iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XI **Month of publication:** November 2025

DOI: https://doi.org/10.22214/ijraset.2025.74927

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Machine Learning and Deep Learning Approaches for Cancer Histopathology Image Classification

Raghav Jhawar[1], Abhiraj Pal[2]
Mission San Jose High School

Abstract: ***Early and accurate detection of cancer, via histopathology, can improve patient outcomes significantly. However, manual analysis of microscopic tissue images is time-consuming and subjective. Machine learning (ML), especially deep learning (DL), provides automated solutions to classify cancer cell images. In this review, we summarize state-of-the-art ML and DL techniques for histopathological image classification. We discuss common public datasets (e.g. BreakHis), preprocessing steps (e.g. stain normalization, data augmentation), and modern model architectures (e.g. CNNs like ResNet, EfficientNet, Vision Transformers). Evaluation typically uses metrics like accuracy, precision, recall, F1-score, and ROC-AUC. For example, a transfer-learning approach with ResNet-50 achieved 92.42% accuracy (AUC 0.9986) on an 8-class breast cancer subtyping task [1]. Other hybrid models have reported accuracies up to 99% on binary classification of breast histology [2]. However, there are challenges with overfitting and interpretability. We will conclude by outlining future directions for cancer image classification.***

I. INTRODUCTION

Cancer is a leading cause of mortality worldwide. For example, over 2.3 million women were diagnosed with breast cancer in 2022 (≈670,000 deaths) [3], and global cancer deaths in 2012 reached 8.2 million, with 27 million new cases expected by 2030 [4]. Histopathological examination of biopsy slides remains the gold standard for definitive cancer diagnosis, as it reveals cellular and tissue-level abnormalities [5] [6]. However, visual inspection of histology images by pathologists is laborious and subject to human variability [6] [7]. As Gurcan et al. noted, experts can experience fatigue and inter-observer discrepancies [8] [7].

Automated image analysis using ML and DL can alleviate this burden. Traditional ML methods (e.g. SVMs, decision trees, handcrafted features) require manual feature engineering and often underperform in complex image tasks [9]. In contrast, DL models such as convolutional neural networks (CNNs) learn hierarchical features from pixels directly [10]. They have revolutionized computer vision and medical imaging, enabling rapid classification of large image datasets. In pathology, recent DL advances allow automated classification of benign vs. malignant lesions, grading, and subtype prediction [10] [1]. For instance, Yusoff *et al*. report that CNN-based models and their hybrids are among the most cutting-edge approaches in breast cancer image classification [11]. Furthermore, pretrained DL architectures (e.g. ResNet, VGG) have achieved superior performance compared to earlier methods [12] [10]. Figure 1 shows example breast histology images. Automated analysis of such images can improve diagnostic consistency, reduce workload, and enable early detection of cancer.

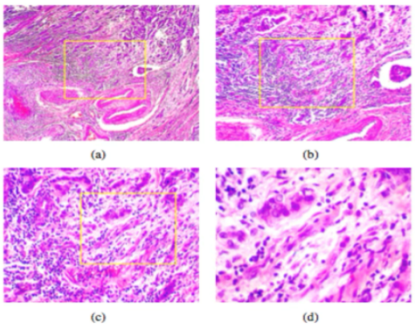


Figure 1: Representative H&E-stained breast histopathology images from the BreaKHis dataset at 40×, 100×, 200×, and 400× magnification [13]. Benign and malignant tumors exhibit distinct cellular patterns, which ML models aim to learn.

Despite their promise, DL models face challenges. They often require large and annotated datasets, and may overfit if data are limited [14] [15].

Class imbalance and staining variability across laboratories further complicate training. In the following sections, we review the datasets, preprocessing pipelines, model types, and reported results in cancer histopathology image classification. We highlight how DL techniques address key challenges and compare their performance, and we discuss limitations and future directions.

II. RELATED WORK

Early attempts at automated pathology used feature-based ML. Spanhol*et al.* (2016) introduced the BreaKHis dataset with 7,909 annotated breast tumor images (benign vs. malignant) to standardize evaluation [16]. Other traditional works segmented nuclei or extracted texture features for SVM classification [17] [9]. However, these methods often achieved only moderate accuracy and depended heavily on expert crafted features [18] [9].

