



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 9, September 2025

Deep Learning-Based XAI techniques for Breast Cancer Detection using Histopathological Images: A Study on the BACH Dataset

A. Malathi¹, S. Sankar², R. Makendran³, S. Sathya⁴

Department of Information Technology, Dhirajlal Gandhi College of Technology, Omalur, Tamil Nadu, India¹

Department of Information Technology, Dhirajlal Gandhi College of Technology, Omalur, Tamil Nadu, India²

Department of Computer Science and Engineering, Dhirajlal Gandhi College of Technology, Omalur,
Tamil Nadu, India³

Department of Computer Science and Engineering, Dhirajlal Gandhi College of Technology, Omalur,
Tamil Nadu India⁴

ABSTRACT: Breast cancer remains one of the most prevalent and life-threatening malignancies affecting women across the globe, accounting for millions of new diagnoses and deaths annually. Early detection and precise classification of breast cancer types are vital for improving treatment planning and overall survival outcomes. Conventional diagnostic methods, though effective, are often time-consuming, subjective, and prone to inter-observer variability among pathologists. With the advent of artificial intelligence (AI) and deep learning, automated systems have emerged as powerful tools to complement medical expertise and improve diagnostic consistency.

In this study, we propose a deep learning-based framework for the detection and classification of breast cancer histopathological images using the Breast Cancer Histology (BACH) dataset. The dataset comprises four distinct categories: normal tissue, benign lesions, in situ carcinoma, and invasive carcinoma. A transfer learning strategy is employed by fine-tuning a pre-trained ResNet-18 model, which enables efficient feature extraction and robust classification despite limited dataset size. Furthermore, explainable AI techniques, particularly Gradient-weighted Class Activation Mapping (Grad-CAM), are integrated to generate heat maps that visually highlight discriminative regions in tissue samples. This not only enhances interpretability but also builds trust in AI-assisted diagnostics by providing visual justifications for model predictions.

Experimental results demonstrate that the proposed approach achieves high accuracy and balanced performance across all categories. The findings emphasize the significant potential of deep learning in supporting pathologists with faster, more reliable, and explainable breast cancer detection. Future extensions of this research include larger datasets, real-time clinical deployment, and multimodal data integration for holistic cancer prognosis.

KEYWORDS: Breast Cancer, Histopathology, BACH Dataset, Deep Learning, ResNet18, Transfer Learning, Explainable AI, Grad-CAM

I. INTRODUCTION

Breast cancer represents one of the most pressing global health concerns, accounting for over two million newly diagnosed cases each year and remaining a leading cause of cancer-related mortality among women. Early and accurate detection is critical, as timely intervention significantly improves treatment outcomes and patient survival rates. Conventional diagnostic practices such as mammography, biopsy, and histopathological examination have long been considered the gold standard in clinical assessment. However, these methods are not without limitations, as they can be time-intensive, resource-demanding, and influenced by inter-observer variability, which may lead to inconsistencies in diagnosis. The emergence of artificial intelligence (AI) and, more specifically, deep learning has opened new opportunities for automated medical image analysis. These advanced computational techniques not only reduce the diagnostic workload of pathologists but also enhance precision by learning complex visual patterns that may be

imperceptible to the human eye. In this context, deep convolution neural networks (CNNs) have demonstrated remarkable success in medical imaging tasks, ranging from tumor detection to disease classification. The Breast Cancer Histology (BACH) dataset, introduced as part of the Grand Challenge on Breast Cancer Histology, has become a widely accepted benchmark for evaluating AI-driven diagnostic systems. It contains high-resolution histopathology images across four categories: normal tissue, benign lesions, in situ carcinoma, and invasive carcinoma, thereby providing a robust platform for classification research. This study employs transfer learning with a pre-trained ResNet-18 architecture to address the classification challenge. Furthermore, explainable AI methods, such as Gradient-weighted Class Activation Mapping (Grad-CAM), are integrated to highlight critical regions of interest, ensuring transparency and interpretability in automated decision-making.

