

DEEP HYBRID CNN-VIT MODEL FOR ACCURATE CANCER DIAGNOSIS FROM HISTOPATHOLOGICAL IMAGES

Sheela s

Department of Computer Science and Engineering
C Byre Gowda Insistute of Technology Kolar
Sheelashamika08@gmail.com

Ayesha sana

Department of Computer Science and Engineering
C Byre Gowda Insistute of Technology Kolar
ayashaafiyah7@gmail.com

Abstract: Cancer diagnosis using histopathological images is One of the most important and difficult tasks in medical picture analysis. mechanisms, transformer-based architectures, and multi-scale learning, a thorough grasp of these methods and their efficacy is still required .Conventional deep learning algorithms ignore the surrounding tissue context, which is crucial for a precise diagnosis, and instead concentrate mostly on tumour regions 41. Convolutional neural networks (CNNs), Vision Transformers (ViTs), and hybrid models are examples of recent developments in deep learning that have made it possible to extract both local and global characteristics from dataimages. This survey paper reviews various deep learning approaches for cancer detection, with a special focus on models that incorporate surrounding tissue information. We analyze different techniques, datasets, performance metrics, and limitations. Furthermore, we highlight current challenges and future research directions for developing efficient and interpretable diagnostic systems.

Keywords: Deep Learning, Histopathology, Cancer Detection, CNN, Vision Transformer, Tissue Context

I. INTRODUCTION

All Early diagnosis is essential to increasing patient survival rates because cancer is still one of the world's top causes of death [1]. [2]. Histopathological image analysis, which involves examining tissue samples under a microscope, is considered the gold standard for cancer detection. Traditionally, this process relies on expert pathologists to analyze complex tissue structures, making it time-consuming and subject to inter-observer variability [3], [4].

The quick development of artificial intelligence, especially deep learning, has led to the widespread use of automated methods for 51 analysis of histopathology images. Convolutional Neural Networks (CNNs) and other deep learning models have demonstrated impressive 10 success in extracting discriminative features from medical pictures, allowing for effective and precise cancer classification [5], [6]. By capturing both local and global features, more recent advanced architectures like Vision Transformers 10 (ViTs) and hybrid CNN-Transformer models have significantly enhanced performance [7], [8]. Despite these advances, a major disadvantage of many current techniques is their concentration on tumour sites while ignoring the surrounding tissue background.

Accurate diagnosis in histopathology relies heavily on the interaction between tumour cells and surrounding tissues, including stromal elements and cellular microenvironments [9]. Ignoring this background knowledge can result into incomplete feature representation and reduced diagnostic accuracy [10]. Therefore, the primary problem addressed in this survey is the lack of effective modelling of surrounding tissue information in traditional deep learning approaches for cancer diagnosis. Although recent methods attempt to incorporate contextual features using attention mechanisms, multi-scale learning, and transformer-based architectures, there is still a need for a comprehensive understanding of these techniques and their effectiveness [11], [12].

With a focus on techniques that take advantage of the surrounding tissue context, this article intends to provide a thorough evaluation of deep learning models for cancer diagnosis using histopathology pictures. CNN-based approaches, transformer-based models, hybrid architectures, patch-based strategies, and multi-instance learning frameworks are all included in this survey. The survey also examines popular datasets, performance indicators, major obstacles, and potential avenues for further study in this field.

II LITERATURE REVIEW

With a focus on how various models handle surrounding tissue context, this section examines contemporary deep learning techniques for using histological pictures to identify cancer.

A. CNN-Based Approaches

Because of their strong ability to extract local spatial features, Convolutional Neural Networks (CNNs) have been utilized widely for the classification of histopathological images. Recent studies (2023–2025) demonstrate that architectures such as ResNet and Dense Net achieve high accuracy in detecting cancerous regions from tissue images [13], [14]. However, CNN-based approaches primarily focus on localized features, which limits their ability to capture long-range dependencies across tissue regions. As a result, these models often fail to incorporate surrounding tissue context effectively, leading to suboptimal performance in complex diagnostic tasks [15].

