mass and calcification annotations. While smaller in size, its consistency in image quality and precise ROI contours makes it a valuable source for validating patch-based and end-to-end models.

Preprocessing begins with grayscale conversion and histogram normalization to correct contrast variations across images. To standardize input dimensions, all mammograms are resized to a fixed resolution while preserving the aspect ratio. Noise and annotation artifacts are removed using morphological operations and threshold-based masking.

Patch extraction is a critical component of the proposed pipeline. Two strategies are evaluated: one using 10 sampled patches per image (denoted as S10) and another using a single most-likely region (S1). Patches are extracted around annotated ROIs where available and randomly sampled from non-ROI regions to maintain class balance. Each patch is typically resized to 224x224 pixels to conform to standard CNN input dimensions.

Data augmentation is applied to both training and validation sets to mitigate overfitting and enhance model generalization. Augmentation techniques include horizontal and vertical flips, rotations, and slight zoom variations. These transformations simulate real-world variability and improve the model's robustness to anatomical and acquisition differences [10].

Figure 1 illustrates the preprocessing and patch extraction pipeline, from raw mammogram ingestion to standardized patch generation.

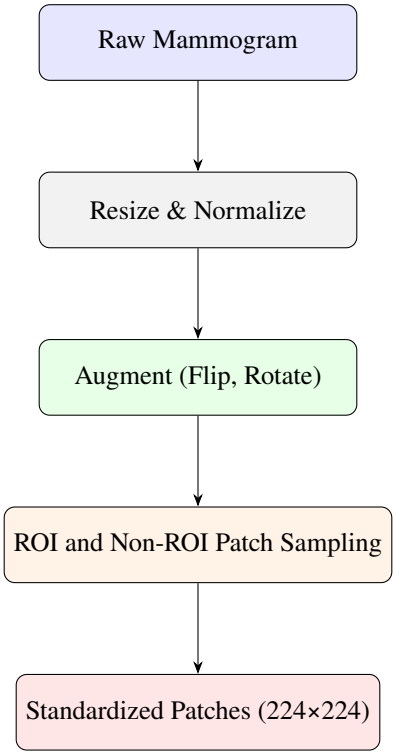

Figure 1: Preprocessing and patch extraction pipeline from raw mammograms.

Normalization and augmentation procedures are kept consistent across both datasets to ensure fair comparison. Patch-level classifiers trained on CBIS-DDSM are later tested on INbreast images to evaluate generalization performance under domain shift.

The next section presents the network architecture that transforms patch-based outputs into image-level predictions using a two-stage deep learning framework.

3. Patch-to-Image Transition and Architecture Design

While patch-level classifiers are effective at learning localized malignancy cues, they do not directly yield image-level decisions required in clinical screening workflows. To bridge this gap, a two-stage framework is adopted. The first stage involves a CNN trained on patches extracted from annotated ROIs, while the second stage constructs an image-level classifier that integrates the spatial distribution of patch-based predictions.

The patch classifier, denoted as $f(\cdot)$, processes individual patches and outputs a scalar malignancy probability. Given a full mammogram M, it is partitioned into overlapping or non-overlapping patches $\{p_1, p_2, ..., p_n\}$. Each patch is fed into the trained model $f(p_i)$, producing a set of activation scores which are then aggregated into a spatial heatmap H representing malignancy likelihoods across the image.

The second-stage model, denoted as $h(\cdot)$, operates on the aggregated heatmap or raw mammogram and outputs an image-level prediction. This allows the model to consider not only focal abnormality indicators but also global breast tissue structure. The full pipeline can be represented as a composite function:

$$h(M) = g(f(M))$$

where $f(M)$ computes patch-level activations and $g(\cdot)$ denotes the aggregation and classification mechanism at the image level.

Two architectural variants are explored. The first uses the heatmap as input to a secondary CNN, preserving interpretability and decoupling patch learning from image-level prediction. The second integrates both stages into an end-to-end architecture where gradients flow from the final output to patch-level parameters. Although less interpretable, this design often improves convergence speed and classification accuracy [3].