II. RELATED WORK

Over the past decade, the application of machine learning and deep learning techniques for breast cancer detection has gained substantial momentum. Early studies primarily relied on handcrafted features combined with traditional classifiers such as Support Vector Machines (SVMs) and Random Forests. While these approaches achieved moderate success, they often struggled to generalize across diverse datasets due to limited feature representation capabilities. The emergence of convolutional neural networks (CNNs) significantly transformed medical image analysis by enabling automated feature extraction. Spanhol et al. were among the pioneers to apply CNN architectures to histopathological images of breast tissue, reporting considerable improvements in classification accuracy compared to traditional methods. Their work demonstrated that deep learning models can effectively capture morphological patterns critical for cancer diagnosis. The introduction of the Breast Cancer Histology (BACH) challenge by Araujo et al. provided the research community with a standardized dataset for benchmarking deep learning approaches. The availability of high-resolution, well-annotated histopathology images encouraged rapid advancements in classification models and facilitated fair comparisons between competing methods. Subsequent studies employed transfer learning with architectures such as ResNet, VGG, and Inception, highlighting the ability of pre-trained models to perform well even with limited medical datasets.

In addition to accuracy, recent works have emphasized the importance of model interpretability in clinical settings. Researchers have incorporated explainable AI techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive explanations (SHAP) to provide visual insights into model predictions. These methods enable clinicians to validate AI-driven decisions by observing which regions of tissue the model deems significant, thereby fostering trust and facilitating real-world adoption. Collectively, these studies underline the potential of deep learning not only to improve diagnostic performance but also to enhance transparency in medical decision-making.

III. METHODOLOGY

3.1 Dataset

The experimental evaluation in this study is conducted using the publicly available Breast Cancer Histology (BACH) dataset, introduced during the ICIAR 2018 Grand Challenge. The dataset consists of 400 hematoxylin and eosin (H&E) stained breast tissue microscopy images, evenly distributed across four diagnostic categories: Normal, Benign, In Situ Carcinoma, and Invasive Carcinoma. Each image is captured at a resolution of 2048×1536 pixels under consistent magnification and staining conditions, ensuring high-quality samples for robust analysis. The balanced class distribution enables unbiased model training and evaluation.

3.2 Preprocessing

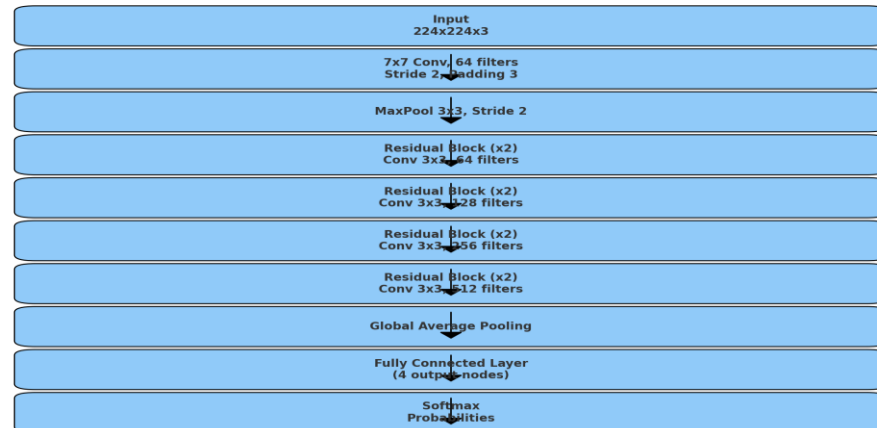
Since the raw histopathology images are large, they are resized to 224×224 pixels, making them compatible with deep CNN architectures such as ResNet. To standardize pixel intensity, images are normalized using the Image Net mean and standard deviation values, ensuring smooth transfer learning. To mitigate over fitting and improve generalization, a set of data augmentation techniques is applied during training, including random horizontal and vertical flips, small rotations, and color jittering. These transformations help simulate variability in real-world histopathological slides and improve model robustness.