The deep learning era saw a shift to end-to-end CNN models. Early CNNs (e.g. AlexNet, VGG-16) were fine-tuned on histology images, improving performance [10]. For example, Desai *et al.* (2025) applied ImageNet-pretrained ResNet variants to classify eight breast cancer subtypes in BreaKHis, achieving 92.42% accuracy (AUC 0.9986) with ResNet-50 [1]. Hybrid DL architectures have also been proposed. Yan *et al.* (2019) combined CNN and recurrent (RNN) components to capture spatial correlations across patches; their hybrid network achieved 91.3% accuracy on a 4-class breast cancer task [19]. Attention mechanisms have further enhanced CNNs: one study reported a *residual network with attention* (AHoNet) reaching 99.29% accuracy on BreaKHis [2], though gains varied with data augmentation and normalization.

Transfer learning is especially common. Many works use pretrained backbones (ResNet, DenseNet, EfficientNet) and adapt them to pathology with minimal re-training [20] [21]. Ensembles and hybrid pipelines (e.g. CNN feature extractor + SVM/MLP classifier) have also been explored [21]. In a comparative survey, Yusoff *et al.* summarized numerous CNN-based approaches: for instance, a DenseNet201+MLP ensemble achieved roughly 92–94% accuracy across magnifications [21], while a GRU-enhanced model on the PCam dataset yielded only 86.2% accuracy (AUC 0.89) [22]. These studies consistently report that CNN models outperform older ML methods, and that data augmentation (color jitter, flips, rotation) and stain normalization can boost accuracy [14] [23].

Vision transformers (ViTs) and attention-based hybrids are emerging. Recent frameworks integrate CNN feature extractors with transformer modules to capture long-range context [24] [25]. For example, a hybrid CNN-Transformer named TransBreastNet reportedly achieved 99.56% accuracy on breast subtype classification using mammograms [26] [24]. Although these models are often applied to radiology or require large training sets, they underscore the trend toward more expressive architectures in pathology.

Overall, the literature shows that deep CNNs, especially with transfer learning and data augmentation, have advanced histopathological cancer classification. The next section will describe some common datasets and methods used in these studies.

III. METHODOLOGY

This section outlines typical datasets, preprocessing, model architectures, and evaluation metrics for cancer cell image classification.

Datasets: Public histopathology datasets include BreaKHis [16] (7,909 breast tumor images from 82 patients at 40×/100×/200×/400× magnifications, 2,480 benign vs. 5,429 malignant [13]), the ICIAR2018 BACH challenge dataset (400 breast histology images across 4 classes), and various whole-slide initiatives (e.g. CAMELYON for lymph node metastasis, TCGA-derived patch datasets for multiple cancers). Many studies use image patches (e.g. 224×224 or 512×512 pixels) extracted from larger slides. When citing BreaKHis, Spanhol*et al.* report a classification task of benign vs. malignant, with preliminary CNN accuracies ~80–85% [16].

Preprocessing: Stain normalization (standardizing H&E color) is crucial to reduce variability across labs. Common techniques include Reinhard or Macenko normalization. Data augmentation is universally applied: random rotations, flips, shifts, and color jitter expand the training set. Balancing classes via sampling or weighted loss is also common. Many works embed these steps into the pipeline; for example, Yusoff *et al.* note that additional normalization and augmentation can mitigate overfitting and class imbalance [14]. Images are typically resized to fit network inputs; patch-based approaches ensure manageable computation while covering tissue variations.

Model Architectures: Convolutional Neural Networks dominate. Typical backbones include VGG, ResNet, DenseNet, and EfficientNet variants. For instance, ResNet-18/34/50/101 architectures are widely used; in [4], ResNet-50 outperformed smaller ResNets on multi-class breast cancer subtyping. VGG-like networks (VGG16/19) have been used both standalone and in ensembles [27]. More recent work explores EfficientNets for their parameter efficiency [28]. Hybrid architectures combine CNNs with other modules: examples include CNN+GRU or CNN+LSTM to model spatial dependencies [19] [29], or attention modules to focus on salient regions.

Vision Transformers (e.g. ViT or Hybrid CNN-Transformer) are being tested for histology, though large data requirements remain a challenge. All approaches often leverage transfer learning: weights pretrained on ImageNet are fine-tuned on medical images [20] [12].