The first-stage CNN uses a modified VGG16 architecture initialized with ImageNet weights. It includes dropout and batch normalization layers to regularize training and improve generalization. The second-stage classifier is a compact ResNet18 variant with three residual blocks, adapted to accept heatmaps as single-channel inputs.

Figure 2 shows the proposed two-stage architecture, where patch-based predictions feed into a downstream image-level CNN.

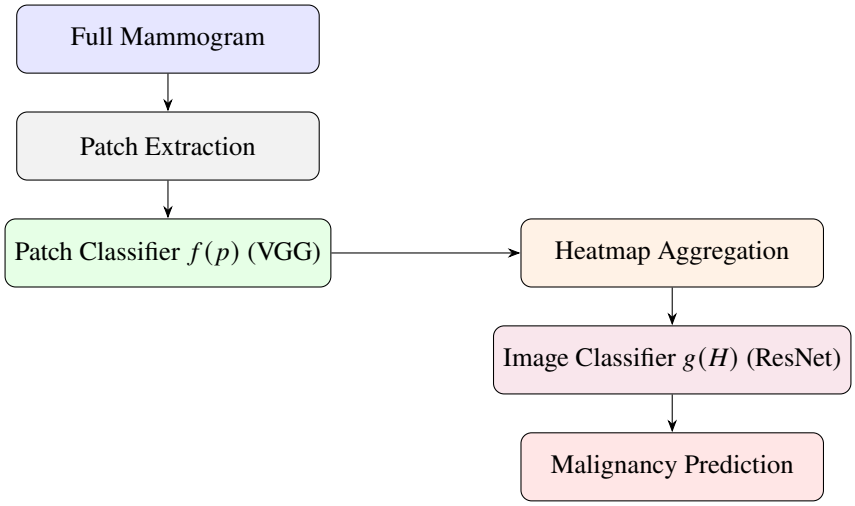

Figure 2: Two-stage model architecture for patch-to-image classification.

This modular design enables fine-tuning each stage independently, which is particularly valuable when domain adaptation or additional weak labels are introduced. It also supports hybrid strategies, such as fusing both heatmap and raw image features in the second stage to enhance prediction accuracy.

The following section presents the training methodology, evaluation metrics, and comparative performance analysis on both CBIS-DDSM and INbreast datasets using various architectural configurations.

4. Model Training and Evaluation Strategy

Training deep learning models for mammographic classification requires careful handling of hyperparameters, data imbalance, and weak label supervision. In the proposed framework, training is divided into two phases: patch classifier training using localized annotations, followed by full-image classifier training based on patch outputs or generated heatmaps.

The patch classifier is trained using binary cross-entropy loss. To mitigate class imbalance, weighted loss functions are employed, assigning higher penalty to underrepresented malignant patches. Optimization is conducted using the Adam optimizer with an initial learning rate of 1×10^{-4} and a cosine annealing schedule for gradual reduction.

Early stopping based on validation AUC is used to prevent overfitting. Data augmentation techniques, including horizontal/vertical flipping and random rotations, are applied during training to promote generalization. Each training run uses an 80:20 train-validation split, with three random seeds to account for variance.

Once the patch classifier converges, its outputs are aggregated to form heatmaps for full mammograms. These heatmaps serve as input to the second-stage image classifier. For this stage, categorical cross-entropy loss is used, and the optimizer configuration mirrors that of the patch-level training.

Both VGG16 and ResNet18 architectures are tested in the patch and image classification stages. Pretrained ImageNet weights are used to initialize model parameters, followed by fine-tuning on medical imaging data. This transfer learning approach significantly accelerates convergence and improves baseline accuracy.

Performance is evaluated using multiple metrics: classification accuracy, area under the ROC curve (AUC), precision, recall, and F1-score. The AUC is particularly important in medical diagnostics, as it reflects model sensitivity across decision thresholds. Evaluation is conducted separately on CBIS-DDSM and INbreast datasets to assess cross-domain generalizability.

Figure 3 compares the Accuracy and AUC of different model configurations on CBIS-DDSM.