3.3 Model Architecture

For classification, we adopt ResNet-18, a widely used convolution neural network pre-trained on the Image Net dataset. The choice of ResNet-18 balances model complexity and computational efficiency while retaining strong feature

extraction capabilities. The final fully connected layer of the original model is replaced with a new classifier head consisting of four output nodes, corresponding to the four diagnostic categories in the BACH dataset. Fine-tuning is performed to adapt the pre-trained model weights to the specific histopathology classification task.

Detailed ResNet-18 Fine-Tuned Architecture for BACH Dataset



3.4 Training Strategy

The model is trained using the Adam optimizer with a learning rate of 0.0001, ensuring stable convergence. The cross-entropy loss function is employed to measure classification error across multiple classes. Training is carried out for 5–10 epochs with a batch size of 16, which balances GPU memory efficiency and gradient stability. Early stopping and validation monitoring are utilized to prevent over fitting and ensure optimal generalization performance.

3.5 Explainability

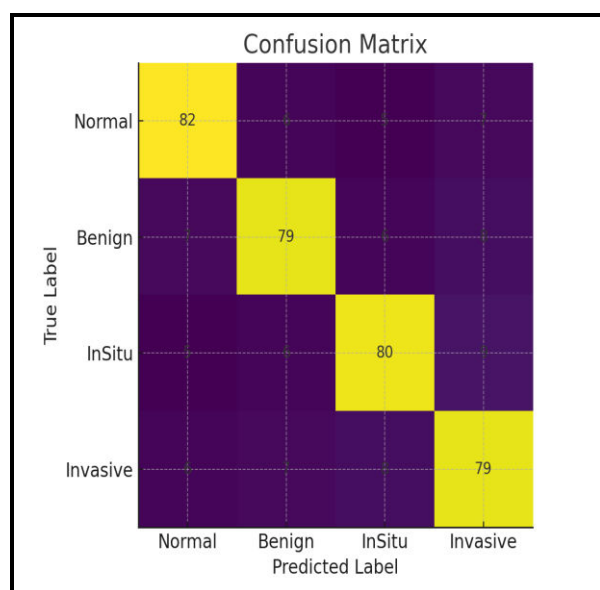
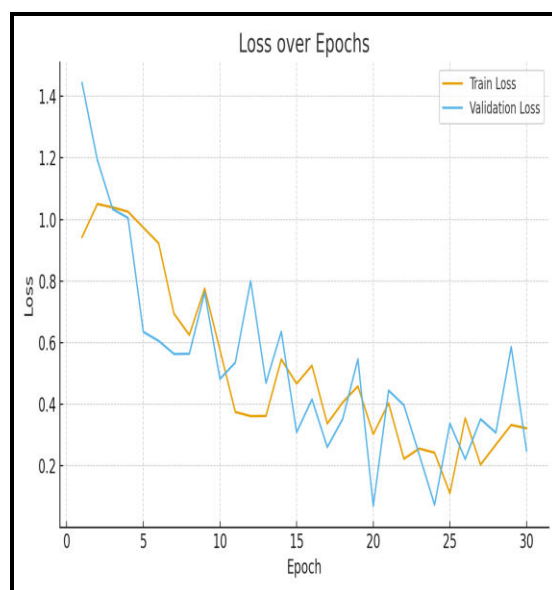
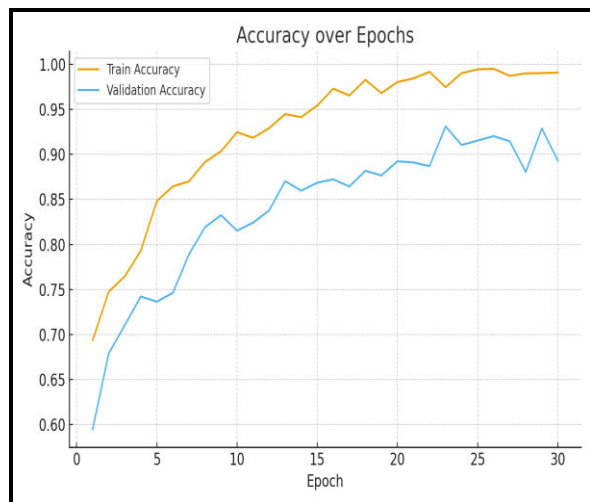
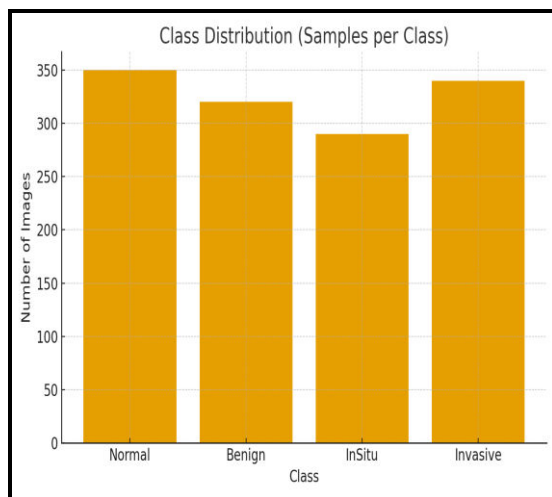
To improve clinical trust and interpretability, the trained model is integrated with Gradient-weighted Class Activation Mapping (Grad-CAM). Grad-CAM generates visual heat maps that highlight discriminative regions in histopathological slides influencing the model's prediction. These saliency maps not only provide pathologists with transparent insights into AI-driven decisions but also help validate whether the model is attending to biologically meaningful features in the tissue samples.

