A Hybrid CNN-Transformer Model for Tumor-Infiltrating Lymphocyte Score Prediction in Breast Cancer Histopathological Image

G. K. Shruthi
Department of Computer Science and Engineering, Adichunchanagiri Institute of Technology, Visvesvaraya Technological University, Jnana Sangama, Belagavi, India
shruthigk@aitckm.in (corresponding author)

Pushpa Ravikumar
Department of Computer Science and Engineering, Adichunchanagiri Institute of Technology, Visvesvaraya Technological University, Jnana Sangama, Belagavi, India
pushparavikumar@aitckm.in

Received: 24 October 2025 | Revised: 24 December 2025 | Accepted: 29 December 2025

 DOI: https://doi.org/10.48084/etasr.15757

ABSTRACT

Tumor-Infiltrating Lymphocytes (TILs) are crucial biomarkers in breast cancer, reflecting the immune response and providing prognostic information. Accurate quantification of TILs from whole-slide histopathology images remains challenging due to large image sizes, heterogeneous cellular distribution, and subtle morphological variations. To address these limitations, this study introduces a hybrid deep learning model, combining ResNet-152 and Vision Transformer (ViT) architectures, for automated TIL score prediction using the TIGER WSITILS dataset. Unlike conventional CNN-based approaches, which primarily capture local texture features, or transformer-only methods, which often lose fine-grained spatial detail, the proposed hybrid model leverages the strengths of both architectures—ResNet152 for rich local morphological representation and ViT for modeling long-range spatial relationships across tissue context. The model achieved a Mean Absolute Error (MAE) of 4.64, Mean Squared Error (MSE) of 48.21, and Root Mean Squared Error (RMSE) of 6.94 for regression-based prediction. In classification, it reached 94.74% accuracy with Pearson and Spearman correlation coefficients of 0.9894 and 0.9671, respectively. These results demonstrate that the proposed ResNet152-ViT hybrid framework effectively bridges local and global feature learning, offering improved accuracy, robustness, and interpretability for TIL assessment in breast cancer histopathology.

Keywords-breast cancer; Tumor-Infiltrating Lymphocytes (TILs); ResNet152; Vision Transformer(ViT); Tiger dataset; TILs score prediction

I. INTRODUCTION

Cancer remains a major health challenge worldwide. Among all types of cancer, breast cancer is one of the most prevalent in women and has a high fatality rate [1]. The most widely used technique for breast cancer analysis is histopathological image analysis [2]. Although digital pathology has modernized slide analysis, histopathological assessment remains subjective, affected by variations in expertise, staining, and analytical techniques [3].

In pathology, one of the most rapidly developing research areas is the use of Deep Learning (DL) techniques for analyzing Hematoxylin and Eosin (H&E)-stained Whole-Slide Images (WSIs), which remain the clinical gold standard for cancer diagnosis [4]. DL models can identify intricate patterns in billion-pixel images that may be challenging for human pathologists to detect. In the context of breast cancer, DL-based evaluation of tissue histomorphology has been investigated for predicting patient outcomes [5]. The immune composition of the primary tumor plays a crucial role in the progression of breast cancer, since it mirrors the immune response at the time of diagnosis [6]. In breast cancer, stromal Tumor-Infiltrating Lymphocytes (sTILs) are primarily located in stromal regions adjacent to tumor margins and account for most of the TIL population, whereas intratumoral TILs (iTILs) are found within the tumor mass and constitute a smaller subset. Both the overall density of TILs and the relative distribution of their subtypes within the peritumoral lymphoid infiltrate have been associated with the prognosis of breast cancer [7]. Beyond their prognostic significance, TILs are considered indicators of response to immunotherapy, chemotherapy, and other targeted therapies, and are believed to represent an active anti-tumor immune response within the tumor microenvironment [8].

