World Journal of Current Medical and Pharmaceutical Research
Content available at www.saap.org.in ISSN: 2582-0222

IMPROVING COLON HISTOPATHOLOGY CLASSIFICATION USING GRAPH-BASED STRUCTURAL FEATURES AND HOG DESCRIPTORS

M.SIVA PARVATHI*[1], SHAHEEN BEGUM[2], N.V.MUTTU LAKSHMI[3], P.V. ANAND KRISHNA[4]

[1*]*Dept. of Applied Mathematics, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India.*
[2]*Associate Professor, Institute of Pharmaceutical Technology, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India.*
[3]*Dept. of Computer Science, Sri Padmavati Mahila Visvavidyalayam, Tirupati, Andhra Pradesh, India.*
[4]*Government Polytechnic, Pilaripattu, Andhra Pradesh, India.*

DOI: https://doi.org/10.37022/wjcmpr.v8i3.427

ARTICLE HISTORY
Received on: 28-07-2026
Revised on: 14-08-2026
Accepted on: 21-09-2026

Keywords:
Histopathology, Graph Theory, HOG, Feature Fusion, Medical Image Analysis, Digital Pathology.

Corresponding Author
Dr. M.Siva Parvathi

ABSTRACT

Reliable classification of colon histopathology images plays an important role in the early detection and diagnosis of colorectal cancer. However, conventional microscopic examination is time-consuming and may be affected by inter-observer variability. This study proposes an interpretable hybrid computational framework that combines Histogram of Oriented Gradients (HOG) features with graph-theoretic descriptors derived from the spatial organization of cell nuclei. To capture complementary structural information, four graph-based representations: Delaunay triangulation, Minimum Spanning Tree (MST), k-Nearest Neighbor (k-NN), and Voronoi diagrams, are systematically examined. The extracted structural and texture features are integrated and used to train a Random Forest classifier. Experiments were performed on an augmented dataset containing approximately 5,000 images per class. The results indicate that the proposed feature-fusion approach provides improved classification accuracy, stability, and sensitivity for cancer detection compared with the use of individual feature descriptors. By combining image texture with interpretable spatial and structural information, the proposed framework offers a balanced approach in terms of predictive performance, interpretability, and computational efficiency, highlighting its potential for computer-assisted analysis of colon histopathology images.

INTRODUCTION

Colorectal cancer is one of the major causes of cancer-related morbidity and mortality worldwide. Early and accurate diagnosis plays an important role in improving patient outcomes and supporting appropriate treatment decisions. Histopathological examination of tissue specimens remains a fundamental method for the diagnosis and assessment of colorectal cancer. The conventional diagnostic process primarily depends on the visual examination of stained tissue sections by experienced pathologists. Although expert evaluation is essential, manual interpretation can be time-consuming and may be influenced by factors such as workload, fatigue, and differences in individual experience, leading to inter-observer variability.

The development of digital pathology and whole-slide imaging has enabled the large-scale digitization, storage, and analysis of histopathological specimens. The increasing availability of digital tissue images has created opportunities for developing computer-assisted diagnostic systems capable of extracting quantitative information from complex tissue structures. Such systems can support pathologists by providing objective and reproducible image-based measurements while reducing the time required for routine image analysis. Machine learning and deep learning techniques have been increasingly investigated for automated histopathology image classification. Deep learning models, particularly convolutional neural networks, have demonstrated considerable potential in learning discriminative features directly from medical images. However, their performance generally depends on the availability of large, well-annotated datasets, and the resulting models can be difficult to interpret. In medical applications, interpretability is particularly important because the features contributing to a classification decision should be understandable and, where possible, related to meaningful tissue characteristics. The morphological and spatial organization of cells provides important information about the underlying tissue architecture. During cancer development, alterations in tissue organization may be reflected in changes in nuclear morphology, spatial distribution, cellular density, and clustering patterns. These

structural characteristics provide a basis for representing histopathological images using mathematical and graph-based models. Graph theory offers a natural framework for describing relationships among individual cells or nuclei, where nuclei can be represented as vertices and their spatial relationships as edges. Different graph constructions can therefore capture complementary aspects of cellular organization and tissue architecture.