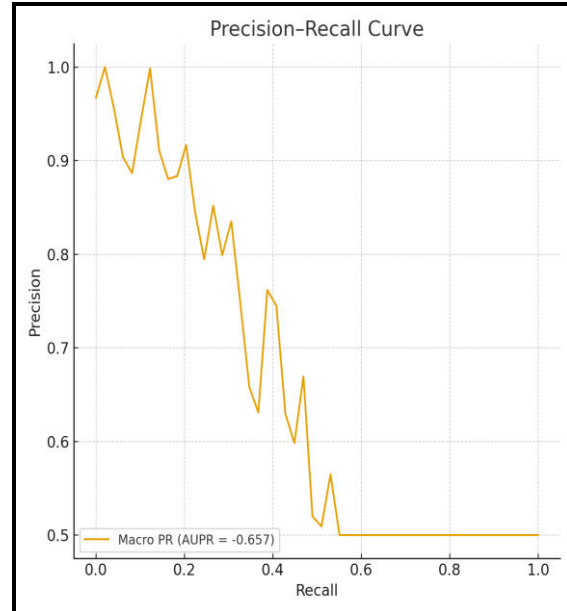
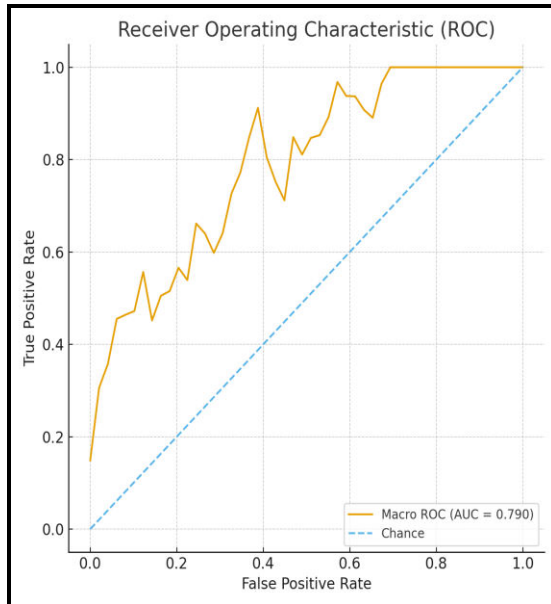
IV. RESULTS AND DISCUSSION