TILs are a type of immune cell capable of infiltrating the tumor microenvironment. They have been widely recognized as an important biomarker in several types of cancer, including breast cancer, due to their positive correlation with treatment outcomes and overall survival [9]. Currently, TIL quantification and scoring on WSIs are largely performed by pathologists, a process that is subjective and prone to human bias. Automated image analysis techniques are therefore desirable to reduce labor, improve reproducibility, and provide accurate and objective evaluations [10]. In simple terms, DL is a machine learning paradigm that leverages representation learning to automatically extract relevant features from input data [11]. Currently, TIL infiltration is commonly assessed through visual inspection of H&E-stained WSIs following the recommendations of the International TILs Working Group (TILs-WG). These guidelines emphasize identifying the invasive tumor boundary and evaluating only the TILs located within this area. TILs-WG specifically advises concentrating on stromal TILs (sTILs), which are well-defined as mononuclear immune cells—mainly lymphocytes—located in the interior of the tumor margin, residing in the stroma between carcinoma cells but not infiltrating the tumor nests directly [12]. The application of DL in pathology is expected to reduce inter-observer variability, as recommended by the International Immuno-Oncology Working Group [13].

Although prior studies on TIL analysis have largely utilized Convolutional Neural Networks (CNNs), these models often struggle to capture long-range contextual relationships due to their limited receptive fields. Consequently, they may inadequately characterize the spatial distribution and clustering of TILs, particularly in the stromal regions that are critical for clinical interpretation. Furthermore, conventional CNN-based methods are frequently trained on relatively small annotated datasets, limiting their generalizability to diverse real-world tissue samples.

To address these challenges, this study proposes a hybrid DL framework that integrates the extraction of features that are present locally using CNNs with the global attention mechanisms of Vision Transformers (ViTs). Using the WSITILs subset of the TIGER dataset, this approach achieves precise patch-level TIL scoring and robust identification of stromal regions with high TIL density, supporting more interpretable and clinically meaningful assessments.

Several studies have explored the classification of breast cancer images using diverse methodologies and techniques. In [14], a self-supervised learning strategy achieved improvements in tissue segmentation, lymphocyte detection, and TIL prediction accuracy. In [15], two DL models were combined for detecting TILs at the cellular level in breast cancer histopathology slides. This method is applicable for TIL scoring in large-scale cancer slides. Statistical evaluation revealed 95% concordance with pathologists' scores, with a reported p-value of 0.045 (n = 63). In [16], four DL techniques were compared in detecting CD3+ and CD8+ cells in WSIs. This study included methods for "learning to segment cells" using fully convolutional networks and U-Net, and approaches for "learning to regress cell locations" using YOLO and a locality-sensitive strategy. In [17], a fine-tuned Faster R-CNN detector was proposed to automatically identify and count lymphocytes across three types of cancer tissue for TIL scoring. In [18], an innovative Boosted Channels Fusion-based CNN model, termed BCF-Lym-Detector, was proposed for detecting lymphocytes in various cancer histology images. Experimental results showed notable improvements, achieving F-scores of 0.93 and 0.84 on the LYSTO and NuClick datasets, respectively.

II. MATERIALS AND METHOD

A. Dataset Description

This study utilized the TIGER Grand Challenge Dataset, a publicly available benchmark collection designed for TIL scoring and breast cancer research [19]. Among the available subsets (WSIROIS, WSIBULK, WSITILS), the WSITILS subset was selected for this study. Unlike other subsets, WSITILS provides slide-level TILs scores, assigned by a board that is certified by breast pathologists succeeding the recommendations of the international TILs working group. This study focused exclusively on this subset because it (i) directly supports TILs score prediction and (ii) contains standardized expert-validated scores.

B. Overall Pipeline for TILs Score Prediction

Figure 1 shows the architectural diagram of the proposed model. The TIL scoring pipeline begins with data preprocessing and patch extraction from the TIGER dataset. WSIs are first color-normalized, and relevant tissue regions are identified. The model is trained to learn the spatial distribution and density of lymphocytes and other cells. Valid patches are extracted using a sliding window technique and passed into a hybrid model combining ResNet-152 for local features [20] and a ViT [21] for global context. The features extracted from both models are concatenated and advanced through a fusion block, which comprises fully connected layers and a Squeeze-and-Excitation (SE) attention block to enhance informative channel responses. The final regressor predicts a continuous TIL score, which can then be mapped to three categories: Low (<10), Moderate (10–50), and High (>50). The model is trained using AdamW optimizer, a CyclicLRscheduler, and MSE loss, and evaluated using MAE, RMSE, and confusion matrices by comparing patch-level and WSI-level predictions against expert pathologist annotations given in a CSV file in the TIGER dataset.

1) Preprocessing and Patch Extraction

To ensure high-quality WSIs before patch extraction, Shannon entropy (image information content) and Laplacian variance (sharpness) were utilized as quality metrics. Thresholds were defined through iterative testing: patches with Laplacian variance <100 and entropy <3.0 were labeled poorly defined, those with variance <250 and entropy <4.5 as moderately defined, and those with variance ≥250 and entropy ≥4.5 as highly defined. Only highly and moderately defined patches were used for the prediction of the TIL score, as recommended by a pathologist, as poorly defined regions lacked sufficient histological detail.

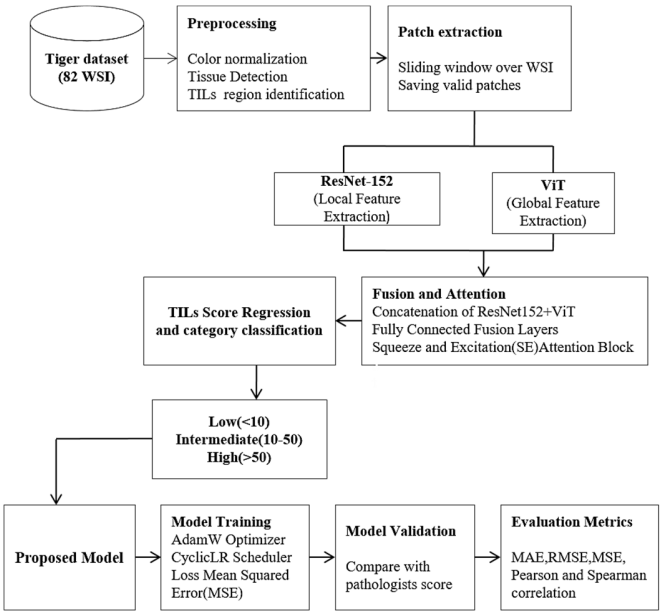

Fig. 1. Architecture diagram of the proposed system.

a) Entropy Calculation

A histogram of the grayscale image pixel intensities (0–255) is created:

$$Gray = 0.299.R + 0.587.G + 0.114.B \quad (1)$$

This histogram is normalized using:

$$p_i = \frac{h(i)}{N} \quad (2)$$

where $h(i)$ denotes the frequency of the i^{th} gray level, and N represents the total number of pixels.

The Shannon Entropy formula is applied, given by:

$$H = -\sum_x p(x) \log_2(p(x)) \quad (3)$$

where H is the entropy of the image, and $p(x)$ is the probability of pixel intensity value x (obtained from the normalized histogram of the image). The summation is over all possible intensity values x (typically 0 to 255 for 8-bit grayscale images). The base of the logarithm is 2, which means that entropy is measured in bits.

b) Laplacian Variance

In the proposed approach, Laplacian kernels are applied to grayscale WSI patches to compute sharpness by emphasizing edges and intensity changes. The variance of the Laplacian response serves as the sharpness score—higher variance indicates better focus. Along with Shannon entropy, this score classifies patches as poorly, moderately, or highly defined for reliable TIL score prediction. The variance of the result can be computed as:

$$Variance = \frac{1}{N}\sum_{i=1}^{N}(I_i - \mu)^2 \quad (4)$$

where I_i is the value of the i^{th} pixel in the Laplacian-filtered image, μ is the mean of all pixel values in the Laplacian-filtered image, and N is the total number of pixels.

c) Extracting Patches from WSI

This step extracts tissue-relevant 224×224 patches from WSIs for tasks such as tumor and TIL classification. Random coordinates are sampled at a chosen resolution (e.g., Level 1), followed by filtering and normalization. Foreground detection removes patches with mostly white background using Otsu's thresholding, retaining only those with less than 85% white pixels to ensure sufficient tissue content. Otsu's method maximizes the between-class variance ($\sigma_b{}^2$), using:

$$\sigma_b^2(t) = w_0(t).w_1(t).[\mu_0(t) - \mu_1(t)]^2 \quad (5)$$

where w_0 and w_1 are the class probabilities and μ_0 and μ_1 are the class means.

Fat detection removes patches dominated by adipose tissue, which appears bright in H&E slides and provides limited diagnostic information. Patches are converted to grayscale and thresholded using Otsu's method, discarding those with more than 80% white pixels. Remaining patches are then color-normalized using Macenko's method, which converts RGB images to Optical Density space to correct staining and scanner variations, calculated using:

$$OD = \log_{10}(\frac{I+1}{25}) \quad (6)$$

where I denotes the observed pixel intensity and I_0 denotes the maximum possible intensity (usually 255 for 8-bit images).

This process estimates stain vectors using Singular Value Decomposition (SVD) and normalizes them to a reference stain matrix, with patches then converted back to RGB. The pipeline ensures that only high-quality, color-consistent, and biologically relevant patches are used for model training and inference.

2) Feature Extraction Using ResNet-152

This study employed ResNet-152 as a deep CNN backbone for extracting robust local features from histopathological patches. ResNet-152 is a widely used architecture known for its deep residual learning framework, which alleviates the vanishing gradient problem by introducing identity shortcut connections. Formally, a residual block learns the mapping $F(X) = H(x) - x$, where $H(x)$ is the desired underlying mapping and x is the input. The output of the block becomes $y = F(x) + x$, enabling effective training of very deep networks.

A ResNet-152 model, pretrained on ImageNet, was employed to leverage transfer learning benefits, accelerating convergence and improving generalization. To adapt the model for feature extraction, the original Fully Connected (FC) classification layer was replaced with an identity function, producing a 2048-dimensional feature vector for each 224×224 input patch. These feature embeddings capture rich local morphological patterns relevant for TILs scoring, providing a strong foundation for subsequent fusion with global contextual features.

3) Feature Extraction Using ViT

The proposed hybrid model utilizes the ViT architecture, specifically the ViT-base-patch16-224 variant, to capture

global contextual information from histopathological image patches for TILs score prediction. ViT uses the self-attention mechanism to simulate long-range dependencies throughout the entire picture patch, which is crucial for capturing spatial linkages between lymphocytes and tumor locations. This is in contrast to traditional CNNs, which mostly concentrate on local information.

The ViT operates by first splitting each input image patch $X = R^{H\times W\times C}$ into a sequence of flattened patches $\{x_1, x_2, \dots, x_N\}$, where each patch $X_i \in R^{p^2.C}$ corresponds to a flattened vector of pixel intensities. These patches are linearly projected into an embedding space. Given an input patch $I = R^{3\times 224\times 224}$, the ViT performs the following steps.

a) Patch Embedding

The image is divided into fixed-size non-overlapping patches of 16×16, forming $N = \frac{224\times 224}{16\times 16} = 196$ patches. Each patch is flattened and projected linearly into a vector $X^i \in R^D$, where $D = 768$.

$$Z_0 = [x_{cls}; x_1E; x_3E; \dots . x_nE] + E_{pos}$$

where $E \in R^{(p^2.C)\times D}$ is the learnable patch embedding matrix and E_{pos} is the positional encoding matrix.

b) Transformer Encoding

A stack of 12 Transformer encoder layers is used to process the embedded sequence. Multi-Head Self-Attention (MHSA), layer normalization, and feed-forward neural networks (FFN) make up each encoder block. MHSA functions as:

$$MHSA\ Attention(Q,K,V) = Softmax = (\frac{QK^T}{\sqrt{d_k}})V$$

These modules enable the model to represent contextual interactions and long-range dependencies among various tissue regions.

c) CLS Token as Global Descriptor

The output representation corresponding to the [CLS] token (i.e., $z_l^0 \in R^{768}$) is extracted and used as a global feature descriptor of the image.

$$f_{vit} = z_l^0 = CLS\ output$$

This vector is then concatenated with the ResNet feature vector and passed to a fusion module for joint learning.

4) Fusion Strategy

To effectively integrate the complementary strengths of convolution- and transformer-based representations, a late fusion strategy combines features extracted from ResNet152 and ViT. This fusion occurs after individual feature extraction, allowing both networks to independently learn rich hierarchical and global contextual representations, respectively. Algorithm 1 outlines the complete fusion pipeline.

Algorithm 1: TIL score prediction using the hybrid ResNet152 and ViT model

Input: Preprocessed image patch $I \in R^{3\times 224\times 224}$

Output: TIL Score $\hat{y} \in R$

1. Extract local features using ResNet152:
   $f_{resnet} \leftarrow ResNet152(I)$ where $f_{resnet} \in R^{2048}$
2. Extract global features using ViT:
   $f_{vit} \leftarrow ViT - Base(I)$ where $f_{vit} \in R^{768}$
3. Concatenate multi-scale features:
   $f_{concat} \leftarrow [f_{resnet}; f_{vit}]$ where $f_{concat} \in R^{768}$
4. Apply feature fusion block (fully connected layers with ReLU and dropout):
   $f_{fused} \leftarrow ReLU(W_2(ReLU(W_1(f_{concat}))))$
5. Apply Squeeze-and-Excitation (SE) block:
   $W_{se} \leftarrow \sigma(W_4.ReLU(W_3.f_{fused})$
   where $W_{se} = R^{256}$, $f_{scaled} \leftarrow f_{fused} \odot W_{se}$
6. Regress final TIL score:
   $\hat{y} \leftarrow W_r.(f_{scaled}) + b_r$

Return $\hat{y}$

III. RESULTS AND DISCUSSION

The proposed model for TIL score prediction from breast histopathology images was assessed using both regression (numeric scores) and classification (Low, Intermediate, High).

A. Quantitative Result

Table I presents the quantitative comparison of ResNet152, ResNet18+ViT, and ResNet152+ViT on the WSITILS dataset. The models were evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE), where lower values indicate better prediction performance. The results show that the ResNet18-only model performed the worst, with substantially higher MAE, MSE, and RMSE values. The integration of the ViT significantly reduced prediction errors, as seen in ResNet18+ViT and ResNet152+ViT. Among these models, ResNet152+ViT achieved the lowest MAE, MSE, and RMSE, indicating its superior ability to capture morphological features and predict TIL scores accurately.

TABLE I. AVERAGE MAE, MSE, AND RMSE ON THE ENTIRE WSITILS DATASET

Model	MAE	MSE	RMSE
ResNet152	19.91	670.78	25.90
ResNet18+ViT	5.17	58.41	7.64
ResNet152+ViT	**4.64**	**48.21**	**6.94**

B. Training and Validation Performance

To evaluate the convergence and stability of the proposed ResNet18+ViT model, MSE, MAE, and RMSE were monitored during training. Figure 2 illustrates the training and validation curves for these three metrics across epochs, showing insights into how well the model learns and whether overfitting occurs during the training process.