B. Patch-Based Approaches

To handle large whole-slide images (WSIs), many methods divide images into smaller patches and perform classification at the patch level. This reduces computational complexity and enables scalable training [16]. However, patch-based approaches process each patch independently, which leads to loss of spatial relationships between neighboring tissues. Recent works highlight that ignoring contextual dependencies can significantly reduce diagnostic accuracy [17].

C. Transformer-Based Approaches

Vision Transformers (ViTs) have recently been introduced for medical image analysis. Unlike CNNs, transformers model images as sequences of patches and capture long-range dependencies across the entire image [18]. Recent studies show that transformer-based models outperform CNNs in capturing global tissue structure and contextual information [19]. However, these models require large datasets and are computationally expensive.

D. Context-Aware and Attention-Based Models

To overcome the limitations of CNNs and patch-based methods, context-aware models use attention mechanisms and multi-scale feature extraction to incorporate surrounding tissue information.

E. Hybrid CNN-Transformer Models

Hybrid models combine CNNs and transformers to leverage the strengths of both approaches. CNN layers extract local features, while transformer layers capture global dependencies. Recent works report that hybrid architectures achieve state-of-the-art performance in histopathological image classification tasks [22], [23]. Recent research (2025) demonstrates that attention-based models significantly improve classification accuracy by focusing on both tumour regions and adjacent tissues [20], [21].

F. Multi-Instance Learning (MIL) Approaches

Multi-Instance Learning (MIL) is widely used for whole-slide image analysis, where labels are available only at the slide level. Attention-based MIL models identify important regions while considering global context [24].

III EXISTING SYSTEM

The existing systems for cancer diagnosis using histopathological images primarily rely on traditional machine learning and early deep learning techniques. Conventional approaches use handcrafted features such as texture, color, and shape descriptors extracted from tissue images. Machine learning algorithms like Random Forests and Support Vector Machines (SVM) are then used to classify these features. Nevertheless, these techniques frequently fall short of capturing intricate patterns found in histopathology data since they mostly rely on human feature extraction. Convolutional Neural Networks (CNNs) have become widely used for automated cancer detection with the growth of deep learning technologies. Because they can automatically extract key characteristics from histopathology pictures, models like ResNet and DenseNet perform well. Nevertheless, the majority of these models primarily concentrate on locating tumor locations and frequently ignore the surrounding tissue, which is crucial for precise diagnosis. Many techniques split big whole-slide photos into smaller pieces. Although this reduces computing complexity, the model's comprehension of the overall tissue structure is limited as a result of the loss of spatial linkages between nearby regions. Furthermore, global interactions between various tissue components are difficult for conventional CNNs to capture.

These current techniques still have problems including poor interpretability, reliance on sizable labeled datasets, and lack of context awareness, although offering acceptable accuracy. These problems highlight the need for more sophisticated models that can efficiently utilize data on the tumor and surrounding tissue to improve cancer diagnosis.

IV PROPOSED SYSTEM

By combining surrounding tissue context with tumor region analysis utilizing cutting-edge deep learning algorithms, the suggested method seeks to increase the accuracy of cancer diagnosis using histopathology pictures.

This system uses a hybrid architecture that combines Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) to capture both local and global information, in contrast to traditional approaches that primarily concentrate on localized tumor traits.

To guarantee data quality and consistency, the histopathology images first go through preprocessing procedures such patch extraction, noise reduction, and normalization.

While the transformer module records long-range dependencies and relationships between various tissue regions, the CNN component is in charge of extracting fine-grained local information, such as texture, shape, and cellular structures. This makes it possible for the model to comprehend how tumor cells interact with surrounding tissues, which is essential for a precise diagnosis. A thorough depiction of the tissue architecture is then created by combining the extracted local and global characteristics using feature fusion techniques. In order to improve classification accuracy and interpretability, attention processes are integrated to highlight diagnostically significant regions and suppress unimportant background information. In order to forecast the existence of cancer or subtype categories, the fused features are then run through fully connected layers and a softmax classifier. Overall, the proposed system provides a more robust, context- aware, and reliable framework for cancer diagnosis compared to existing approaches.