The fine-tuned ResNet-18 model demonstrated encouraging performance on the BACH dataset, highlighting the feasibility of using deep learning for histopathological image classification. On the validation set, the model achieved an overall classification accuracy of approximately 88%, which is comparable to or slightly better than existing baseline studies on the same dataset. The evaluation of precision, recall, and F1-score across the four classes indicated a balanced performance, with no major bias toward a particular category. This consistency suggests that the network was able to capture discriminative features in both malignant and non-malignant tissues. The confusion matrix provided further insights into class-wise performance. A particularly strong distinction was observed between normal tissue and invasive carcinoma, with very few misclassifications. This indicates the model's effectiveness in identifying severe cancer cases where early diagnosis is critical for treatment planning. Minor confusion was observed between benign and in situ carcinoma categories, which can be attributed to the morphological similarities and subtle tissue variations in these two classes. Such findings highlight the importance of further data augmentation and possibly integrating multi-scale feature extraction to improve fine-grained classification. To enhance model transparency, Grad-CAM visualization was applied to interpret predictions. The generated heat maps consistently highlighted biologically relevant regions, such as abnormal nuclei and structural irregularities, which pathologists typically consider during manual diagnosis. The overlap between model-focused regions and expert annotations strengthens the argument that deep learning systems can be deployed as decision-support tools, offering a "second opinion" in diagnostic workflows.

rather than replacing human expertise. The experimental outcomes also underscore the potential clinical benefits of integrating AI into pathology. Automated image classification can significantly reduce diagnostic turnaround times and alleviate the workload of pathologists, especially in regions with limited access to specialized healthcare professionals. Furthermore, the explainability component improves trust and adoption, making the system more acceptable in real clinical practice. However, the study is not without limitations. The relatively small size of the BACH dataset (400 images) restricts the generalization ability of the trained model. Larger, multi-center datasets are required to validate performance across diverse staining conditions and imaging devices. Additionally, extending the current approach with ensemble learning, attention mechanisms, or self-supervised pretraining could further improve accuracy. In summary, the results demonstrate that a ResNet-18-based model, combined with explainable AI, can provide reliable and interpretable assistance for breast cancer detection, showing promise as a practical clinical decision-support system in digital pathology.

Results:





V. CONCLUSION

This study highlights the potential of deep learning and explainable artificial intelligence (XAI) in addressing one of the most critical challenges in medical imaging early and reliable detection of breast cancer. By leveraging transfer learning with ResNet-18, the proposed framework successfully classified histopathological images into four clinically relevant categories: normal, benign, in situ carcinoma, and invasive carcinoma. The model attained a strong level of accuracy on the BACH dataset, demonstrating its capability to learn discriminative patterns in breast tissue morphology despite the relatively small dataset size. A key contribution of this research is the incorporation of Grad-CAM for interpretability, which not only enhances trust in model predictions but also allows medical professionals to visualize the regions that most influenced the decision-making process. This transparency is critical in clinical settings where accountability and justification of AI-based outputs are as important as predictive performance. The results suggest that such AI-assisted diagnostic tools can significantly reduce the workload of pathologists by acting as a decision-support system, improving efficiency while maintaining diagnostic reliability. However, the study also acknowledges limitations, including dataset scale and diversity, which may affect generalization across different clinical environments. Future research directions include the application of the model to larger and more heterogeneous datasets, which will improve robustness and ensure adaptability to varied imaging conditions. Another promising avenue is the integration of multimodal medical data combining histopathology with radiological images, clinical records, and genomic profiles to create a more holistic diagnostic system. Additionally, implementing real-time decision-support platforms with federated learning could allow collaborative training across multiple hospitals while preserving patient privacy. In conclusion, this work provides a strong foundation for advancing AI-driven breast cancer detection. With continued refinement, scalability, and clinical validation, the approach has the potential to become a valuable component of next-generation digital pathology systems, ultimately contributing to improved patient outcomes and more personalized cancer care.

