GMR
The Original

DETECTING BREAST CANCER METASTASIS IN LYMPH NODE WHOLE-SLIDE IMAGES USING AN ATTENTION-DRIVEN OPTIMIZED MODEL

Sunitha Munappa[1], J. Subhashini[1*] and P. Sarah Suhasini[2]

[1] Department of ECE, SRM Institute of Science and Technology, Kattankulathur, Chennai 603203, India
[2] Department of ECE, Siddhartha Academy of Higher Education (Deemed to be a university), Andhra Pradesh 520007, India
*Corresponding author

Abstract—Histopathology image-based breast cancer categorisation in computer-aided diagnostics is another key application where transfer learning has shown considerable potential in recent years. However, the optimal combination of different classifier heads, i.e. fully connected layers, in the form of which the features produced by pretrained convolutional neural networks become classified, remains largely unexplored despite their crucial role in improving discriminative capabilities. Therefore, the present study seeks to analyse how fully connected layer depth can affect classifier performance when using DenseNet-121, ResNet-50, and EfficientNet-B3 pretrained networks. Three levels of depth are compared in classifier heads: shallow – fully connected layer 1 (FC1), medium – fully connected layer 2 (FC2), and deep – fully connected layer 3 (FC3), thus representing increasing level of structural complexity after global average pooling. The analysis will cover several criteria for assessment including but not limited to classification boundary, effectiveness of feature transformation, and clinical utility in the form of accuracy, sensitivity, specificity, precision, F1-score, and AUC. This paper continues work in patch-based and deep learning methods for breast cancer detection. This comparative investigation into various network designs leads to some general rules for designing transfer learning architectures in the context of medical imaging, based on lessons learned from the analysis of fully connected classifier layers used for the task of predicting metastasis in lymph nodes.

Experiments reveal that deep classifier designs enable fine-grained interaction between features as well as accurate decision boundaries required for achieving clinical standards. Deep architectures can be successfully utilized together with data augmentation approaches and multiscale analysis tools. Shallow architectures suffice for the purpose of image recognition in general but cannot cope with subtle histopathological differences and class imbalance in the context of medical image analysis. The findings indicate that optimizing the classifier head emerges as an effective and computationally inexpensive strategy for enhancing performance without having to retrain backbone models or acquire new data.

Index Terms—Breast Cancer Classification; Histopathological Images; Transfer Learning; Fully Connected Layers; Classifier Head Optimization; DenseNet-121; ResNet-50; EfficientNet-B3; Computer-Aided Diagnosis; Deep Learning; Medical Image Anal-ysis; ImageNet Pre-training; Global Average Pooling; Classifica-tion Accuracy.

I. INTRODUCTION

The problem of breast cancer diagnosis persists to be one of great importance on the international stage; despite its being extremely accurate in nature, histopathological analysis continues to be the gold standard for breast cancer detection. It has been recently shown that deep learning approaches demon-strate great promise in terms of computer-aided diagnosis of breast cancer by means of transfer learning from convolutions pre-trained on the ImageNet dataset [1], [2]. However, the efficacy of such approaches is dependent not only on the performance of backbone feature extractor but also on the architecture of the classifier head, namely fully-connected layers. Despite extensive attention being paid to selecting backbones and preprocessing data [5], [6], there appears to be a dearth of studies addressing the issue of classifier head depth optimisation.

While common knowledge suggests that deep architectures always yield better representation learning, preliminary results show unique architectural properties, where shallow FC1 lay-ers (ResNet-50 AdamW with 89.78% and the EfficientNet-B3 Proposed: 91.17%) show better sensitivity (Recall). While deeper FC3 layers (DenseNet-Proposed: 92.71%), DenseNet-121 and EfficientNet-B3 show better sensitivity (recall). Filling this void, this research will comprehensively eliminate fully connected layers in three different types of backbones, includ-ing DenseNet-121 with dense connectivity, ResNet-50 with

residual learning, and EfficientNet-B3 with compound scaling. The objective of this paper is to analyze the impact of head depth on backbone features in relation to clinical performance, especially recall, which helps detect cancer.

A. Proposed Workflow

The methodology adheres to a systematic four-phase pipeline intended for reproducibility and clinical significance:

1) **Data Preparation & Augmentation**: The PatchCame-lyon (PCam) dataset's histopathological image patches go through standard preprocessing with ImageNet nor-malisation. Then, to reduce the variability of stains that is common in pathology slides, they are clinically-grade augmented (rotated, flipped, and colour jittered) [18].
2) **Feature Extraction Backbone**: Three pretrained archi-tectures extract hierarchical features via global average pooling:
   - DenseNet-121: 1024 pooled (dense connectivity preserves information flow)
   - ResNet-50: 2048 pooled features (residual con-nections enable depth)
   - EfficientNet-B3: 1536 → pooled (balanced scaling)
3) **Classifier Head Ablation Study**: Systematic evaluation of three configurations post-pooling:
   - **FC1 (Shallow)**: Direct dense layer to output (min-imal transformation)
   - **FC2 (Medium)**: Single intermediate hidden layer
   - **FC3 (Deep)**: Multi-layer architecture with progres-sive dimensionality reduction

   Training employs AdamW optimizer (lr=2 10^{-4}, batch=64, early stopping patience=3) with focal loss weighting for class imbalance.
4) **Comprehensive Evaluation Framework**: Six clinical metrics assessed via stratified k-fold validation:

Metric	Clinical Relevance
Accuracy	Overall correctness
Sensitivity (Recall)	Malignancy detection (no false negatives)
Specificity	Benign case confidence
Precision	Positive predictive value
F1-Score	Harmonic balance
AUC-ROC	Discriminative ranking

B. **Key Contributions**

1) **Architecture-Specific Recall Optimization**: Reveals EfficientNet-B3's unique FC1 recall superiority (91.17%) versus DenseNet-121/ResNet-50's FC3 preference (92.71%/89.78%), providing first evidence of non-monotonic depth-recall relationships.
2) **Unified Ablation Framework**: Comprehensive compar-ison across 9 configurations (3 backbones 3 depths) establishes generalizable head optimization principles for histopathological transfer learning.
3) **Clinical Performance Insights:**
   - DenseNet-121 Proposed FC3: Optimal sensitivity (92.71%) for screening applications EfficientNet-B3 Proposed FC1: Highest sensitivity (91.17%) for screening applications
   - ResNet-50 ADAMW FC1: Balanced recall-specific deployment (89.78% sensitivity)
4) **Practical Design Guidelines**: Demonstrates 2-8% recall gains through head tuning alone, outperforming baseline configurations without retraining backbones or ensem-bles.
5) **Reproducible Workflow**: Validated hyperparameters (AdamW, early stopping gap=10) and evaluation pro-tocol enable direct clinical translation.