TABLE1- COMPARISION OF DEEP LEARNING MODELS

Model Type	Example Model	Accuracy (%)	Complexity	Context Awareness
CNN-Based	ResNet, DenseNet	85–90	Low–Medium	Low
Patch-Based CNN	Patch CNN	86–91	Medium	Low
Transformer-Based	(ViT)	90–94	High	High
Attention-Based	Attention CNN	91–95	Medium–High	Medium–High
Hybrid CNN–Transformer	CNN + ViT	93–97	High	High
MIL-Based	Attention MIL	90–94	Medium	Medium

The comparative analysis table-1 presents a detailed evaluation of various deep learning approaches used for cancer diagnosis based on histopathological images, considering key parameters such as accuracy, computational complexity, context awareness, and overall performance. It is evident from the comparison that conventional CNN-based models offer more effective local feature extraction and have less computational complexity. However, its diagnostic accuracy is comparatively low due to its inability to gather global context information. Patch-

based approaches improve computational efficiency by dividing large images into smaller regions, but suffer from the loss of spatial relationships between patches, limiting their effectiveness in capturing the surrounding tissue context.

Traditional CNN models are computationally efficient and very good at extracting local features from images. However, they cannot capture the overall context of the image, which can reduce diagnostic accuracy. Patch-based methods improve efficiency by dividing large images into smaller parts, but they often lose the spatial relationships between regions, making it harder to understand surrounding tissue information.

Transformer-based models, on the other hand, are better at capturing long-range relationships and the overall structure of the image, leading to higher accuracy. However, they require more computational power and larger datasets. Attention-based methods improve performance by focusing on important regions of the image, helping in better interpretation of complex patterns.

Overall, the hybrid CNN-Transformer model performs the best by combining the strengths of both approaches. It effectively captures both local details and global context, resulting in improved accuracy and robustness. However, its performance depends on an effective aggregation strategy.



Figure 1- Proposed Deep Learning Models Targeting Surrounding Tissue for Accurate Cancer Diagnosis Using Histopathological Images.

Figure 1 shows the proposed context-aware deep learning framework for cancer diagnosis using histopathological images. The process begins by capturing a high-resolution whole slide image (WSI), which is first subjected to preprocessing steps such as point normalization, denoising, and image resizing to ensure data consistency.

These processed images are divided into smaller parts during the chunk extraction phase, allowing efficient processing of large images while preserving important spatial information. Data augmentation techniques such as rotation, flipping, and scaling are used to improve model generalization. The framework then uses two feature extraction mechanisms. Convolutional neural networks (CNNs) are used to extract local features by capturing details such as cellular structures and texture patterns.

In parallel, a Transformer or attention-based module processes image fragments to model the overall context and learn relationships between different tissue regions. The results of both modules are then combined in a feature fusion step.

In this step, local and global representations are integrated into a single feature vector using an attention-based method. Finally, the combined features are passed through a fully connected layer for classification, and the system predicts whether the tissue is cancerous or not, and also estimates the probability.

The figure shows that integrating local and global feature learning improves diagnostic accuracy, reliability, and interpretability, making the proposed system suitable for clinical decision support.

ACKNOWLEDGMENT

The authors would like to express their deep sense of gratitude to everyone who has contributed, directly or indirectly, to the successful completion of this survey paper. We are especially indebted to our respected guide/supervisor for their continuous support, expert guidance, and constructive feedback throughout the development of this work. Their valuable insights and encouragement have played a crucial role in enhancing the quality and depth of this study.

We also extend our sincere thanks to our institution and the faculty members of the department for providing the necessary academic environment, infrastructure, and resources required to carry out this research. Their support has been instrumental in enabling us to explore and understand the advancements in deep learning and medical image analysis.