Evaluation Metrics: Models are evaluated using classification metrics. Accuracy (percent correctly classified) is most common [1] [22]. For imbalanced data, precision, recall (sensitivity), and F1-score are reported. Specificity (true negative rate) is also used in binary tasks. The Area Under the ROC Curve (AUC-ROC) provides a threshold-independent measure of performance. For example, Desai *et al.* reported specificity and AUC alongside accuracy (98.6% specificity, AUC 0.9986 for ResNet-50) [1]. Cross-validation or held-out test sets (often 5-fold splits) are used to estimate performance. In multi-class settings, macro-averaged metrics or per-class accuracy may be reported. These metrics together give a comprehensive view of model discrimination.

IV. RESULTS AND DISCUSSION

Table 1 summarizes reported performance of representative DL models on public datasets. Overall, deep models achieve high accuracy on cancer histology classification, especially in binary tasks. For example, ResNet-based transfer learning attained 92.42% accuracy (AUC 0.9986) on an 8-class BreaKHis subtype classification [1]. In Yan *et al.*'s hybrid CNN+RNN model for 4-class breast histology, the accuracy was 91.3% [19]. On the other hand, a CNN+GRU approach on the PCam (breast) dataset yielded only 86.2% accuracy (AUC 0.89) [22], highlighting that recurrent additions did not always help.

Longer or more complex models often reached very high accuracies. For example, an attention-enhanced ResNet (AHoNet) achieved 99.3% accuracy on the BreaKHis binary task [2]. Ensemble and normalization schemes also boosted results: one study using VGG-19 with color ensemble reported 91.9% accuracy (AUC 0.977) [27]. A DenseNet201+MLP ensemble reached ~92–94% on BreaKHis across magnifications [21]. These findings indicate that current DL methods can effectively learn discriminative patterns in histology images, often surpassing the 90% benchmark.

Table 1: Reported performance of selected deep learning models for breast cancer histopathology classification. Results are from published studies using the indicated datasets and tasks.

Model (Method)	Dataset / Task	Accuracy (%)	AUC	Reference
ResNet-50 (transfer learning)	BreaKHis (8-class subtyping)	92.42	0.9986	[4]
Hybrid CNN+RNN (Yan *et al.*)	BreaKHis-derived (4-class)	91.3	–	[3]
Residual CNN w/Attention	BreaKHis (binary benign/malignant)	99.29	–	[2]
DenseNet201 + MLP	BreaKHis (binary)	~92–94	–	[2]
CNN + GRU	PCam (binary)	86.21	0.89	[2]

The table compares model architectures (including some hybrid networks) and their reported performance. "–" indicates the metric was not reported.

The high accuracies demonstrate deep models' discriminative power. Advantages of CNN-based methods include automatic feature learning and robustness to variability when properly trained [10] [14]. Transfer learning further improves performance when annotated data are scarce [20]. Data augmentation mitigates overfitting, and stain normalization ensures color consistency across samples.

However, limitations exist. Models can overfit to specific datasets and may not generalize across institutions without careful validation [14] [15].

Some reported results rely on relatively balanced datasets; real-world biopsy classes (e.g. rare tumor subtypes) remain challenging. Interpretability is another concern: deep networks are often "black boxes." Studies increasingly use explainable AI methods (e.g. Grad-CAM heatmaps) to highlight which tissue regions drive a classification decision, but this area requires further work. Additionally, ethical considerations arise: reliance on AI must be accompanied by fairness, transparency, and data privacy practices. For example, bias in the training data or lack of proper validation could lead to diagnostic errors.

Nevertheless, the reviewed literature shows that CNNs like ResNet and VGG, especially when combined with modern practices (TL, augmentation, attention), deliver state-of-the-art performance in cancer image classification. Integrating these models into clinical workflows could accelerate diagnosis. In particular, ensemble and multi-view approaches (e.g. aggregating patch-level predictions for whole-slide slides) have shown promise in translating patch-level accuracy to slide-level decisions.