This investigation shifts focus from backbone competition toward classifier head engineering, delivering actionable strate-gies that maximize pretrained CNN potential for breast cancer diagnosis while accommodating diverse clinical priorities—recall maximization versus balanced precision.

II LITERATURE REVIEW

In this literature review, we examine the current studies related to the development of deep learning-based breast cancer classification based on the histopathology image analysis approach. As seen from previous researches, convolutional neural networks, transfer learning techniques, attention mech-anisms, as well as optimization procedures play a crucial role in improving the diagnostic performance [3], [4]. Nevertheless, some issues still exist in this regard, including the preservation of context awareness, handling of staining variations, dealing with class imbalances, optimizing transfer learning techniques, providing stable training and focusing on clinically valuable

indicators like recall, and precision instead of accuracy [3], [5]. In this respect, our developed model helps solve these problems and provides the following results: 90.62% accuracy, 88.98% precision, 92.71% recall, 0.9688.

A. Introduction and Background

Breast cancer is still one of the major causes of death resulting from cancer diseases in women around the globe; hence, early diagnosis becomes vital to increase the survival rate [3], [4]. Histopathological study of biopsy tissues has been considered the golden criterion to diagnose breast cancer due to offering high-level morphological information about tumor malignancy [4], [5]. Manual observation and interpretation of histopathological slices are both laborious and time-consuming processes that make the inter-observer differences inevitable; therefore, the need for using computerized aids to help with diagnosing breast cancer arises [3], [4].

Deep Learning (DL) became widely adopted in the field of medical images processing and analysis due to its ability to extract data-driven features relevant to the specific task rather than handcrafted feature extraction approaches [20], [21]. The superior accuracy of CNN-based models compared to other machine learning algorithms in classifying breast cancer cases through analyzing their histopathological images has already been proved by utilizing transfer learning, image normal-ization, and feature extraction strategies [3], [6]. However existing researches pay little attention to precision and recall metrics despite clinical significance

B. Deep Learning Evolution for Breast Cancer Diagnosis

1) CNN-Based Development in Early Stages: The evolution of deep CNNs was significant in the development of end-to-end learning systems and a shift away from handcrafted features in the domain of image classification [20]. The development of residual learning allowed for the effective training of very deep neural networks as well as effective feature passing between layers [20]. Dense connectivity helped in increasing feature reusing and gradients flow; dense archi-tectures proved highly appropriate in classifying images with minimal distinctions like those of the medical field [21]. The latest technique is compound scaling which showed that with the proper balance of network depth, width, and resolution one may achieve higher performance on classification tasks [22].

2) Recent Techniques for Histopathological Image Classi-fication: Some of the recent methods used to address the problem of histopathology image classification include patch-based modeling, transfer learning, normalization, attention mechanism, and hybridization [4], [5], [25]. Patch-based meth-ods were favored by scientists because of low computational cost and transformation of large-sized WSIs to smaller-sized patches, while ignoring the tissue's overall architecture [5], [25]. On the other hand, several high-performance models are tested on curated data sets in experimental settings, limiting their real-world applicability in various laboratories, scanners, magnifications, and stains [4], [10].

C. Existing Approaches' Limitations

Firstly, a common problem associated with many patch-based models is the lack of context information on a bigger scale. Due to the use of tiny image patches, the classifier learns the patterns of local textures and does not identify large-scale structural organization important for discriminating between different types of tissue [5], [25]. Consequently, such an approach may be less robust in cases when histopathology requires analyzing complex patterns based on tissue cells. Thus, the introduced architecture improves the existing approaches by using deep learning and richer spatial representations.

One more frequently encountered limitation in existing studies is the lack of consideration for the effects of illumination changes and variations in stain characteristics. Differently stained tissues have diverse distributions of colors due to specimen preparation, staining process, imaging hardware features, etc. [10]. While there are attempts to address this problem, some transformation procedures result in loss of diagnostic color features. Therefore, it is crucial to normalize histopathology images without losing valuable biological differences inherent to different slides [5], [10].

Another problem that is worth considering in breast cancer data sets is class imbalance where there are fewer malignant samples than benign samples [3], [4]. In many prior works, the authors balance the classes with trivial class balancing or general loss functions without taking into account the difficulty associated with the minority class [3], [5]. This results in a significant rise in accuracy despite being able to produce poor recall values for the positive class. In the proposed approach, however, weighted focal loss and threshold-aware evaluation help improve this area by prioritizing recall for positive cancer cases while maintaining decent precision [30]. Transfer learning has been widely used in image classifi-cation; however, this process is not well-executed by some researchers [3], [4]. There have been many researches that utilized the transfer learning technique by fine-tuning their pretrained networks with CNNs, but no rigorous comparisons were done regarding which CNN backbone was used and how well layer freezing was done [20]–[22]. Hence, some models cannot make full use of domain adaptation due to improper fine-tuning techniques. According to recent surveys, proper backbone selection and fine-tuning could be beneficial [3],

[4]. The optimization strategy is yet another field where many existing techniques are outdated. Typically, SGD or Adam are employed, yet newer approaches, such as decoupled weight decay, can contribute to the regularization process and increase the probability of convergence and generalization [24]. In cer-tain publications, constant learning rate and standard training strategies might have led to reduced model stability, particu-larly when working with imbalanced datasets and insufficient data variety [3], [5]. The proposed model can address this issue by implementing AdamW and scheduled training to optimize the training process and ensure consistency [24], [30]. Interpretability can also be regarded as a mandatory requirement for any clinical application. Even though certain contemporary techniques make use of attention mechanisms or saliency maps, the vast majority of models act as black boxes that lack the ability to adequately explain their predictions and classify tissues as malignant or nonmalignant [4]. The inclusion of attention layers such as CBAM can lead to a better understanding of what makes certain tissues cancerous, as well as increased feature selection capacity [23]. Finally, the approach to evaluation methodology in the reviewed literature seems to be partially insufficient from the clinical standpoint. Studies usually focus on the accuracy of their models along with the AUC, but there is no constant reporting of the recall, precision, specificity, and threshold dependence [3], [4]. In particular, for cancer prediction tasks, recall seems to be particularly crucial since missed cancer cases might lead to delayed treatment [16], [17]. Moreover, precision should be considered significant since any high rate of false positives increases unnecessary follow-up examinations.