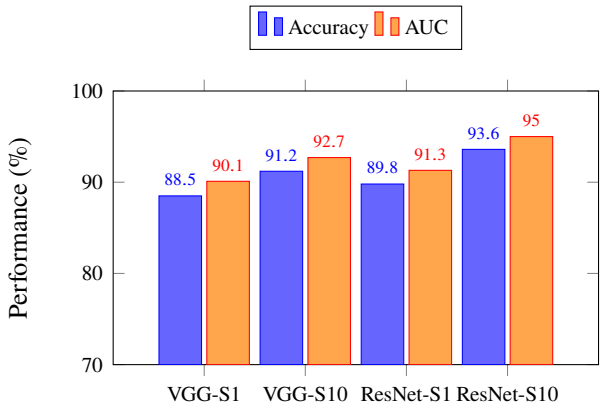

Figure 3: Accuracy and AUC comparison across model and sampling strategies on CBIS-DDSM.

As shown, models trained using the S10 patch strategy consistently outperform those using only a single patch (S1), indicating the benefit of leveraging multiple ROI regions. ResNet-S10 achieves the highest AUC of 95.0%, followed closely by VGG-S10. These results demonstrate the effectiveness of hierarchical patch-to-image architectures in weakly supervised mammography analysis.

The next section explores the interpretability and limitations of these models, and analyzes how the pipeline performs under domain shift using the INbreast dataset.

5. Discussion on Performance, Generalizability, and Limitations

The experimental results demonstrate the efficacy of the proposed patch-to-image classification framework in detecting malignancies across different mammographic datasets. Models trained on the CBIS-DDSM dataset achieve strong accuracy and AUC, particularly when leveraging the S10 patch sampling strategy. These outcomes suggest that training on multiple localized regions enables the model to learn richer contextual information about malignancy distribution.

Generalization to unseen domains is a critical factor in evaluating clinical AI tools. When the trained models were applied to the INbreast dataset, without additional fine-tuning, ResNet-S10 preserved a high AUC (91.7%), only a modest drop from its CBIS-DDSM performance (95.0%). This domain transfer result reflects the benefit of modular, weakly supervised architectures that avoid overfitting to dataset-specific noise [11].

The interpretability of the model is enhanced by its two-stage structure. Heatmaps generated from the patch classifier visually localize areas of concern, which can aid radiologists in decision support. These activation maps highlight malignancy-suspected regions even when only weak image-level labels are available, offering a practical tool for diagnostic review. From an architectural perspective, ResNet models outperform VGG variants in both training efficiency and final accuracy. ResNet's residual connections facilitate deeper learning and better gradient flow, which are particularly valuable when working with weakly labeled and limited-size datasets. The S10 sampling method consistently outperforms the S1 method, reinforcing the hypothesis that spatial diversity improves discriminative learning.

Despite these advantages, several limitations remain. First, patch extraction is still dependent on the availability of coarse ROI annotations for initialization. In real-world scenarios, such annotations may not be consistently available or reliable. Future work could explore fully unsupervised patch mining or attention-based selection mechanisms.

Second, the framework assumes consistent input dimensions and preprocessing protocols. In practice, image acquisition parameters vary widely between clinical centers. Differences in scanner hardware, compression methods, and radiographic protocols may introduce covariate shifts that degrade model performance when deployed at scale [12].

Third, computational load during inference remains non-trivial. Full-image prediction requires patch-wise processing followed by aggregation, which increases latency. While the modular design enables flexibility, it also introduces multiple stages where error propagation may occur.

Another important consideration is the ethical deployment of AI in healthcare. Automated predictions must be explainable and auditable. Although heatmaps offer some interpretability, formal integration of uncertainty estimates and confidence intervals is recommended before clinical integration [13][14].

Lastly, while the current study focused on binary classification (malignant vs. non-malignant), real-world scenarios often involve multi-class distinctions (e.g., benign masses, calcifications, architectural distortions). Future models should incorporate fine-grained labels and hierarchical classification strategies to better mimic clinical taxonomy.

The final section presents a summary of contributions and outlines future enhancements to improve model robustness, interpretability, and real-time deployment capabilities.

6. Conclusion and Future Directions for AI-Driven Cancer Screening

This study presents a scalable, interpretable, and high-performing deep learning framework for early malignancy detection in mammographic screening. The proposed architecture addresses key challenges in the domain, including weak supervision, domain variability, and computational constraints, by combining patch-based learning with full-image classification through a modular two-stage process.