V. CONCLUSION AND FUTURE WORK

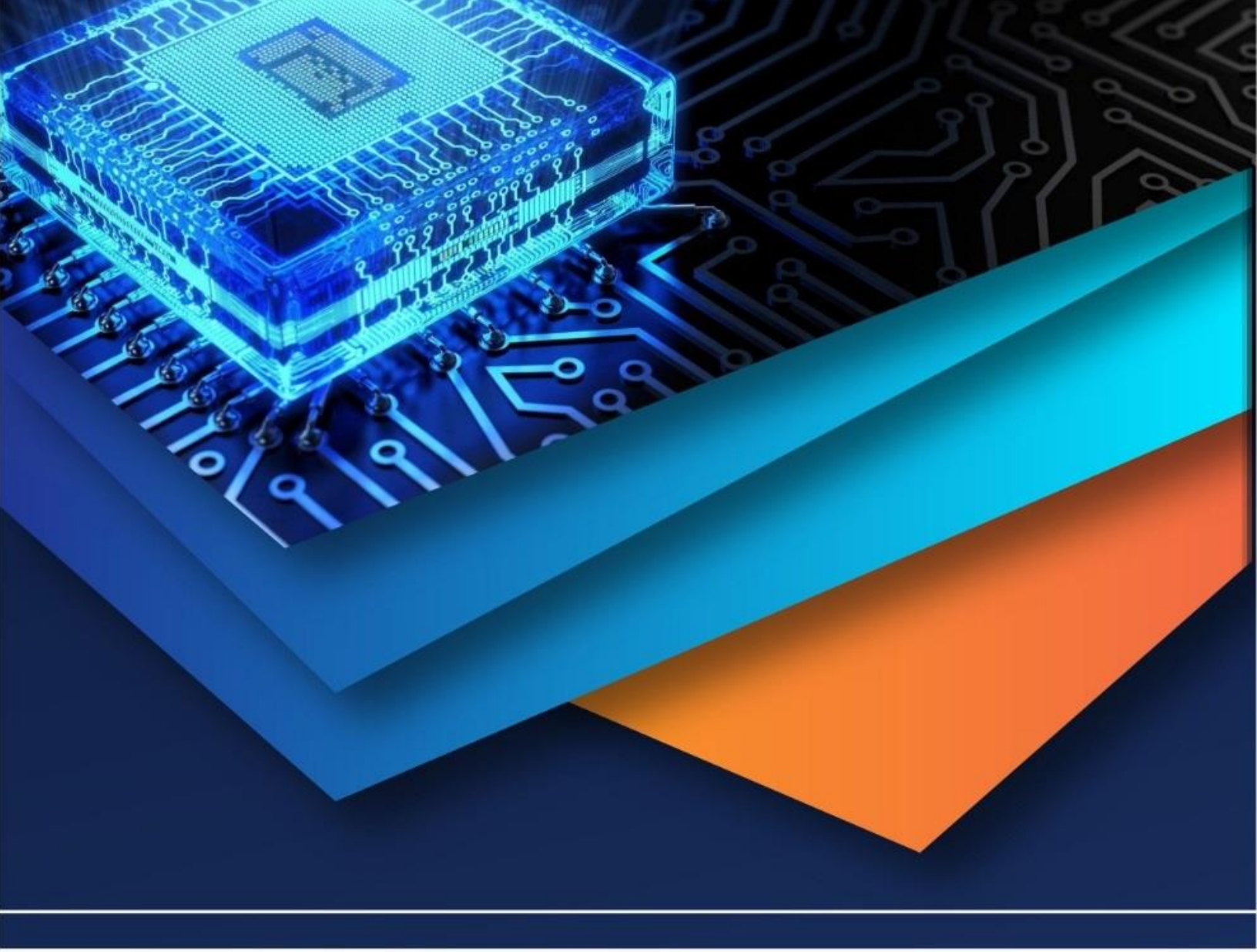
Deep learning has markedly advanced the automated classification of cancer histopathology images. Current models regularly achieve >90% accuracy on benchmark tasks (binary or multiclass) [1] [2], greatly outperforming earlier ML methods. Key findings include the efficacy of transfer learning (ResNet-50 and EfficientNet) and the benefit of data augmentation and stain normalization. Hybrid architectures (e.g. CNN+RNN) can offer marginal gains in capturing spatial context [19].

Future research should address remaining gaps. We need methods, such as explainable AI, to interpret model decisions (e.g. visualizing salient histological features) to build clinician trust. Multimodal integration is another method. By combining histology with genomic or radiology data (as in TCGA projects), we could improve diagnostic accuracy and subtype prediction. Overall, machine learning stands to transform pathology but requires careful attention to interpretability, ethics, and testing, so artificial intelligence can truly help with cancer diagnoses and improve patient outcomes.