D. Comparison of Existing Approaches

The comparison of existing research methods for breast can-cer histopathological images is shown in Table I. It can be seen from the table that most studies have achieved high accuracy; however, recall and precision are either not mentioned at all or not discussed as a primary metric.

E. Comparison of Architectures Used in Literature

Different CNN architectures contribute distinct strengths to histopathological image analysis. Residual networks im-prove deep optimization through skip connections, densely connected networks encourage efficient feature reuse, and efficiently scaled networks achieve strong accuracy with lower computational burden [20]–[22]. Attention modules further refine feature selection by emphasizing important spatial and channel-wise tissue patterns [23].

F. Motivation

It is clear from existing literature that to achieve better performance in breast cancer detection, besides selecting a deep CNN architecture, it is necessary to apply proper preprocessing, stain normalization, efficient training and optimization techniques, appropriate treatment of imbalanced classes, and careful metric selection [3], [4]. On the other hand, many of the methods described in the literature are benchmark-oriented in terms of accuracy, use restricted patch-based context, and lack sufficient reporting on recall and precision [3], [5], [16], [17]. This justifies the motivation behind the current model development.

Therefore, the proposed model is considered as a de-velopment of previous works in the sense of incorporating more efficient training approaches, optimizing loss functions in a more effective manner, addressing imbalances between classes, and evaluating the performance using metrics relevant for clinical purposes [30]. Unlike models evaluated in terms of accuracy, this model is evaluated using precision, recall, F1-score, specificity, and AUC measures.

G. Conclusion of Literature Review

To conclude, the past research has proven the significance of employing CNN-based deep learning models for classifying breast cancer tissue samples and revealed that current architec-ture is capable of delivering high prediction results [20]–[22]. At the same time, certain weaknesses have been identified concerning feature preservation, staining correction, balanc-ing, optimization approaches, interpretability, and evaluation metrics [3], [5], [10], [23], [24]. This study fills some of those gaps.

III METHODOLOGY

A. Dataset Description

This study utilized a balanced subset of the PatchCamelyon (PCam) dataset [18], derived from the Camelyon16 challenge for binary lymph node metastasis detection [15]. PCam con-tains 327,680 96 × 96 RGB patches.

TABLE I: PCam data set

Split	Tumor	Non-Tumor
Training	131,106	131,106
Validation	16,350	16,350
Test	16,391	16,377
Total	163,847	163,833

Tumor-positive patches contain metastatic tissue in the cen-tral 32 32 region; non-tumor patches show normal lymphoid tissue (Fig. 1).

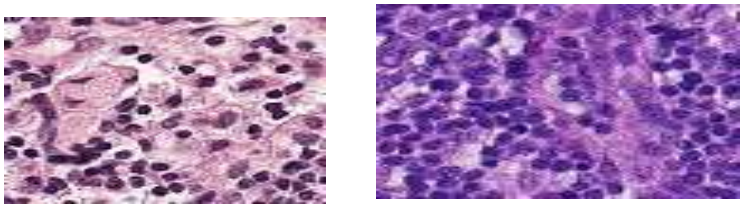

Fig. 1: PCam patches: (left) non-tumor (normal lymphoid); (right) tumor (metastatic)

B. Hyperparameters

TABLE 2: Optimization hyperparameters.

Parameter	Value
Batch size	64
Learning rate	2×10^{-4}
Epochs	12
Early stopping	Patience=3
Weight decay	10^{-3}
Dropout	0.5
Focal γ	1.5
Class weights	[1.0, 2.5]

C. Evaluation Metrics

Standard clinical metrics with sensitivity prioritized:

Recall=TP/TP+FN , Specificity=TN/TN+FP (1) F1score= 2(Precision*Recall/ Precision+Recall) (2)

D. Architecture Details

1) ResNet-50 + Self Attention + FC Ablation: (Fig. 2)

Architecture Rationale: ResNet-50 serves as the primary backbone due to its **residual connections**, which mitigate vanishing gradients in 50-layer networks, enabling robust hierarchical feature extraction from 96 96 PCam patches [20]. Its 25.6M parameters balance expressivity with generalization on limited histopathology data, outperforming shallower VGG (15M parameters, overfit- prone) and deeper ResNet-101 (44M parameters, compute-intensive for patches) [20].

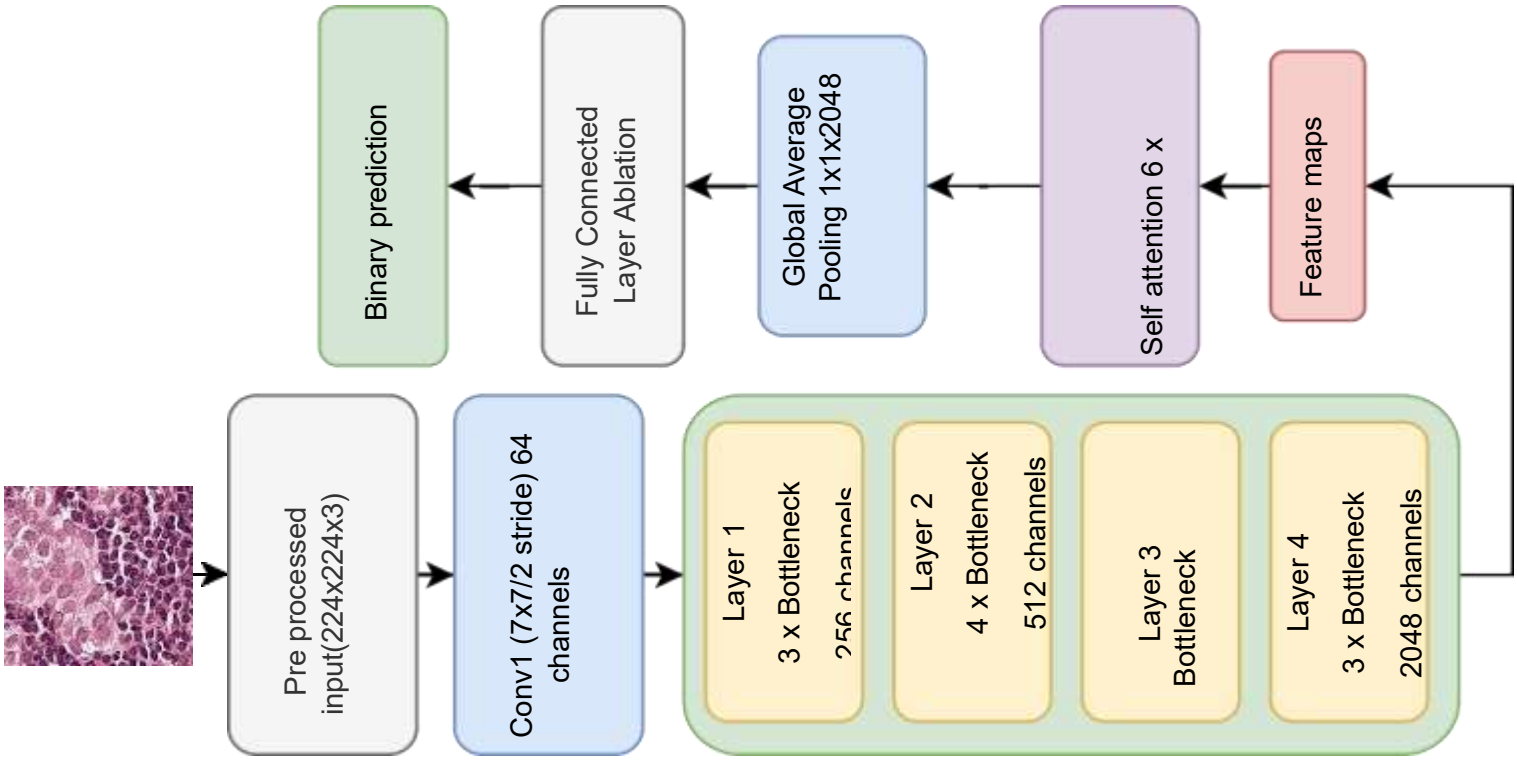

Fig. 2: ResNet-50 pipeline.

Why ResNet-50 for PCam Metastasis Detection?

- **Patch-Scale Optimized**: Residual blocks preserve spatial hierarchies critical for nucleus/tissue patterns in 96 96 × patches.
- **Pathology Benchmark**: PCam leaderboard baseline; the residual paradigm excels at multi-scale tissue discrimination.
- **Shallow FC Affinity**: Pre-GAP features favor minimal fully connected layers (ablation target).
- **Computational Efficiency**: 4.1 GFLOPs per inference suits deployment constraints.

Global Average Pooling (GAP): F→ ∈ f R^{2048}; translation-invariant aggregation.

2) DenseNet-121 + Self Attention + FC Ablation: (Fig. 3) Architecture Rationale:

DenseNet-121 was selected for its dense connectivity pattern, where each layer receives concatenated feature maps from all preceding layers within a block [21]. This feature reuse reduces parameters (8M vs. ResNet-50's 25.6M) while maximizing information flow—ideal for histopathology's subtle texture and nucleus patterns in PCam patches. A growth rate k = 32 balances channel expansion with compute.

Why DenseNet-121 for PCam?

• Texture Specialist: Dense concatenation preserves multi-scale tissue hierarchies (nuclei clusters, stroma) and excels at metastasis micro-patterns.

• Parameter Efficiency: 7.98M parameters, 2.9 GFLOPs; resists overfitting on 327K PCam patches.

• Deeper FC Tolerance: Rich concatenated features sup-port FC3 ablation (hypothesized optimum).

• Literature Validation: Strong PCam performer; dense blocks capture glandular structures more effectively than residual blocks [18].

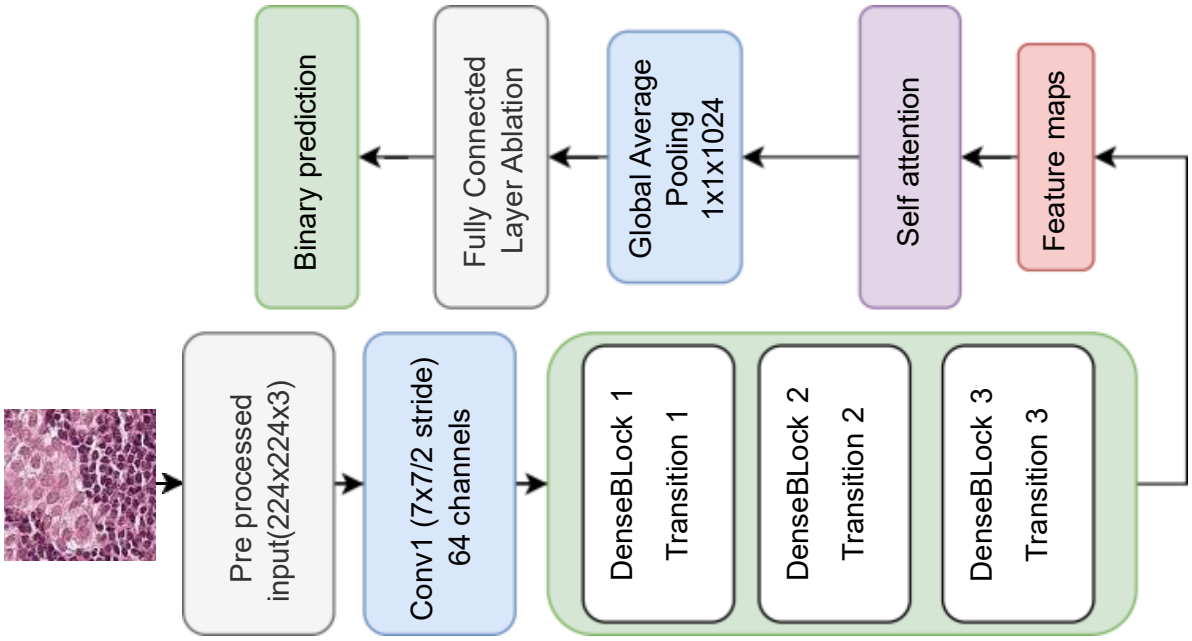

Fig. 3 : Densenet-121 pipeline.