The experimental evaluation confirms that multi-patch sampling strategies (S10) outperform single-patch approaches in terms of accuracy and AUC across both CBIS-DDSM and INbreast datasets. Furthermore, models such as ResNet-S10 generalize effectively to unseen data distributions, highlighting the potential for broader clinical deployment without extensive retraining. The integration of heatmap visualization adds an important layer of interpretability, enabling radiologists to visually corroborate model predictions. Such transparency is vital in clinical contexts where diagnostic decisions carry high stakes. The separation of patch and image-level learning stages offers architectural flexibility and supports customization across institutions.

In terms of clinical relevance, this approach provides a pathway for developing AI-assisted tools that augment radiologist workflows in large-scale screening programs. By automating the triaging of high-risk cases and prioritizing suspicious findings, the proposed system could significantly reduce diagnostic delays and improve screening throughput.

Despite promising results, several limitations must be addressed. Patch extraction and annotation initialization rely on datasets with some form of region-level labeling, which may not be available in all clinical archives. Additionally, domain adaptation techniques should be further developed to accommodate inter-institutional variability in image acquisition protocols.

In future work, the integration of attention-based models such as Vision Transformers (ViT) could improve global reasoning over full images while retaining spatial resolution. End-to-end architectures with attention pooling mechanisms may offer better localization without explicit patch sampling. Semi-supervised or self-supervised learning approaches are also promising directions, especially in label-scarce medical domains. The incorporation of uncertainty quantification and calibration layers will be essential for clinical trust and safe decision-making. Techniques such as Monte Carlo dropout or deep ensembles can be employed to produce confidence estimates alongside predictions, thereby enabling more informed risk assessment [15][16].

From an engineering perspective, deployment on edge devices or integration into radiology workstations will require further optimization. Model compression and inference acceleration techniques can reduce computational overhead without compromising diagnostic performance.

In summary, the proposed deep learning framework for mammographic screening demonstrates high diagnostic performance, strong generalizability, and modular design suitable for clinical integration. With continued refinement, such systems can contribute meaningfully to global efforts in early cancer detection and equitable healthcare delivery.

7. Ethical and Practical Considerations

7.1 Ethical Considerations

The deployment of artificial intelligence in medical imaging carries profound ethical responsibilities. While the proposed framework improves diagnostic accuracy, it must be implemented with safeguards to prevent algorithmic bias, ensure equitable performance across demographic subgroups, and maintain human oversight. Models trained predominantly on publicly available datasets (CBIS-DDSM, INbreast) may not fully represent the diversity of real-world populations, potentially leading to disparities in screening outcomes. Therefore, rigorous external validation across heterogeneous clinical settings is essential before clinical adoption. Moreover, the interpretability provided by heatmaps must be accompanied by uncertainty estimates to avoid over-reliance on automated predictions. Transparent reporting of model development, validation protocols, and failure modes is necessary to foster trust and accountability in AI-assisted screening programs.

7.2 Data Privacy and Security

The proposed framework relies on large-scale mammography data, which constitute sensitive protected health information. All experiments were conducted using de-identified public datasets, but real-world deployment would require adherence to regulations such as HIPAA and GDPR. Key considerations include: (i) anonymization of imaging metadata before storage; (ii) secure on-premises or federated learning architectures to prevent data leakage; (iii) strict access controls and audit trails for model inference; and (iv) compliance with institutional review board approvals for prospective evaluation. Any transfer of model parameters or aggregated heatmaps should avoid exposing patient-level information.

7.3 Clinical Integration and Workflow

Successful integration of AI tools into radiology workflows demands more than high AUC scores. The proposed two-stage architecture can be embedded as a decision-support system that prioritizes suspicious cases for expedited review. To minimize disruption, the system should generate predictions within the existing picture archiving and communication system (PACS) interface, providing radiologists with overlays of malignancy probability heatmaps. Workflow considerations include: (i) real-time inference latency < 2 seconds per mammogram) (ii) adjustable sensitivity thresholds to accommodate different screening settings (e.g., higher specificity for recall reduction); and (iii) seamless fallback mechanisms when the AI output is uncertain or when technical failures occur. Pilot implementations with iterative feedback from clinicians are recommended to refine usability and trust.