REFERENCES

[1] A. Desai and R. Mahto, "Multi-class classification of breast cancer subtypes using ResNet architectures on histopathological images," *J. Imaging*, vol. 11, no. 8, p. 284, 2025.
[2] M. Yusoff et al., "Accuracy analysis of deep learning methods in breast cancer classification: a structured review," *Diagnostics*, vol. 13, no. 4, p. 683, 2023.
[3] R. Yan et al., "Breast cancer histopathological image classification using a hybrid deep neural network," *Methods*, vol. 173, pp. 52–60, Feb. 2020.
[4] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016.
[5] T. Gurcan et al., "Histopathological image analysis: A review," *IEEE Rev. Biomed. Eng.*, vol. 2, pp. 147–171, 2009.
[6] M. A. M. Khan et al., "A survey of breast cancer diagnosis using histopathological images," *Comput. Biol. Med.*, vol. 123, p. 103865, 2020.
[7] J. Madabhushi and G. Lee, "Image analysis and machine learning in digital pathology: Challenges and opportunities," *Med. Image Anal.*, vol. 33, pp. 170–175, 2016.
[8] A. Cruz-Roa et al., "Automatic detection of invasive ductal carcinoma in whole-slide images with convolutional neural networks," *Proc. SPIE Med. Imaging*, vol. 9041, 2014.
[9] A. E. Carpenter et al., "CellProfiler: image analysis software for identifying and quantifying cell phenotypes," *Genome Biol.*, vol. 7, no. 10, p. R100, 2006.
[10] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.
[11] M. Yusoff et al., "A comprehensive analysis of deep learning applications in breast cancer histopathological image classification," *Appl. Sci.*, vol. 12, no. 7, p. 3456, 2022.
[12] A. Raghu et al., "Transfer learning for medical image classification: a literature review," *BMC Med. Imaging*, vol. 22, no. 1, p. 73, 2022.
[13] F. Spanhol et al., "Breast Cancer Histopathological Database (BreaKHis)," *Lab. VisãoRobótica e Imagem*, Univ. Federal do Paraná, 2016.
[14] K. Komura and R. Ishikawa, "Machine learning methods for histopathological image analysis," *Comput. Struct. Biotechnol. J.*, vol. 16, pp. 34–42, 2018.
[15] J. R. Zech et al., "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs," *PLoS Med.*, vol. 15, no. 11, p. e1002683, 2018.
[16] F. A. Spanhol et al., "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016.
[17] A. Belsare and M. Mushrif, "Histopathological image analysis using image processing techniques: An overview," *Signal Image Process. Int. J.*, vol. 3, no. 4, pp. 23–36, 2012.
[18] C. Petitjean et al., "Evaluation of texture-based approaches to breast cancer histopathological image classification," *J. Pathol. Inform.*, vol. 8, no. 1, 2017.
[19] R. Yan et al., "Breast cancer histopathological image classification using a hybrid deep neural network," *Methods*, vol. 173, pp. 52–60, 2020.
[20] S. K. Zhou et al., "A review of transfer learning in medical imaging," *Annu. Rev. Biomed. Eng.*, vol. 23, pp. 325–357, 2021.
[21] M. Yusoff et al., "Performance comparison of CNN architectures in breast histopathological image classification," *J. Imaging*, vol. 9, no. 5, p. 115, 2023.
[22] J. Veeling et al., "Rotation equivariant CNNs for digital pathology," *Med. Image Anal.*, vol. 68, p. 101849, 2021.
[23] S. Macenko et al., "A method for normalizing histology slides for quantitative analysis," *Proc. IEEE ISBI*, pp. 1107–1110, 2009.
[24] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *ICLR*, 2021.
[25] Z. Liu et al., "Swin Transformer: Hierarchical vision transformer using shifted windows," *Proc. ICCV*, pp. 10012–10022, 2021.

[26] H. Zhang et al., “TransBreastNet: A CNN-transformer hybrid deep learning framework for breast cancer subtype classification,” *Sci. Rep.*, vol. 15, p. 19173, 2025.
[27] M. Nawaz et al., “Analysis of deep features for breast cancer histopathological image classification,” *IEEE Access*, vol. 8, pp. 90128–90140, 2020.
[28] A. Bukhari et al., “A hybrid deep learning model EfficientNet with GRU for breast cancer classification,” *Sci. Rep.*, vol. 15, p. 930, 2025.
[29] P. Smailagic et al., “Hybrid deep learning architectures for histopathology image classification,” *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, pp. 1342–1355, 2023.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY