Optimizing Patch-based Whole-Slide Image Classification through Advanced Selection Methods and Feature Extraction

Sümeyye Zülal Dik[1], Hatice Çağlar[1], Büşra Sinem Karadeniz[1], Musa Aydin[2], Zeki Kuş[3]
[1] Department of Computer Engineering, Fatih Sultan Mehmet Vakıf University, İstanbul, Türkiye
[2] Department of Artificial Intelligence and Data Engineering, Samsun University, Samsun, Türkiye
[3] Department of Artificial Intelligence and Data Engineering, Fatih Sultan Mehmet Vakıf University, İstanbul, Türkiye
Email: szdik@fsm.edu.tr, hatice.caglar@stu.fsm.edu.tr, busrasinem.karadeniz@stu.fsm.edu.tr,
musa.aydin@samsun.edu.tr, zkus@fsm.edu.tr

2026 11th International Conference on Machine Learning Technologies (ICMLT) | 979-8-3315-4865-0/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICMLT69916.2026.11689086

***Abstract*—Whole-slide histopathology images (WSIs) pose significant challenges for automated classification due to their large size and structural complexity. In this study, we propose an end-to-end system that divides WSIs into patches, extracts embeddings, selects informative patches, and classifies them using machine learning models. We evaluate Random Forest, XGBoost, and a Multi-Layer Perceptron (MLP) on the EBHI dataset, which contains 5532 images across five classes. XGBoost consistently yields the best performance, reaching an accuracy of 0.84 and demonstrating strong robustness to complex data structures. We systematically analyze the effects of patch selection methods (cosine similarity and k-means clustering), the number of patches, and data preprocessing. The results show that discarding non-informative patches and using pre-trained feature extractors such as KimiaNet improve efficiency and accuracy. Nonetheless, certain classes, particularly Normal and High-grade IN, remain difficult to classify, indicating a need for further refinement. Overall, this work highlights the importance of patch selection, data quality, and model choice in optimizing histopathology image classification. To support reproducibility, we publicly release the source code and EBHI train/test splits at GitHub.**

***Index Terms*—Whole-slide histopathology imaging, Patch-based image classification, Xgboost, Feature embedding and selection, Ebhi dataset**

I. Introduction

Deep learning has become a core technology for biomedical image analysis, often outperforming traditional feature engineering and, in specific tasks, human experts [1], [2]. Convolutional neural networks (CNNs) can automatically learn hierarchical representations from large, high-dimensional medical imaging datasets, enabling scalable and fast computer-aided diagnosis [3], [4]. Whole-slide histopathology images (WSIs) are giga-pixel scans of tissue slides that capture detailed morphology but pose substantial computational and methodological challenges. Diagnostic regions may occupy only small parts of the slide, and class imbalance is common because abnormal tissue can be rare [5], [6]. In recent years, deep neural networks have shown strong promise for automated WSI classification, typically using a patch-based paradigm in which WSIs are divided into smaller patches processed individually. Tellez et al. [7] combined CNNs with multiple instance learning to improve breast cancer metastasis detection under weak labels, and Wang et al. [8] proposed a weakly supervised framework for lung cancer WSIs with limited annotations. Poudel et al. [9] introduced a two-step patch-based approach for joint classification and segmentation to address gigapixel-scale images. Lu et al. [10] developed a multi-scale CNN that fuses information across magnifications, while Xiang et al. [11] designed a multi-scale attention network capturing both local and global features. More recently, Xu et al. [12] presented Prov-GigaPath, a foundation model trained on $\sim$ 1.3 billion pathology images, achieving state-of-the-art cancer classification on WSIs, and Fu et al. [13] proposed a self-learning sampling strategy that resolves non-differentiability in WSI down-sampling and enables end-to-end training. Collectively, these studies highlight the substantial impact of deep learning on WSI classification and provide robust, scalable solutions that improve accuracy and efficiency in computational pathology [14].

In this study, we investigate a patch-based classification pipeline for WSIs from the EBHI dataset, which contains 5532 images across five classes (Normal, Polyp, Low-grade IN, High-grade IN, Adenocarcinoma). Each WSI is divided into patches and represented using pre-trained KimiaNet embeddings. We systematically examine how (i) the number of patches per WSI, (ii) patch selection strategies (cosine similarity and k-means clustering), and (iii) pre-processing (discarding non-informative patches) affect downstream performance. Three classifiers—Random Forest, XGBoost, and a Multi-Layer Perceptron (MLP)—are evaluated using accuracy, precision, recall, and F1 score. XGBoost achieves the best performance, with accuracy up to 0.84 and strong robustness to complex histopathology patterns. Random Forest attains up to 0.74 accuracy, while the MLP underperforms overall (best accuracy 0.64) but benefits from larger numbers of patches, indicating sensitivity to data density. The effect of discarding non-informative patches is model-dependent: Random Forest often prefers using all patches, XGBoost shows mixed be-

Authorized licensed use limited to: FATIH SULTAN MEHMET UNIVERSITY. Downloaded on November 11,2026 at 07:58:40 UTC from IEEE Xplore. Restrictions apply.