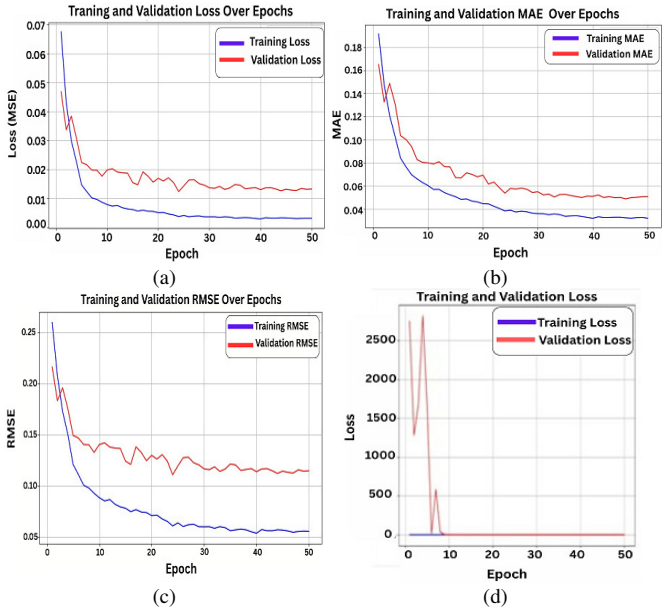

Fig. 2. Training and validation performance of ResNet18+ViT (a-c) and ResNet152+ViT (d) models over 50 epochs.

Figure 2 (a-c) shows that for ResNet18+ViT, MSE, MAE, and RMSE steadily decreased with epochs, with training and validation curves converging after ~20 epochs, indicating good learning and reduced overfitting. Figure 2(d) shows ResNet152+ViT initially unstable but quickly stabilizes near zero, demonstrating faster convergence and better generalization.

C. Per-WSI Prediction Analysis

To evaluate the per-slide prediction quality of the proposed models, the predicted and ground-truth TIL scores were compared for each WSI in the dataset. Scatter plots were generated to visually assess the agreement between predictions and actual values. Figure 3 presents these plots for the ResNet18+ViT and ResNet152+ViT models. The diagonal line in each plot represents the ideal prediction line, where predicted scores would perfectly match the ground-truth scores. Both models closely follow the reference line, indicating strong agreement. ResNet18+ViT exhibited slight underestimation at higher scores, while ResNet152+ViT showed tighter clustering and higher correlations, demonstrating improved feature representation and predictive accuracy.

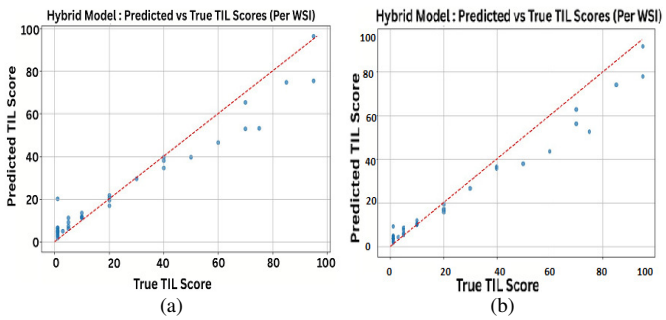

Fig. 3. Scatter plots for (a) ResNet18+ViT and (b) ResNet152+ViT.

D. Correlation Analysis

Agreement between predicted and ground-truth TIL scores was evaluated using Pearson and Spearman correlations. ResNet18+ViT achieved Pearson 0.9801 and Spearman 0.9394, while ResNet152+ViT showed higher correlations (Pearson 0.9894, Spearman 0.9671). For categorical TIL classification, ResNet18+ViT reached 92.11% accuracy, and ResNet152+ViT achieved 94.74%, demonstrating robust performance in clinically relevant TIL groups.

E. Discussion

The ResNet152+ViT model outperformed ResNet18+ViT across all regression metrics (Table I), with lower MAE/RMSE and higher Pearson and Spearman correlations. Scatter plots (Figure 3) demonstrate predictions closer to the ideal line, and the categorical accuracy of 94.74% validates reliable TIL stratification. This improvement is attributed to ResNet152's higher representational capacity combined with ViT's global context modeling.

In comparison with existing TIL assessment approaches, many prior studies primarily focused on cell-level lymphocyte detection or patch-level classification using CNNs, which often lack global contextual understanding across whole-slide images. Recent transformer-based methods attempt to model long-range dependencies but may lose fine-grained morphological details when used independently. In contrast, the proposed hybrid ResNet152+ViT framework effectively integrates local morphological feature extraction with global context modeling, enabling direct slide-level TIL score prediction. Table II summarizes a qualitative comparison with representative existing methods.