Texture information provides another important source of discriminative features in histopathology images. In particular, Histogram of Oriented Gradients (HOG) can describe local intensity and edge-orientation patterns that characterize the visual texture of tissue regions. However, texture descriptors alone may not adequately represent the spatial relationships among cells. Conversely, graph-based descriptors can capture structural organization but may not fully describe local image appearance. Integrating these complementary sources of information can therefore provide a more comprehensive representation of histopathological tissue patterns.

Motivated by these considerations, this study develops an interpretable hybrid framework for colon histopathology image classification by integrating texture-based and graph-theoretic structural features. HOG features are combined with descriptors derived from multiple graph representations, including Delaunay triangulation, Minimum Spanning Tree, k-Nearest Neighbor, and Voronoi diagrams. The fused feature representation is subsequently used with a Random Forest classifier to distinguish between histopathological classes. The proposed approach aims to exploit both visual texture and nuclear spatial organization while maintaining a balance between classification performance, interpretability, and computational efficiency.

CONTRIBUTIONS

- Development of a hybrid framework integrating HOG and graph-based structural descriptors.
- Systematic evaluation of multiple graph constructions.
- Integration of interpretable spatial and texture features.
- Comprehensive experimental validation.
- Detailed analysis of limitations and future directions

RELATED WORK

Over the past decade, computational histopathology has progressed from conventional handcrafted image descriptors to machine learning and deep learning-based approaches. Early studies mainly employed texture, statistical, and morphological descriptors such as Gray-Level Co-occurrence Matrix (GLCM), run-length statistics, fractal measures, Local Binary Patterns (LBP), Gabor filters, and wavelet transforms to characterize tissue morphology and heterogeneity. Histogram of Oriented Gradients (HOG) has also been used to capture local edge orientations and structural patterns. Such handcrafted descriptors remain relevant because they provide compact and relatively interpretable representations of histological appearance. Recent work has continued to investigate combinations of handcrafted features, including HOG and other spatial and frequency-domain descriptors, for colorectal tissue classification. The increasing availability of whole-slide imaging has subsequently accelerated the adoption of machine learning and deep learning for computational pathology. Convolutional neural networks (CNNs), vision transformers, weakly supervised learning, and multiple-instance learning have been applied to tissue classification, tumour detection, segmentation, prognosis prediction, and molecular characterization. Unger and Kather (2024) reviewed the application of deep learning to cancer genomics and histopathology and highlighted its potential for extracting clinically relevant information from histological and genomic data. Similarly, Bahadir et al. (2024) provided a broad review of artificial intelligence applications in histopathology, discussing CNNs, vision transformers, graph neural networks, weak and self-supervised learning, and the challenges associated with annotation, computational requirements, interpretability, and cross-site variability.

Although deep learning models have achieved considerable success, conventional CNN architectures may not explicitly capture long-range spatial relationships among cells and tissue components. This limitation has motivated increasing interest in graph-based representations, in which cells, nuclei, tissue regions, or image patches can be represented as nodes and their spatial relationships as edges. Such representations provide a natural framework for describing cellular organization and tissue topology. Meng and Zou (2023) reviewed clinical applications of Graph Neural Networks (GNNs) in computational histopathology and reported their use in classification, segmentation, and prognosis-related tasks. More recently, Brussee et al. (2025) reviewed emerging trends in GNN-based histopathology and emphasized that graph representations can explicitly model pairwise interactions and topological relationships within tissue structures, addressing some of the spatial limitations of conventional CNN-based approaches.

Graph-based modelling has also been investigated specifically in colon cancer histopathology. A spatially aware GNN framework was developed to transform whole-slide images into graph-structured representations for modelling spatial heterogeneity in colon cancer and predicting molecular profiles across multiple cohorts. The study demonstrated the potential of spatial graph representations for capturing tissue-level organization beyond conventional image features. Other recent work has investigated graph-based representations of nuclei, where appearance and geometric relationships between nuclei are used to identify meaningful cellular communities, further demonstrating the relevance of explicit nuclear spatial modelling in histopathological analysis.