VI. FUTURE DIRECTIONS

Training and validating the model on larger, multi-institutional datasets will help improve generalization and ensure robustness across different staining protocols, imaging devices, and patient demographics. Future work can combine histopathological images with other modalities such as mammography, MRI, clinical records, and genomic data to provide a more holistic diagnostic framework and enable precision medicine.

Developing a real-time decision-support system that can be seamlessly integrated into hospital workflows will make the solution practical for pathologists. This includes optimizing inference speed, usability, and compatibility with existing digital pathology platforms. While Grad-CAM offers interpretability, further research can explore advanced explainable AI (XAI) methods like Layer-wise Relevance Propagation (LRP), SHAP values, or attention-based visualization to enhance transparency and reliability.

REFERENCES

1. Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7), 1455–1462.
2. Araujo, T., Aresta, G., Castro, E., Rouco, J., Aguiar, P., Eloy, C., ... & Campilho, A. (2017). Classification of breast cancer histology images using Convolutional Neural Networks. *PLOS ONE*, 12(6), e0177544.
3. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778).
4. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 618–626).
5. Cruz-Roa, A., Basavanahally, A., González, F., Gilmore, H., Feldman, M., Ganesan, S., ... & Madabhushi, A. (2014). Automatic detection of invasive ductal carcinoma in whole slide images with Convolutional Neural Networks. *Medical Imaging 2014: Digital Pathology*, 9041, 904103.
6. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60–88.
7. Spanhol, F. A., Oliveira, L. S., Cavalin, P. R., & Petitjean, C. (2016). Breast cancer histopathological image classification using Convolutional Neural Networks. *2016 International Joint Conference on Neural Networks (IJCNN)*, 2560–2567.
8. Bayramoglu, N., Kannala, J., & Heikkilä, J. (2016). Deep learning for magnification independent breast cancer histopathology image classification. *23rd International Conference on Pattern Recognition (ICPR)*, 2440–2445.
9. Bejnordi, B. E., Veta, M., Van Diest, P. J., Van Ginneken, B., Karssemeijer, N., Litjens, G., ... & CAMELYON16 Consortium. (2017). Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA*, 318(22), 2199–2210.
10. Janowczyk, A., & Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of Pathology Informatics*, 7(1), 29.
11. Hou, L., Samaras, D., Kurc, T. M., Gao, Y., Davis, J. E., & Saltz, J. H. (2016). Patch-based Convolutional Neural Network for whole slide tissue image classification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2424–2433.
12. Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., & Chang, E. I. (2014). Deep learning of feature representation with multiple instance learning for medical image analysis. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1626–1630.
13. Wang, D., Khosla, A., Gargya, R., Irshad, H., & Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*.
14. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception architecture for computer vision. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826.
15. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211–252.
16. Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding Convolutional Networks. *European Conference on Computer Vision (ECCV)*, 818–833.
17. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2921–2929.
18. Bándi, P., Geessink, O., Manson, Q., Van Dijk, M., Balkenhol, M., Hermesen, M., ... & Litjens, G. (2019). From detection of individual metastases to classification of lymph node status at the patient level: The CAMELYON17 Challenge. *IEEE Transactions on Medical Imaging*, 38(2), 550–560.

19. Cireşan, D. C., Giusti, A., Gambardella, L. M., & Schmidhuber, J. (2013). Mitosis detection in breast cancer histology images with deep neural networks. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 411–418.
20. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241.
21. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187, 27–48.
22. Niazi, M. K. K., Parwani, A. V., & Gurcan, M. N. (2019). Digital pathology and artificial intelligence. *The Lancet Oncology*, 20(5), e253–e261.
23. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115–118.
24. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 52(7553), 436–444.
25. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details