REFERENCES

- [1]. Global Cancer Observatory, "Global Cancer Statistics," 2025.
- [2]. A. Madabhushi and G. Lee, "Artificial intelligence in digital pathology," *Medical Image Analysis*, vol. 65, 2023.
- [3]. S. Bera et al., "Artificial intelligence in digital pathology — new tools for diagnosis," *Nature Reviews Clinical Oncology*, vol.20, 2023.
- [4]. L. Zhang et al., "Deep CNN-based histopathology image classification," *IEEE Access*, vol. 12, 2024.
- [5]. M. D. Nayeem et al., "Cross-platform multi-cancer histopathology classification using deep learning," *Scientific Reports*, 2025.
- [6]. H. Chen et al., "Patch-based deep learning for whole-slide image classification," *IEEE Transactions on Medical Imaging*, vol. 43, 2024.
- [7]. P. Sharma et al., "Context loss in patch-based histopathological image analysis," *Pattern Recognition Letters*, vol. 172, 2025.
- [8]. A. Dosovitskiy et al., "An image is worth 16x16 words: Vision Transformer," 2023 (medical adaptations).
- [9]. R. Mosalpuri et al., "DeepHistoViT: Vision Transformer for histopathological classification," *arXiv*, 2026.
- [10]. R. Singh et al., "Attention-based deep learning for cancer detection," *IEEE Access*, 2025.
- [11]. X. Wang et al., "Multi-scale context-aware networks for cancer classification," *IEEE Access*, 2025.
- [12]. A. Brahmareddy et al., "Graph-transformer based histopathology classification," *Pattern Recognition Letters*, 2025.
- [13]. D. L. Oliveira et al., "Hybrid CNN–Transformer models in histopathology image analysis," *IEEE Access*, 2025.
- [14]. X. Li et al., "Multi-feature fusion CNN–Transformer for cancer classification," 2026.
- [15]. Y. Liu et al., "Limitations of CNN-based histopathology classification," *Medical Image Analysis*, 2024.
- [16]. J. Wang et al., "Context-aware deep learning techniques for histopathology analysis," *Journal of Medical Systems*, 2026.
- [17]. M. Ilse et al., "Attention-based multiple instance learning for histopathology," updated applications, 2025.

- [18]. S. Kumar et al., "Deep learning for breast cancer histopathology classification," IEEE Access, 2025.
- [19]. L. Zhao et al., "Transformer-based medical image analysis: A review," 2024.
- [20]. G. Ramkumar, M. Alluwaici, T. Vijetha, G. H, A. Venkatraman and R. M, "A Reliable System Performance Through NeuToMa with Spiking Neural Network Transformation in Network on Chip (NoC) Communication," 2026 International Conference on Communication, Computing and Emerging Technologies (IC3ET), Vasai, India, 2026, pp. 215-219, doi: 10.1109/IC3ET64989.2026.11467379.
- [21]. V. R, P. R. Kumar, D. Das, G. H, W. Hanbin and E. A, "Machine Learning-Driven Predictive Maintenance for Wind Turbines," 2026 International Conference on Electric Power and Renewable Energy (EPREC), Durg, India, 2026, pp. 1-6, doi: 10.1109/EPREC66546.2026.11412064.
- [22]. A. Tiwari, J. D. Kumar, W. Syed and G. H, "Efficient FPGA-Based CNN Accelerator for Real-Time Traffic Sign Classification on PYNQ-Z2," 2026 IEEE International Conference on AI Engineering and Innovations (AIEI), NIT Jamshedpur, India, 2026, pp. 1-4, doi: 10.1109/AIEI69164.2026.11496958.
- [23]. G. Ramkumar, M. Alluwaici, T. Vijetha, G. H, A. Venkatraman and R. M, "A Reliable System Performance Through NeuToMa with Spiking Neural Network Transformation in Network on Chip (NoC) Communication," 2026 International Conference on Communication, Computing and Emerging Technologies (IC3ET), Vasai, India, 2026, pp. 215-219, doi: 10.1109/IC3ET64989.2026.11467379.
- [24]. Girish H., Srinivas S., Kavya M., ...Kabbur S. "Swift Autonomous Fire Suppression Robot", 16th International Conference on Advances in Computing Control and Telecommunication Technologies Act 2025
- [25]. Girish H., Soujanya S., "A Comprehensive Cadence-based Design Flow for RISC-V UART SOC IN SCL 180NM", 16th International Conference on Advances in Computing Control and Telecommunication Technologies Act 2025
- [26]. Girish H., Rakesh S., Prasanna Raj C.P., Singh A.S., Yathisha L, "Deep Learning for Diabetic Retinopathy: Automated Detection and Classification from Retinal Images", 16th International Conference on Advances in Computing Control and Telecommunication Technologies Act 2025
- [27]. Girish H., Shyam Sundar V., Priyanka C., Renuka Hebballi R.S, "RISC V Steel Processor Implementation for Communication between Two Chips using AXI Lite", 16th International Conference on Advances in Computing Control and Telecommunication Technologies Act 2025
- [28]. M. H N, G. H and N. K N, "Simulation and Verification of SPI Master and Slave Communication," 2025 6th International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 2025, pp. 1531-1534, doi: 10.1109/ICOSEC67334.2025.11459684. <https://ieeexplore.ieee.org/document/11459684>
- [29]. S. S. V, P. C and G. H, "Design and Analysis of a D-Band LC Bandpass Filter and Low-Noise Amplifier for 6G Front-End Applications," 2026 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), Gwalior, India, 2026, pp. 1-6, doi: 10.1109/IATMSI68868.2026.11466038. <https://ieeexplore.ieee.org/document/11466038>
- [30]. "A Comprehensive Cadence-based Design Flow for RISC-V UART SOC IN SCL 180NM" , Girish H, Soujanya, Grenze International Journal of Engineering and Technology, 2025