The importance of deep learning and computational image analysis has also been demonstrated across other cancer types. Jiang et al. (2024) reviewed deep learning applications in breast cancer histopathology, including diagnosis, classification, prognosis, and treatment-related applications, and discussed continuing challenges associated with data quality, model interpretation, and clinical translation. These developments, together with the broader review by Unger and Kather (2024), indicate a continuing transition toward computational pathology methods that combine predictive

performance with clinically meaningful information.

Despite these advances, there remains scope for computationally efficient and interpretable approaches that explicitly combine texture information with nuclear spatial organization**.** Deep learning and GNN approaches can provide powerful representations but may require substantial computational resources and can be difficult to interpret at the individual feature level. In contrast, handcrafted descriptors can provide transparent measurements but may not adequately represent complex cellular relationships. Therefore, a hybrid representation that combines both types of information is of particular interest. Motivated by this gap, the present study integrates Histogram of Oriented Gradients (HOG) with graph-theoretic descriptors derived from Delaunay triangulation, Minimum Spanning Tree (MST), k-Nearest Neighbor (k-NN), and Voronoi diagrams**.** The resulting feature representation is classified using a Random Forest model to capture complementary information from local texture and nuclear spatial organization while maintaining computational efficiency and interpretability.

4. DATA DESCRIPTION

The study utilized a publicly available Lung and Colon Cancer Histopathological Images dataset available from **Kaggle.** In the present work, only the colon tissue subset was considered for experimentation. The dataset contains two categories: normal colon tissue images and colon adenocarcinoma tissue images, organized into separate folders as colon_n and colon_aca, respectively. Initially, the dataset consisted of 250 benign colon tissue images and 250 colon adenocarcinoma images. To improve dataset diversity and support robust model training, image augmentation was performed using the Augmentor package, increasing the dataset size to 5,000 images per class. All histopathological images are provided in JPEG format with a spatial resolution of 768×768 pixels.

In Data Preprocessing, the histopathological images were initially loaded and standardized to ensure uniformity during feature extraction and classification. Wherever required, images were resized to maintain consistent input dimensions across the dataset. The class labels were encoded numerically, where normal colon tissue images were assigned the label 000 and adenocarcinoma tissue images were assigned the label 111. To reduce computational overhead during preliminary experimentation, a smaller subset containing approximately 1,000 images from each class was first used for analysis and model validation. After optimizing the workflow, the experiments were extended to the complete augmented dataset consisting of 5,000 images per class. Stratified sampling with a 75:25 ratio is used. Five-fold cross-validation is performed. Ground-truth annotations are provided by expert pathologists.

5. METHODOLOGY

The processing pipeline consists of preprocessing, segmentation, feature extraction, and classification. Preprocessing includes color normalization, Gaussian filtering, and morphological enhancement. Nuclei are segmented using adaptive thresholding and watershed transformation. Connected component analysis extracts centroid coordinates. HOG descriptors encode distributions of oriented intensity gradients. Spatial graphs are constructed using Delaunay, MST, k-NN, and Voronoi methods. Structural features include degree statistics, clustering coefficients, edge length moments, density mea sures, and spectral eigenvalues.