7.4 Limitations and Future Directions

Despite strong performance, several limitations warrant further investigation. First, the current model is trained for binary malignancy classification, whereas clinical practice involves multi-class distinctions (benign masses, calcifications, architectural distortions). Expanding the framework to multi-label or hierarchical classification would better align with clinical taxonomy. Second, patch extraction still relies on coarse region annotations during training; future work should explore attention-based or fully self-supervised methods to eliminate this dependency. Third, domain adaptation techniques, such as unsupervised adversarial training or style transfer, could improve robustness to variations in imaging equipment and acquisition protocols. Finally, rigorous prospective studies measuring the impact on

radiologist performance, patient outcomes, and cost-effectiveness are necessary to validate the clinical utility of this AI-assisted screening approach.

References

[1] Rebecca L Siegel, Kimberly D Miller, Narges S Wagle, and Ahmedin Jemal. Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73(1):17–48, 2023.

[2] S. A. Qureshi, L. Hussain, T. Sadiq, S. T. H. Shah, A. A. Mir, M. A. Nadim, and S. A. H. Shah. Breast cancer detection using mammography: Image processing to deep learning. *IEEE Access*, 13:60776–60801, 2024.

[3] Liang Shen, Laurie R Margolies, Joseph H Rothstein, Eugene Fluder, Robyn McBride, and Weiva Sieh. Deep learning to improve breast cancer detection on screening mammography. *Scientific Reports*, 9(1):1–12, 2019.

[4] L. Wang. Mammography with deep learning for breast cancer detection. *Frontiers in Oncology*, 14:1281922, 2024.

[5] G Aresta, T Araújo, S Kwok, S Chennamsetty, M Safwan, V Alex, et al. Bach: Grand challenge on breast cancer histology images. In *Medical Image Analysis*, volume 56, pages 122–139, 2019.

[6] A. Sahu, P. K. Das, and S. Meher. An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images. *Biomedical Signal Processing and Control*, 87:105377, 2024.

[7] M. N. Yeasmin, M. Al Amin, T. J. Joti, Z. Aung, and M. A. Azim. Advances of AI in image-based computer-aided diagnosis: A review. *Array*, 23:100357, 2024.

[8] Dezső Ribli, Anna Horváth, Zsuzsa Unger, Péter Pollner, and István Csabai. Detecting and classifying lesions in mammograms with deep learning. *Scientific Reports*, 8(1):1–7, 2018.

[9] W. Gómez-Flores, M. J. Gregorio-Calas, and W. Coelho de Albuquerque Pereira. BUS-BRA: A breast ultrasound dataset for assessing computer-aided diagnosis systems. *Medical Physics*, 51(4): 3110–3123, 2024.

[10] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48, 2019.

[11] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding transfer learning for medical imaging. *Advances in Neural Information Processing Systems*, 32: 3342–3352, 2019.

[12] John R Zech, Marcus A Badgeley, Min Liu, Arthur B Costa, Joseph J Titano, and Eric K Oermann. Variable generalization performance of deep learning models for chest radiograph interpretation. *NPJ Digital Medicine*, 1(1):1–8, 2018.

[13] S. Zafar, A. Hafeez, H. Shah, I. Mutiullah, A. Ali, K. Khan, and G. Leyva-Gómez. Emerging biomarkers for early cancer detection and diagnosis: Challenges, innovations, and clinical perspectives. *European Journal of Medical Research*, 30(1):760, 2025.

[14] Marzyeh Ghassemi, Lauren Oakden-Rayner, and Andrew L Beam. A review of challenges and opportunities in machine learning for health. *NPJ Digital Medicine*, 4(1):1–11, 2021.

[15] Alireza Mehrtash, William M Wells, Clare M Tempany, Purang Abolmaesumi, and Tina Kapur. Confidence calibration and predictive uncertainty estimation for deep medical image classifiers. *Medical Image Analysis*, 59:101627, 2020.

[16] D. A. Milner Jr. and J. K. Lennerz. Technology and future of multi-cancer early detection. *Life*, 14 (7):833, 2024.