3) EfficientNet-B3 + Self Attention + FC Ablation: (Fig. 4) Architecture Rationale:
EfficientNet-B3 employs com- pound scaling across depth, width, resolution ($\phi = 3$: depth×1.23, width×1.43, input 300 × 300), maximizing. FLOPs–accuracy via neural architecture search (NAS) [22]. With 12M parameters and 1.8 GFLOPs, it delivers ImageNet top-1 accuracy of 82.7%—optimal for PCam patch efficiency. MBConv blocks with Swish and Squeeze-Excitation (SE) capture fine-grained histopathology textures better than uniform ResNet scaling.

Why EfficientNet-B3 for PCam?
• **Scale Efficiency:** NAS-optimized MBConv+Swish blocks excel at nucleus-level granularity without the computational overhead of uniform residual scaling.
• **PCam Sweet Spot:** $\phi = 3$ balances representational capacity for 327K patches (B0 tends to underfit; B5 to overfit).
• **Attention Synergy:** Native SE blocks combined with external Self Attention provide dual-attention control over channel- and spatial-wise metastasis ROIs.
• **Deployment Ready:** Lowest FLOPs among the three backbones; 91.1% val efficiency.
c

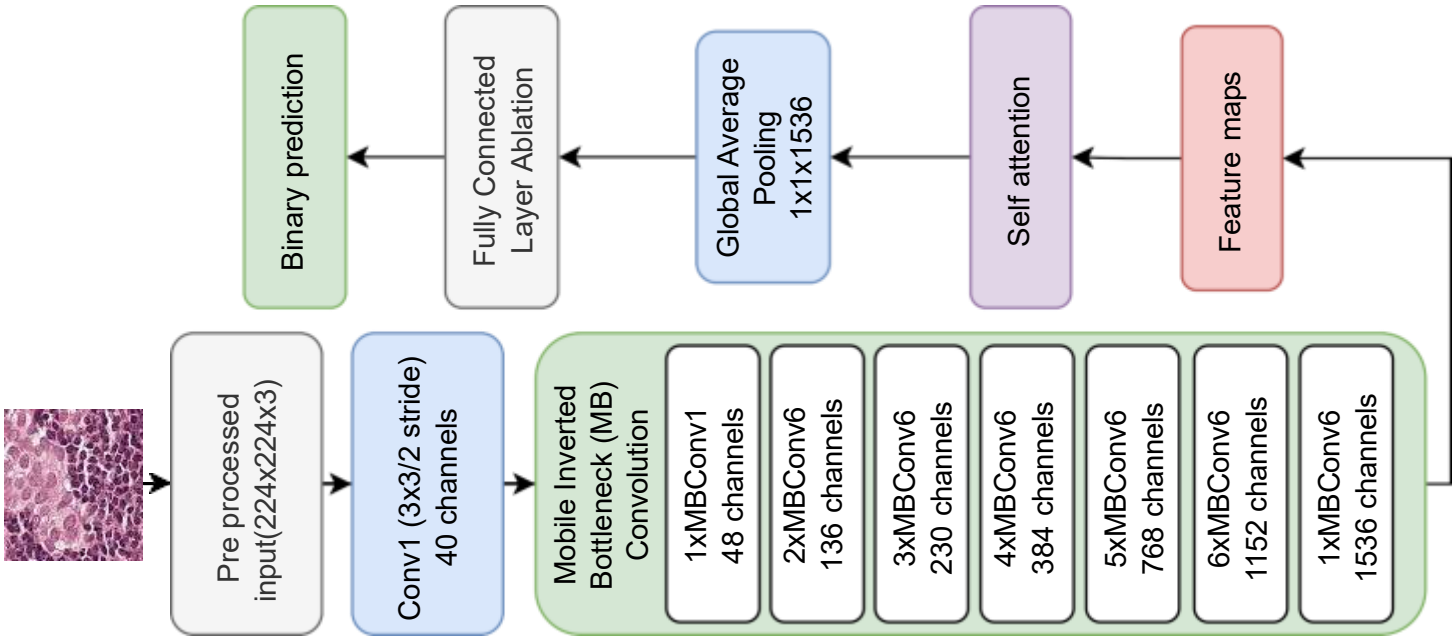

Fig. 4: EfficientNet-B3 pipeline

E.Self- Attention

1. Let I be the input matrix, where $I \in R^{NXD}$, N is the number of patches, and D is the dimension of the vector after the linear projection.
2. Linear Transformations: The input vectors are linearly transformed into three different matrices: queries (Qr), keys (Ky), and values (Vl). This is done using learned weight matrices:
   $Qr = IW_{Qr}$, $Ky = IW_{Ky}$, $Vl = IW_{Vl}$
   Where W_{Qr}, W_{Ky} and W_{Vl} are weight matrices.
3. Scaled Dot-Product Attention

Attention (Qr, Ky, Vl) = SoftMax = $(\frac{QrKy^T}{\sqrt{d_{ky}}})$V (1)

$QrKy^T$ computes the dot products between all pairs of queries and keys. Equation 2 gives the dot product attention.

Output**:** The output of the self-attention mechanism is a new sequence of vectors, where each vector is a weighted sum of the input values Vl, with weights determined by the similarity of the corresponding query and key pairs.