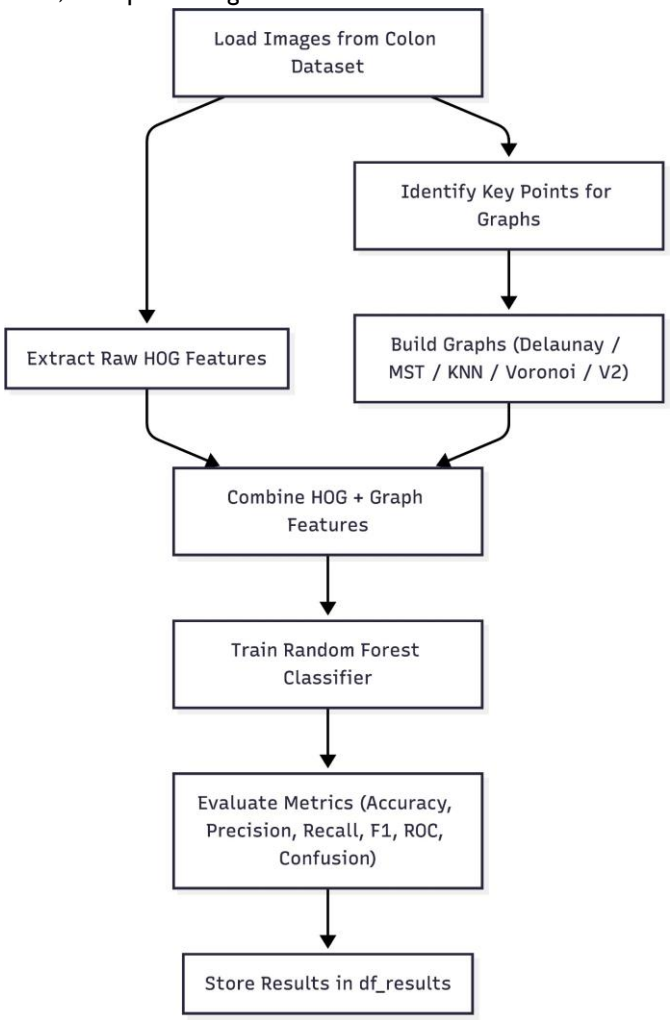

Figure 01: Methodology pipeline for Colon Tissue Classification

5.1 Histogram of Oriented Gradients (HOG)

HOG is a classical feature descriptor used in image analysis. It captures the distribution of gradientorientationsinlocalizedportionsofanimage,effectively encodingedgeandtexture information. HOG features are widely used in object detection and medical imaging tasks due to their robustness to lighting and minor deformations.

5.2 Graph-TheoreticFeatures

Graph-theoretic models are effective for representing the spatial organization and connectivity patterns present in histopathological images. In these approaches, image components such as nuclei, cells, or key feature points are treated as nodes, while the spatial relationships between them are represented as edges. This representation helps in capturing tissue architecture, cellular arrangement, and local structural irregularities that are often difficult to describe using conventional texture features alone.

5.2.1 Delaunay Triangulation

Delaunay Triangulation constructs triangles between neighboring points in such a way that no point lies inside

the circumcircle of any triangle. This method preserves local geometric relationships and is useful for analyzing cellular distribution, neighborhood density, and tissue organization patterns.

5.2.2 Minimum Spanning Tree (MST)

The Minimum Spanning Tree connects all nodes in the graph using the minimum possible total edge length without forming cycles. MST-based features describe the overall connectivity structure of the tissue and help in identifying clustering behavior, structural compactness, and spatial continuity among cells.

5.2.3 k-Nearest Neighbors (KNN) Graph

In the KNN graph, each node is connected to its kkk nearest neighboring nodes based on Euclidean distance. This approach focuses on local neighborhood interactions and is particularly useful for capturing short-range spatial dependencies and micro-architectural variations in tissue regions.

5.2.4 Voronoi Diagram

The Voronoi Diagram partitions the image plane into regions around seed points such that each region contains all locations closest to its corresponding point. These regions provide information about spatial influence, territorial distribution, and variations in cellular dominance across the tissue structure.

5.2.5 Enhanced Graph Variants (MST_V2 and KNN_V2)

The enhanced variants, namely MST_V2 and KNN_V2, extend the basic graph models by incorporating additional statistical, topological, and structural descriptors derived from node and edge properties. These variants provide a more detailed characterization of tissue organization and improve the discriminative capability of the extracted graph-based features.

5.3 AI and Medical Imaging

The proposed work can be considered a significant contribution to artificial intelligence–based medical image analysis as it integrates machine learning techniques with advanced feature extraction methods for tissue classification. The study combines handcrafted image descriptors such as Histogram of Oriented Gradients (HOG) with graph-theoretic structural features to capture both texture and spatial organization within histopathological images. By incorporating graph-based representations, the work explores the relationship between tissue architecture and classification performance, providing a more comprehensive understanding of structural patterns.