TABLE II. QUALITATIVE COMPARISON BETWEEN THE PROPOSED METHOD AND EXISTING TIL ASSESSMENT APPROACHES

Study	Model/Approach	Prediction level	Dataset	Key limitation
[16]	CNN-based lymphocyte detection	Cell-level	IHC images	Does not directly predict slide-level TIL score
[13]	CNN-based feature extraction	Patch-level	Breast cancer WSI	Limited global contextual modeling
[14]	Transformer-based aggregation	Patch aggregation	WSI	High computational complexity
Proposed method	**ResNet152+ViT**	**Slide-level**	**TIGER (WSITILs)**	-

IV. CONCLUSION

This study presented a novel hybrid deep learning framework that integrates ResNet-152 with a ViT for automated quantification of TILs in breast cancer whole-slide histopathological images. The proposed model effectively captures both fine-grained morphological patterns and long-range contextual dependencies across tissue regions, addressing the limitations of conventional CNN-based and transformer-

only architectures. Experimental evaluation on the WSITILS dataset demonstrated that the proposed model achieved superior predictive performance, with a Mean Absolute Error (MAE) of 4.64, Mean Squared Error (MSE) of 48.21, and Root Mean Squared Error (RMSE) of 6.94 for regression tasks. In addition, the model achieved a classification accuracy of 94.74%, with Pearson and Spearman correlation coefficients of 0.9894 and 0.9671, respectively, indicating strong agreement with pathologist-annotated TIL scores. These findings validate the capacity of transformer-based hybrid architectures to deliver robust, scalable, and clinically relevant analysis of the tumor microenvironment. Future work will focus on evaluating generalizability across multi-institutional datasets, incorporating explainable AI techniques, and exploring integration into digital pathology workflows.

REFERENCES

[1] Y. Zhao, J. Zhang, D. Hu, H. Qu, Y. Tian, and X. Cui, "Application of Deep Learning in Histopathology Images of Breast Cancer: A Review," *Micromachines*, vol. 13, no. 12, Dec. 2022, https://doi.org/10.3390/mi13122197.

[2] R. Gurumoorthy and M. Kamarasan, "Breast Cancer Classification from Histopathological Images using Future Search Optimization Algorithm and Deep Learning," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 12831–12836, Feb. 2024, https://doi.org/10.48084/etasr.6720.

[3] R. Rashmi, K. Prasad, and C. B. K. Udupa, "Breast histopathological image analysis using image processing techniques for diagnostic purposes: A methodological review," *Journal of Medical Systems*, vol. 46, no. 1, Dec. 2021, Art. no. 7, https://doi.org/10.1007/s10916-021-01786-9.

[4] M. Tafavvoghi, L. A. Bongo, N. Shvetsov, L. T. R. Busund, and K. Møllersen, "Publicly available datasets of breast histopathology H&E whole-slide images: A scoping review," *Journal of Pathology Informatics*, vol. 15, Dec. 2024, Art. no. 100363, https://doi.org/10.1016/j.jpi.2024.100363.

[5] B. Jiang, L. Bao, S. He, X. Chen, Z. Jin, and Y. Ye, "Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis," *Breast Cancer Research*, vol. 26, no. 1, Sept. 2024, Art. no. 137, https://doi.org/10.1186/s13058-024-01895-6.

[6] A. Fiorin, C. López Pablo, M. Lejeune, A. Hamza Siraj, and V. Della Mea, "Enhancing AI Research for Breast Cancer: A Comprehensive Review of Tumor-Infiltrating Lymphocyte Datasets," *Journal of Imaging Informatics in Medicine*, vol. 37, no. 6, pp. 2996–3008, Dec. 2024, https://doi.org/10.1007/s10278-024-01043-8.

[7] C. Coleman *et al.*, "Harnessing Tumor-Infiltrating Lymphocytes in Triple-Negative Breast Cancer: Opportunities and Barriers to Clinical Integration," *International Journal of Molecular Sciences*, vol. 26, no. 9, May 2025, https://doi.org/10.3390/ijms26094292.

[8] K. El Bairi *et al.*, "The tale of TILs in breast cancer: A report from The International Immuno-Oncology Biomarker Working Group," *npj Breast Cancer*, vol. 7, no. 1, Dec. 2021, Art. no. 150, https://doi.org/10.1038/s41523-021-00346-1.