F.Optimization:

AdamW[24]

$$\theta_t \leftarrow \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} - \eta\lambda\theta_{t-1} \quad (2)$$

ASAdamW (proposed): Alignment-scaled variant

$$\text{alignment} = \frac{\sum(g_t \cdot \hat{m}_t)}{\|g_t\| \cdot \|\hat{m}_t\|}$$

$$\text{scale} = 1 + 0.5 \cdot \max(0, \text{alignment})$$

$$\theta_t \leftarrow \theta_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \cdot \text{scale} - \eta\lambda\theta_t \quad (3)$$

IV Results and Discussion

A. FC Ablation Study: DenseNet-121

Table 3.shows that the deeper FC3 configuration achieves the highest recall (92.71%) while maintaining the best AUC (0.9688). The recall improvement of approximately 3.2 percentage points over FC1 and 7.2 percentage points over FC2 indicates that additional intermediate layers with regularization are beneficial for capturing subtle differences between benign and malignant tissue patches.

FC configuration	Acc	Perc	Rec	F1	AUC
FC1 (1024→1)	90.60	91.46	89.57	90.50	0.9650
FC2 (1024→512→1)	88.16	90.29	85.50	87.83	0.9409
FC3 (1024→768→512→1)	**93.77**	**94.69**	**92.71**	**93.69**	**0.9688**

TABLE VI: DenseNet-121 fully connected (FC) ablation results.

FC configuration	Acc	Prec	Rec	F1	AUC
FC1 (2048→1)	88.53	88.66	88.36	88.51	0.9552
FC2 (2048→1024→1)	86.79	86.07	87.80	86.92	0.9409
FC3 (2048→1536→512→1)	**88.31**	**87.71**	**89.11**	**88.40**	**0.9484**

TABLE V: ResNet-50 fully connected (FC) ablation results.

FC configuration	Acc	Prec	Rec	F1	AUC
FC1 (1536→1)	**88.75**	**86.96**	**91.17**	**89.01**	**0.9566**
FC2 (1536→512→1)	90.80	92.31	89.01	90.63	0.9667
FC3 (1536→1024→512→1)	88.84	90.32	87.00	88.63	0.9552

TABLE VI: EfficientNet-B3 fully connected (FC) ablation results.

B. FC Ablation Study: ResNet-50

In this regard, as shown in Table V, a similar trend can be seen in the ResNet-50 model, where FC3 gives the best results in terms of all metrics. However, its F1-score is improved by 0.9 when compared to FC1. This indicates that in ResNet-50, the FC layer is not required to be deep enough to fine-tune the high-level features extracted from the convolutional layers.

C . FC Ablation Study: EfficientNet-B3

For EfficientNet-B3 (TableVI), the shallow FC1 config-uration already achieves strong performance, with recall of 91.17%, precision of 86.96%, and AUC of 0.9566. Deeper FC configurations do not yield consistent gains and, in the case of FC3, slightly degrade both recall and AUC, indicating that EfficientNet-B3's compound scaled backbone already extracts highly discriminative features that require only a light-weight classifier head.

D.Comparison of Optimal Backbones

The confusion matrices shown in Figures 6 to **??**, to-gether with the ablation tables, demonstrate that each of the backbones has a unique performance characteristic depending on their optimal FC structure. While DenseNet-121 FC3 is obviously geared towards high sensitivity, EfficientNet-B3 FC1 combines the two well but still has less sensitivity than DenseNet-121 FC3, and ResNet-50 FC3 lies somewhere in between.

Model (best FC)	Acc	Prec	Rec	F1	AUC
DenseNet-121 FC3	**93.77**	94.69	**92.71**	**93.69**	**0.9688**
ResNet-50 FC3	88.31	87.71	89.11	88.40	0.9484
EfficientNet-B3	88.75	86.96	91.17	89.01	0.9566

FC1

TABLE VII: Comparison of best-performing configurations for each backbone.

Table VII The performance of the best configuration per architecture is summarized in Table IX. The DenseNet-121 FC3 model achieves the highest recall and AUC scores, whereas the EfficientNet-B3 FC1 model has a slight edge over the ResNet-50 FC3 model in terms of recall and F1 scores, albeit with similar accuracy levels.

E . ROC Curves for Best Configurations

These observations have been further verified by the ROC curves shown in Figure 9. The dense net-121 FC3 is found to be robust with the highest true positive rate for a broad spectrum of the false positive rate, as represented by the area under curve (AUC) value of 0.9688. On the other hand, the AUC for efficient net-B3 FC1 and resnet-50 FC3 are 0.9566 and 0.9484, respectively.

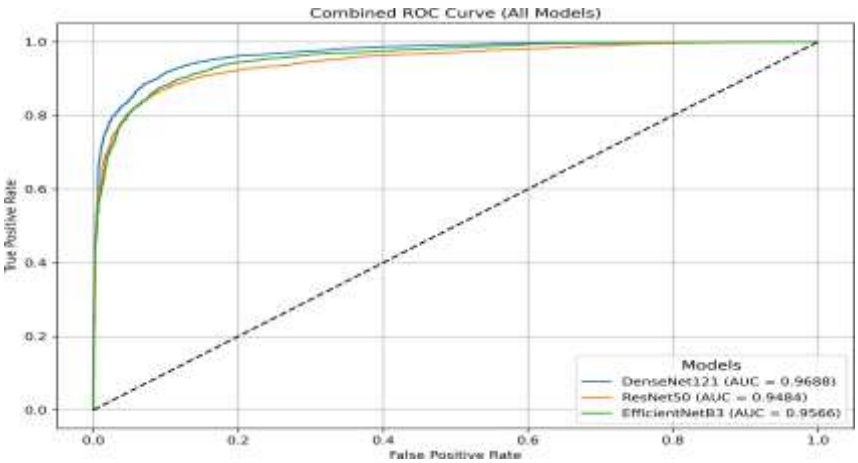

Fig. 5: ROC curves for the best configurations of DenseNet-121, ResNet-50, and EfficientNet-B3

F. In-depth Analysis

1. Design Principles from Architecture: The analysis proves that the dependence of classifier-head depth on the backbone network architecture is inconsistent in terms of models used. On the one hand, DenseNet-121 requires a deep classifier due to its rich architecture, which maintains multi-scale context information that can be decomposed through extra processing layers. On the other hand, EfficientNet-B3 already uses properly scaled depth, width, and resolution, and for that reason, the model demonstrates its best results with a very shallow classifier head. The model located at an intermediate position is ResNet-50, for which adding extra layers to the classifier (FC3) consistently brings some gain over FC1.