5.4 Combined Feature Representation

For each histopathological image, the extracted HOG descriptors and graph-theoretic features were integrated to form a unified feature representation. The HOG features captured local texture patterns, edge orientations, and intensity variations present in the tissue structure, while the graph-based features represented the spatial organization and connectivity relationships among cellular components. The combined feature vector was constructed by concatenating both feature sets into a single high-dimensional representation:

This integration allowed the model to utilize both texture-based and structural information simultaneously. During experimentation, graph-only representations were found to capture tissue organization effectively but showed instability in certain classification cases due to structural variability across samples. By incorporating HOG descriptors along with graph features, the representation became more robust and discriminative. The combined framework improved class separability between normal and adenocarcinoma tissues and produced more stable classification performance across different experimental settings.

5.5 Classification Model

The classification stage was carried out using the Random Forest algorithm implemented through the scikit-learn library. Random Forest was selected because of its robustness, ability to handle high-dimensional feature spaces, and effectiveness in reducing overfitting through ensemble learning. The classifier operates by constructing multiple decision trees during training and generating predictions based on majority voting.

Before classification, feature normalization was performed using the StandardScaler technique to standardize the feature distributions and ensure that all extracted features contributed uniformly during model training. The dataset was divided using a stratified train-test split in a ratio of 75:25, preserving the proportion of normal and cancerous tissue samples in both training and testing subsets.

The performance of the proposed framework was evaluated using several quantitative measures to provide a comprehensive assessment of classification quality. Accuracy metrics were computed for overall classification performance as well as separately for normal and cancer classes. Precision, recall, and F1-score were calculated to analyze the balance between false positives and false negatives. A confusion matrix was generated to visualize correct and incorrect predictions for each class.

In addition, Receiver Operating Characteristic (ROC) analysis and Area Under the Curve (AUC) values were used to evaluate the discriminative capability of the classifier across different decision thresholds. Feature importance analysis was also performed to identify the contribution of individual features toward classification performance. To assess computational efficiency, timing-related measures including total execution time, model fitting time, and average processing time per image were recorded during experimentation.

The effectiveness of the proposed framework is validated through quantitative evaluation using multiple performance measures including accuracy, precision, recall, F1-score, ROC analysis, confusion matrix, and computational efficiency. The observed improvement in classification performance demonstrates the capability of the proposed hybrid approach for reliable and efficient tissue image analysis.

6. RESULTS AND DISCUSSIONS

This section presents a comprehensive evaluation of the proposed hybrid framework using multiple quantitative and qualitative measures. The experimental results indicate that graph-theoretic features are capable of capturing meaningful structural and spatial information from histopathological images. Features derived from graph representations effectively describe tissue organization, cellular connectivity, and neighborhood relationships. However, when used

independently, graph-based features were not sufficient to consistently outperform the baseline HOG descriptors, which provide strong texture and edge-based representations.

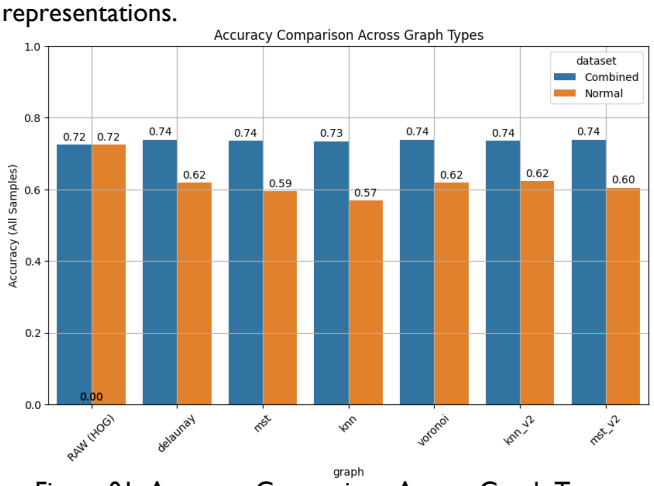

Figure 01: Accuracy Comparison Across Graph Types

This figure compares the classification accuracy obtained using different graph-based feature extraction methods. The results are presented separately for overall accuracy, cancer tissue accuracy, and normal tissue accuracy. The RAW HOG baseline achieved balanced performance, while the combined graph-based methods demonstrated noticeable improvements in several cases. Delaunay, Voronoi, and MST_V2 produced the highest overall accuracies among the combined models. The figure also shows that graph-enhanced representations improve the discrimination capability of the classifier, particularly for cancerous tissue samples. Although certain graph methods slightly reduced normal tissue accuracy, the overall trend indicates that combining structural graph features with HOG descriptors improves classification stability and consistency.