[9] R. Perera *et al.*, "Annotation-efficient deep learning for breast cancer whole-slide image classification using tumour infiltrating lymphocytes and slide-level labels," *Communications Engineering*, vol. 3, no. 1, July 2024, Art. no. 104, https://doi.org/10.1038/s44172-024-00246-9.

[10] Z. Lu *et al.*, "Deep-Learning–Based Characterization of Tumor-Infiltrating Lymphocytes in Breast Cancers From Histopathology Images and Multiomics Data," *JCO Clinical Cancer Informatics*, no. 4, pp. 480–490, May 2020, https://doi.org/10.1200/CCI.19.00126.

[11] M. Nasser and U. K. Yusof, "Deep Learning Based Methods for Breast Cancer Diagnosis: A Systematic Review and Future Direction," *Diagnostics*, vol. 13, no. 1, Jan. 2023, https://doi.org/10.3390/diagnostics13010161.

[12] M. Verdicchio, V. Brancato, C. Cavaliere, F. Isgrò, M. Salvatore, and M. Aiello, "A pathomic approach for tumor-infiltrating lymphocytes classification on breast cancer digital pathology images," *Heliyon*, vol. 9, no. 3, Mar. 2023, https://doi.org/10.1016/j.heliyon.2023.e14371.

[13] S. Choi *et al.*, "Deep learning model improves tumor-infiltrating lymphocyte evaluation and therapeutic response prediction in breast cancer," *npj Breast Cancer*, vol. 9, no. 1, Aug. 2023, Art. no. 71, https://doi.org/10.1038/s41523-023-00577-4.

[14] S. Kim *et al.*, "Improving Tumor-Infiltrating Lymphocytes Score Prediction in Breast Cancer with Self-Supervised Learning," *Life*, vol. 14, no. 1, Jan. 2024, https://doi.org/10.3390/life14010090.

[15] M. Yosofvand *et al.*, "Automated Detection and Scoring of Tumor-Infiltrating Lymphocytes in Breast Cancer Histopathology Slides," *Cancers*, vol. 15, no. 14, July 2023, https://doi.org/10.3390/cancers15143635.

[16] Z. Swiderska-Chadaj *et al.*, "Learning to detect lymphocytes in immunohistochemistry with deep learning," *Medical Image Analysis*, vol. 58, Dec. 2019, Art. no. 101547, https://doi.org/10.1016/j.media.2019.101547.

[17] I. K. Evangeline, J. G. Precious, N. Pazhanivel, and S. P. A. Kirubha, "Automatic Detection and Counting of Lymphocytes from Immunohistochemistry Cancer Images Using Deep Learning," *Journal of Medical and Biological Engineering*, vol. 40, no. 5, pp. 735–747, Oct. 2020, https://doi.org/10.1007/s40846-020-00545-4.

[18] Z. Rauf, A. R. Khan, A. Sohail, H. Alquhayz, J. Gwak, and A. Khan, "Lymphocyte detection for cancer analysis using a novel fusion block based channel boosted CNN," *Scientific Reports*, vol. 13, no. 1, Aug. 2023, Art. no. 14047, https://doi.org/10.1038/s41598-023-40581-z.

[19] "TIGER - Grand Challenge," *grand-challenge.org*. https://tiger.grand-challenge.org/.

[20] Z. Wang *et al.*, "ResNet for Histopathologic Cancer Detection, the Deeper, the Better?," *Journal of Data Science and Intelligent Systems*, vol. 2, no. 4, pp. 212–220, 2024, https://doi.org/10.47852/bonviewJDSIS3202744.

[21] M. L. Abimouloud, K. Bensid, M. Elleuch, M. B. Ammar, and M. Kherallah, "Advancing breast cancer diagnosis: token vision transformers for faster and accurate classification of histopathology images," *Visual Computing for Industry, Biomedicine, and Art*, vol. 8, no. 1, Jan. 2025, Art. no. 1, https://doi.org/10.1186/s42492-024-00181-8.