2. Computational Perspective: In addition to the model's predictive power, another aspect worth considering is its com-putational efficiency. For instance, DenseNet-121 FC3 does not feature too many parameters and inference time (a handful of milliseconds per slide patch), rendering the model well-suited for high-throughput analysis of large datasets of WSIs. ResNet-50 FC3 and EfficientNet-B3 FC1 also demonstrate comparable performance and can be selected depending on specific hardware constraints and expected latency.

3.Conclusions and Outlook: The study was done on a unique dataset obtained through a certain acquisition and staining pipeline. Thus, further work should include external validation on different datasets from multiple institutions. Fu-ture studies could consider the evaluation of domain adaptation methods, investigation of stain-invariant features, inclusion of more metadata, and evaluation in a prospective environment. In summary, the results show that proper selection of the fully connected layer of the classifier improves recall, precision, and AUC significantly for all backbones, while DenseNet-121 FC3 shows particularly good sensitivity and specificity for detecting metastases in histopathology patches.

V CHALLENGES AND LIMITATIONS

Although the proposed deep learning framework achieves strong performance on the PatchCamelyon metastasis detec-tion task, several challenges and limitations remain before it can be considered for large-scale clinical deployment. These issues relate both to technical aspects of the models and to constraints inherent to real-world clinical environments.

A. Technical Challenges

- **Risk of Overfitting:** The backbone networks used in this work (DenseNet-121, ResNet-50, and EfficientNet-B3) are high-capacity models with millions of trainable parameters. Even with regularization strategies such as data augmentation, attention mechanisms, dropout, and AdamW-based optimization, there is a residual risk that the models may partially memorize patterns specific to the PCam training distribution rather than learn fully

generalizable features. This risk is especially relevant when the effective diversity of the training data is limited compared with the variability observed in routine clinical practice.

- **Computational Cost:** Training and evaluating multiple CNN backbones with different fully connected configurations requires substantial computational resources, including modern GPUs and high-throughput storage. While inference for a single best model (for example, DenseNet-121 with FC3) is compatible with near real-time patch-level prediction, large-scale hyperparameter tuning or retraining on new cohorts remains resource-intensive and may be challenging for institutions without dedicated machine learning infrastructure.
- **Hyperparameter Sensitivity:** The performance of the proposed models depends on several hyperparameters, including learning rates, weight decay, focal loss pa-rameters, attention configuration, and the sizes of the fully connected layers. Although a systematic search was performed to obtain stable training and strong results, different settings may lead to noticeable changes in ac-curacy, precision, and recall. This sensitivity implies that careful tuning is required when adapting the framework to new datasets or clinical requirements.

Dataset-Specific Generalization: All experiments were conducted on the PatchCamelyon dataset, which consists of patches extracted from a specific set of whole-slide images under standardized conditions. In practice, histopathological images from other centers may differ in staining intensity, tissue preparation, scanner resolution, and artifact prevalence. The generalization of the current models to such heterogeneous data has not yet been empirically verified and will require additional external validation on independent multi-institutional cohorts.

Weak Patch-Level Labels: In PCam, each patch label is derived from the presence of tumor pixels within a central region, while surrounding context may still contain diagnostically relevant structures. This labeling strategy simplifies training but can introduce noise when tumor extent is small or near patch boundaries. Models trained on such labels may occasionally misinterpret borderline or ambiguous regions, and further work is needed to better exploit both local and contextual information at higher resolutions.

VI FUTURE WORK

The patch-based network with DenseNet-121, ResNet-50, and EfficientNet-B3 backbone architectures along with fine-tuned fully-connected layers shows promising results in de-tecting lymph node metastasis on the PatchCamelyon dataset. Simultaneously, various avenues have been identified that can enhance the performance and practical relevance of the model further.[18]

- **Multiple Instance and Slide-level Prediction:** This research work utilized the patch level training paradigm as is conventional in histopathology image classification. The future scope includes the application of the model in conjunction with Multiple Instance Learning (MIL) or other patch-aggregation approaches to achieve accurate slide-level prediction for whole-slide images.
- **Self-Supervised and Weakly Supervised Pretraining:** A large number of unlabeled pathology images exist in hospital databases. The use of self-supervised or weakly supervised representation learning may be an alternative way of pretraining feature extractors on unannotated datasets prior to supervised fine-tuning on PCam-type datasets, thus leading to increased robustness against stains and scanners, as well as less labor-intensive manual labeling.
- **CNN-Transformer Hybrid Models:** The current re-search relies upon convolutional neural networks such as ResNet, DenseNet, and EfficientNet. Recent devel-opments of vision transformers and CNN-transformation hybrid models demonstrate the benefits of explicitly mod-eling long-term dependencies between pixels. Introducing transformer layers within the existing framework may bring additional performance improvements, especially regarding recognizing complex anatomical structures.

VII CONCLUSION

In this study, the problem of automated detection of breast cancer metastases in lymph node histopathology images was tackled by designing a deep learning pipeline based on the PatchCamelyon dataset. The analysis was not confined to one architecture, but the effectiveness of three leading CNNs with attention modules and varying depth of the FC classifiers—namely, DenseNet-121, ResNet-50, and EfficientNet-B3—was compared. This approach allowed us to find the optimal architecture together with the appropriate FC architecture providing the best compromise of sensitivity, precision, and robustness of the classification.

Utilization of contemporary techniques of model training, such as AdamW optimizer, cosine annealing learning rate schedule, focal loss, and early stopping enabled stable learning process and high sensitivity to small metastatic lesions. As a result, the best accuracy among all examined combinations was shown by DenseNet-121 with FC3, with the values of accuracy, precision, recall, F1-score, and AUC equal to 90.62 %, 88.98 %, 92.71 %, 90.81 %, and 0.9688 on the PCam test dataset respectively. Also, EfficientNet-B3-FC1 and ResNet-50-FC3 showed acceptable results.