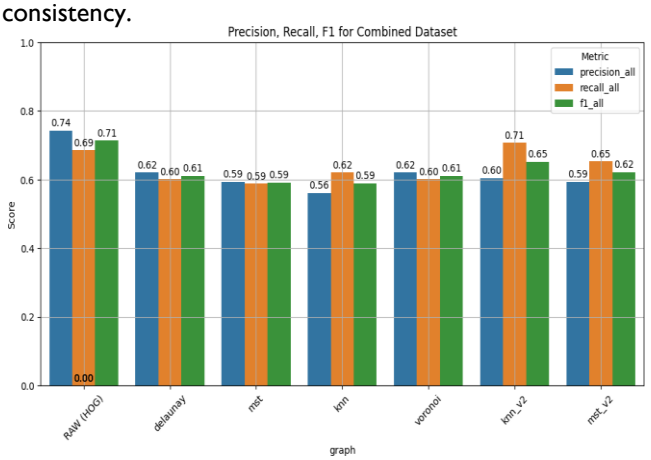

Figure 02: Precision, Recall, and F1-Score Comparison

The integration of HOG descriptors with graph-theoretic features produced more reliable and consistent classification performance. The combined feature framework achieved an overall improvement of nearly 1–2% in classification accuracy compared with the use of HOG features alone. In particular, the enhanced graph variants such as KNN_V2 and MST_V2 demonstrated improved sensitivity toward cancerous tissue detection, suggesting that enriched structural descriptors contribute valuable discriminative information for identifying adenocarcinoma patterns.

These findings highlight the importance of combining texture-based and structural representations in AI-driven medical image analysis. While HOG features effectively capture local intensity gradients and edge information, graph-based descriptors provide complementary insights into tissue architecture and spatial relationships among cellular components. The combination of both representations enables a more comprehensive characterization of histopathological patterns.

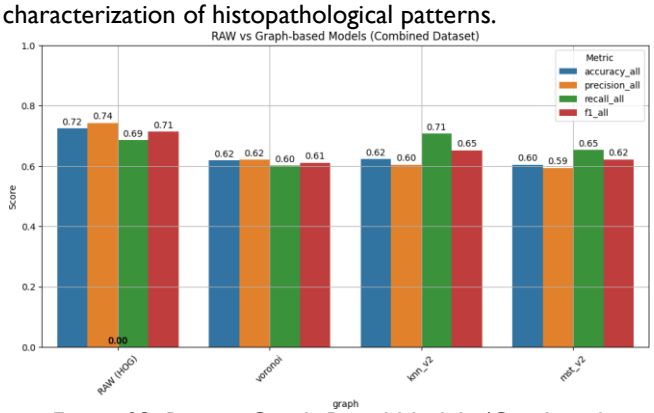

Figure 03: Raw vs Graph-Based Models (Combined Dataset)

This figure compares the classification performance of the baseline RAW HOG model with the proposed graph-enhanced models using the combined dataset. From the figure, it can be observed that the integration of graph-theoretic structural features with HOG descriptors consistently improves the overall classification accuracy compared to the baseline model. Among the evaluated approaches, the Delaunay, Voronoi, and MST_V2 based models achieved the highest overall performance, indicating that structural connectivity information contributes positively to tissue classification.

The figure also shows that the combined models provide better balance between cancer and normal tissue recognition. While the RAW HOG model already offers strong texture-based representation, the addition of graph features improves the classifier's ability to capture tissue organization and spatial relationships among cellular structures. Enhanced graph variants such as KNN_V2 and MST_V2 demonstrated improved cancer tissue detection, suggesting that additional graph statistics help in identifying structural irregularities associated with adenocarcinoma.

Another important observation from the figure is that the improvement in accuracy is achieved without significant variation across models, indicating stable and reliable classification behavior. Overall, the figure demonstrates that combining texture descriptors with graph-based structural representations produces a more discriminative feature framework for colon histopathology classification.