- [31]. "Swift Autonomous Fire Suppression Robot" Girish H et. Al., Grenze International Journal of Engineering and Technology, 2025
- [32]. "Deep Learning for Diabetic Retinopathy: Automated Detection and Classification from Retinal Images" , Girish H, Cyril Prasanna Raj P , Grenze International Journal of Engineering and Technology, 2025
- [33]. "RISC V Steel Processor Implementation for Communication between Two Chips using AXI Lite" Shyam Sundar V, Dr Girish H, Priyanka C, Renuka S Hebballi Grenze International Journal of Engineering and Technology, 2025
- [34]. Sujatha.S, Mahendra BM, & Dr. Girish H. (2024). Smart Irrigation System for Optimal Water Conservation. *Revista Electronica De Veterinaria*, 25(1), 2613 - 2619. <https://doi.org/10.69980/redvet.v25i1.1311>
- [35]. Mamatha K R, Thejaswini S, Rashmi N, & Dr. Girish H. (2024). Laser Fencing Surveillance System With Email Alert. *Educational Administration: Theory and Practice*, 30(2), 1539–1544. <https://doi.org/10.53555/kuey.v30i2.7919>
- [36]. Thejaswini S, Mamatha K R, Rashmi N, & Dr. Girish H. (2024). Performance Evaluation of AI Models in Early Heart Disease Diagnosis. *Educational Administration: Theory and Practice*, 30(2), 1531–1538. <https://doi.org/10.53555/kuey.v30i2.7918>
- [37]. Sushma K Sattigeri, Seema Srinivas, Shridhar Kabbur, Girish H, "Revolutionizing Urban Waste Management: An AI-Powered Autonomous Solution for Smart Cities", *Tuijin Jishu/Journal of Propulsion Technology* ISSN: 1001-4055, Vol. 45 No. 02 (2024)
- [38]. Girish H, Vajjala Sai Sree Hari,Ponde Lakshmi Narasimha,P N Shreya,Dadigala Bhargavi, "THE DESIGN AND VERIFICATION OF A 32-BIT RISC V PROCESSOR USING VEDIC MATHEMATICS", *International Journal of Creative Research Thoughts (IJCRT)*, ISSN:2320-2882, Volume.12, Issue 5, pp.1130-1135, May 2024, Available at :<http://www.ijcrt.org/papers/IJCRTAB02172.pdf>
- [39]. IoT based home automation using Blynk cloud, J. Chetan Naik, B. S. Deeksha, H. Girish, Gurusharan Kashetty, T. G. Shivapanchakshari, Manohar Sai Polakam, *Journal: AIP Conference Proceedings, AIP Conf. Proc. 2965, 020006 (2024)*, <https://doi.org/10.1063/5.0215427>, Published: July 2024
- [40]. IoT based fire alarm system using Blynk app, T. Bhargava, H. Kishor Kumar, Pramod Mahalingappa Godi, H. Girish, J. Chetan Naik, *Journal: AIP Conference Proceedings, AIP Conf. Proc. 2965, 020007 (2024)*, <https://doi.org/10.1063/5.0212325>, Published: July 2024