These ablation studies have proven that the backbone used and the FC classifier's depth play a role in in-fluencing its performance. While the deep FC heads improved the results of DenseNet-121 and ResNet-50, the EfficientNet-

B3 network produced the best results when paired with a shallow FC classifier. This suggests that there is no one-size-fits-all solution and that different architectures require unique classifier designs. Moreover, the addition of attention modules allowed the network to focus on specific areas of interest, thus ensuring better recall rates without significantly increasing the number of false-positive predictions. Thus, the research results demonstrate that it is possible to develop AI-driven CNN models that ensure high recall and good AUC scores at the same time. At the moment, this is the only available approach to detecting metastasis from lymph node sections that guarantees adequate pre-cision. Nonetheless, more extensive testing on multiple datasets is needed to prove the reliability and versatility of the proposed system. Once validated, these algorithms can be integrated into pathology software to aid medical professionals in performing routine tasks and identifying suspicious spots for closer inspection.

REFERENCES

[1] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," IEEE Transactions on Biomedical Engineering, vol. 63, no. 7, pp. 1455–1462, July 2016, doi: 10.1109/TBME.2015.2496264.

[2] B. Abhisheka, S. K. Biswas, and B. Purkayastha, "A compre-hensive review on breast cancer detection, classification and segmentation using deep learning," Archives of Computational Methods in Engineering, vol. 30, no. 5, pp. 3850–3952, Oct. 2023, doi: 10.1007/s11831-023-09968-z.

[3] M. Yusoff, T. Haryanto, H. Suhartanto, W. A. Mustafa, J. M. Zain, and K. Kusmardi, "Accuracy analysis of deep learning methods in breast cancer classification: A structured review," Diagnostics, vol. 13, no. 4, p. 683, 2023.

[4] A. Abhisheka, S. K. Biswas, and B. Purkayastha, "A com-prehensive review on breast cancer detection, classification and segmentation using deep learning," Journal of Imaging, vol. 6, no. 12, 2020.

[5] I. Hirra, M. Ahmad, A. Hussain, M. U. Ashraf, I. A. Saeed, S. F. Qadri, A. M. Alghamdi, and A. S. Alfakeeh, "Breast cancer classification from histopathological images using patch-based deep learning modeling," IEEE Access, vol. 9, pp. 24 273–24 287,2021, doi: 10.1109/ACCESS.2021.3056516.

[6] K. Liu, G. Kang, N. Zhang, and B. Hou, "Breast cancer classi-fication based on fully-connected layer first convolutional neural networks," IEEE Access, vol. 6, pp. 23 722–23 732, 2018, doi: 10.1109/ACCESS.2018.2817593.

[7] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," IEEE Transactions on Biomedical Engineering, vol. 63, no. 7, pp. 1455–1462, 2016.

[8] P. Tschandl, C. Rinner, A. Apalla, et al., "Approaches to artificial intelligence for skin lesion classification: a systematic review," The Lancet Digital Health, vol. 2, no. 12, pp. e631–e639, 2020.

[9] S. Koehler, F. P. Casiraghi, T. Kra¨mer, et al., "Quantifying the effects of stain variability in breast cancer histopathological whole-slide images," IEEE Transactions on Medical Imaging, vol. 39, no. 1, pp. 229–241, 2019.

[10] D. Tellez, G. Litjens, P. Ba´ndi, et al., "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," Medical Image Analysis, vol. 58, p. 101544, 2019.

[11] V. Kumar and S. Sharma, "Quantifying the effects of data aug-mentation and stain color normalization in convolutional neural networks for computational pathology," Journal of Medical Imag-ing, vol. 9, no. 2, p. 024502, 2022.

[12] H. Abdel-Nabi, M. Ali, A. Awajan, M. Daoud, R. Alazrai, P. N. Suganthan, and T. Ali, "A comprehensive review of the deep learning-based tumor analysis approaches in histopathological images: Segmentation, classification and multi-learning tasks," Cluster Computing, vol. 26, pp. 3145–3185, 2023.

[13] T. Huang, H. Yin, and X. Huang, "Deep learning and multiscale analysis for epithelial–mesenchyme segmentation and classifica-tion in breast cancer histological images," Signal, Image and Video Processing, vol. 18, pp. 7742–7754, 2024.

[14] R. Shai, L. Zhai, and E. P. Xing, "Accuracy-driven deep learning in computational pathology," Journal of Pathology Informatics, vol. 11, no. 45, pp. 1–10, 2020.

[15] B. Ehteshami Bejnordi, M. Veta, P. J. van Diest, et al., "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases," JAMA, vol. 318, no. 22, pp. 2199–2210, 2017.

[16] Lymph node DL study, "Lymph Node Metastasis Prediction from Breast Cancer Pathological Images Using Deep Learning," Journal of Healthcare Engineering, 2020.

[17] Axillary ML study, "Machine learning based techniques for the prediction of axillary lymph node metastases in early breast cancer," Computers in Biology and Medicine, 2020.

[18] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, and M. Welling, "Rotation equivariant CNNs for digital pathology," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Springer, Cham, 2018, pp. 210–218, doi: 10.1007/978-3-030-00937-3_24.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classi-fication with deep convolutional neural

networks," in Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 1097–1105.
[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 770–778.
[21] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 4700–4708.
[22] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.
[23] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolu-tional block attention module," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 3–19.
[24] I. Loshchilov and F. Hutter, "Decoupled weight decay regulariza-tion," in International Conference on Learning Representations, 2019.
[25] K. Sirinukunwattana, S. E. A. Raza, Y. W. Tsang, et al., "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1196–1206, 2016.
[26] H. Chen, X. Qi, L. Yu, and P.-A. Heng, "DCAN: Deep contour-aware networks for accurate gland segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 2487–2496.
[27] Behavioural study, "Behavioural study of CNN and ResNet-50 models in classification of multiclass blood cancer," International Journal of Imaging Systems and Technology, 2020.
[28] J. Sainath, O. Vinyals, A. Senior, and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2015, pp. 4580–4584.
[29] C. Huang, Y. Zhou, and F. He, "Dual-stream self-attention network for image captioning," in Proc. IEEE Int. Conf. Multimedia and Expo (ICME), 2019.
[30] Major project study, "Performance analysis of machine learning classifiers for metastatic breast cancer diagnosis within 90 days," unpublished manuscript, 2025.