The heatmap visualizes classification accuracies for different graph methods across the combined, cancer, and normal tissue categories. Darker color regions correspond to higher accuracy values. The combined dataset models generally produced stronger performance than graph-only approaches, particularly for Delaunay, Voronoi, and MST_V2 methods. The heatmap clearly highlights the improved balance between cancer and normal tissue

recognition achieved by the hybrid feature framework.

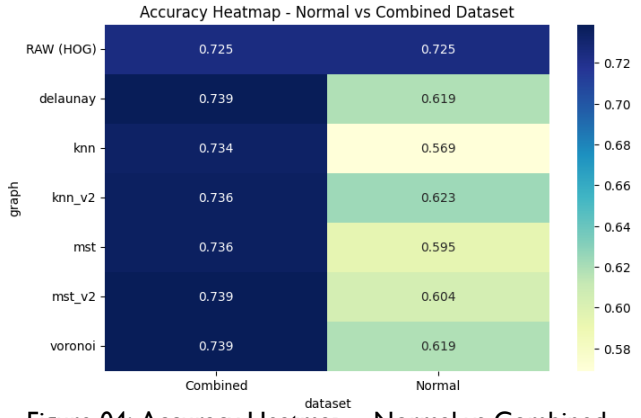

Figure 04: Accuracy Heatmap – Normal vs Combined Dataset

7. CONCLUSION

The experimental analysis demonstrates that graph-theoretic features provide meaningful structural information from histopathological tissue images. However, graph-based descriptors alone were not sufficient to outperform the RAW HOG baseline consistently. The integration of HOG descriptors with graph representations produced the most reliable results by combining texture information with spatial tissue organization.

Among the evaluated graph methods, Delaunay, Voronoi, and MST_V2 consistently achieved improved overall performance when combined with HOG features. The enhanced variants, particularly MST_V2 and KNN_V2, improved cancer tissue detection by incorporating additional graph statistics such as node degree distribution, clustering coefficients, graph density, and edge-length characteristics.

Another important observation is the computational efficiency of the proposed framework. Despite the inclusion of graph construction and structural feature extraction, the average processing time per image remained low. This indicates that the proposed methodology can potentially support real-time or computer-assisted clinical diagnostic systems.

Overall, the results confirm that combining texture-based descriptors with graph-theoretic structural information provides a more comprehensive representation of tissue architecture and improves colon histopathology classification performance.

ACKNOWLEDGEMENTS

We sincerely thank Pradhan Mantri Uchchatar Shiksha Abhiyan (PM-USHA) Project, under the Multi-Disciplinary Education and Research Universities (MERU) grant sanctioned to Sri Padmavati MahilaVisvavidyalayam, Tirupati, Andhra Pradesh.

REFERENCES

1. Brussee S, Buzzanca G, Schrader AMR, Kers J. Graph neural networks in histopathology: emerging trends and future directions. Med Image Anal. 2024;101:103444.
2. Bahadir CD, Omar M, Rosenthal J, Marchionni L, Liechty B, Pisapia DJ, et al. Artificial intelligence applications in histopathology. Nat Rev Electr Eng. 2024;1:93-108.
3. Dimitriou N, Arandjelović O, Caie PD, Tsironi E. Deep learning on histopathological images for colorectal cancer diagnosis: a systematic review. Diagnostics (Basel). 2022;12(4):837.
4. Jiang B, Bao L, He S, Chen X, Jin Z, Ye Y. Deep learning applications in breast cancer histopathological imaging: diagnosis, treatment, and prognosis. Breast Cancer Res. 2024;26:137.
5. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med. 2019;25:1054-1056.
6. Kather JN, Krisam J, Charoentong P, Weis CA, Althammer S, Luedde T, et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med. 2019;16:e1002730.
7. Meng Y, Zou L. Clinical applications of graph neural networks in computational histopathology: a review. Comput Biol Med. 2023;164:107201.
8. Shapcott M, Hewitt KJ, Rajpoot N. Deep learning with sampling in colon cancer histology. Front Bioeng Biotechnol. 2019;7:52.
9. Unger M, Kather JN. Deep learning in cancer genomics and histopathology. Genome Med. 2024;16:44.
10. Unger M, Kather JN. A systematic analysis of deep learning in genomics and histopathology for precision oncology. BMC Med Genomics. 2024;17:48.