havior, and the MLP exhibits both improvements and declines depending on the configuration. Cosine similarity and k-means clustering yield comparable overall results. Class-wise analysis reveals that all models perform well on some classes (e.g., Adenocarcinoma) but struggle with others, particularly Normal and High-grade IN, underscoring the need for more balanced data and targeted strategies for challenging categories. Overall, this work provides practical guidance for optimizing patch-based WSI classification and highlights the interplay between patch selection, preprocessing, and classifier choice. The main contributions of this study are:

- We propose and analyze a patch-based pipeline for WSI classification that leverages pre-trained KimiaNet embeddings to efficiently process large histopathology images.
- We systematically evaluate Random Forest, XGBoost, and MLP under different numbers of patches and with/without discarding non-informative patches, showing how model choice and preprocessing jointly affect performance and computational cost.
- We compare cosine similarity and k-means clustering for patch selection, quantifying their impact on classification performance and offering guidance on how many and which patches to retain.
- We provide detailed class-wise results, identify persistent difficulties for Normal and High-grade IN, and publicly release our train/validation/test splits and source code to facilitate reproducibility and future comparisons.

II. Material and Methods

A. Extreme Gradient Boosting Algorithm

Extreme Gradient Boosting (XGBoost) is a gradient-boosting ensemble of decision trees that iteratively refines predictions by minimizing a regularized objective function [15]. Trees are grown level-wise, selecting features and thresholds that best reduce the loss at each split. Second-order Taylor expansion and explicit regularization help control overfitting while maintaining high accuracy and efficiency. In this study, we use the XGBClassifier implementation with default parameters as a strong baseline for patch-level classification.

B. Random Forest Algorithm

Random Forest (RF) is an ensemble of decision trees trained on bootstrap samples of the data with random feature selection at each split. For classification, the final prediction is obtained by majority voting across trees. This strategy improves robustness, reduces variance, and handles high-dimensional, nonlinear problems effectively [16]. RF is widely used in medical image analysis as a competitive baseline due to its stable performance and interpretability. Here, it serves as a classical tree-based comparator to XGBoost.

C. Multi-Layer Perceptron (MLP)

The MLP is a feed-forward neural network used for supervised classification. It comprises an input layer (dimension determined by the 1024-dimensional KimiaNet embeddings), multiple fully connected hidden layers with nonlinear activations, and an output layer with one neuron per class. Key hyperparameters include hidden layer sizes, learning rate, and number of epochs [17]. In this work, we use an MLP with layers 1024–512–512–1024–512–5, ReLU activations in the hidden layers, and a softmax output for five-class prediction.

TABLE I
Dataset Scale of EBHI.

Magnification	40x	100x	200x	400x	Total
Normal	17	29	61	79	186
Polyp	119	165	254	304	842
Low-grade IN	204	341	603	660	1808
High-grade IN	47	80	130	161	418
Adenocarcinoma	205	471	790	812	2278
Total	592	1086	1838	2016	5532

D. Proposed Method

We propose a patch-based pipeline for WSI classification comprising five stages: patch generation, preprocessing, embedding extraction, patch selection, and classification (Figure 1). First, each WSI is divided into smaller patches to make processing feasible and to capture local tissue patterns at high resolution. Patches are then optionally filtered to remove tissue-free regions, producing two variants: a *discarded* (cleaned) set and a *not-discarded* set that retains all patches. Next, each patch is passed through the pre-trained KimiaNet network [18] to obtain a 1024-dimensional feature vector from the penultimate layer. For each WSI, this yields a set of patch embeddings of size $N \times 1024$, where N is the number of patches. Patch selection is then performed using either k-means clustering or cosine similarity. K-means groups embeddings into clusters from which representative patches are sampled, whereas cosine similarity directly ranks embeddings based on their similarity. The selected embeddings are forwarded to the classifiers (Random Forest, XGBoost, MLP) to predict one of five classes. WSI-level labels are obtained by majority voting over patch-level predictions (see Subsection 3.5), enabling efficient slide-level diagnosis from patch-level decisions.

III. Experimental Studies

A. Dataset

We use the EBHI dataset [19], which contains 5532 H&E-stained histopathological images of size 2048×1536 (JPG) across five classes (Normal, Polyp, Low-grade IN, High-grade IN, Adenocarcinoma) and four magnifications (40×, 100×, 200×, 400×) (Table I). Images are acquired at the Cancer Hospital of China Medical University and labeled according to predefined histological criteria.

For our experiments, we use only 200× magnification images. Per class, 80% of the images are used for training and 20% for testing, preserving class ratios. Exact file lists and splits are publicly available on GitHub to support reproducibility. Additionally, we define a validation set for the MLP by reserving the last five patches from the training data as validation and using the remaining patches for training.

Authorized licensed use limited to: FATIH SULTAN MEHMET UNIVERSITY. Downloaded on November 11,2026 at 07:58:40 UTC from IEEE Xplore. Restrictions apply.

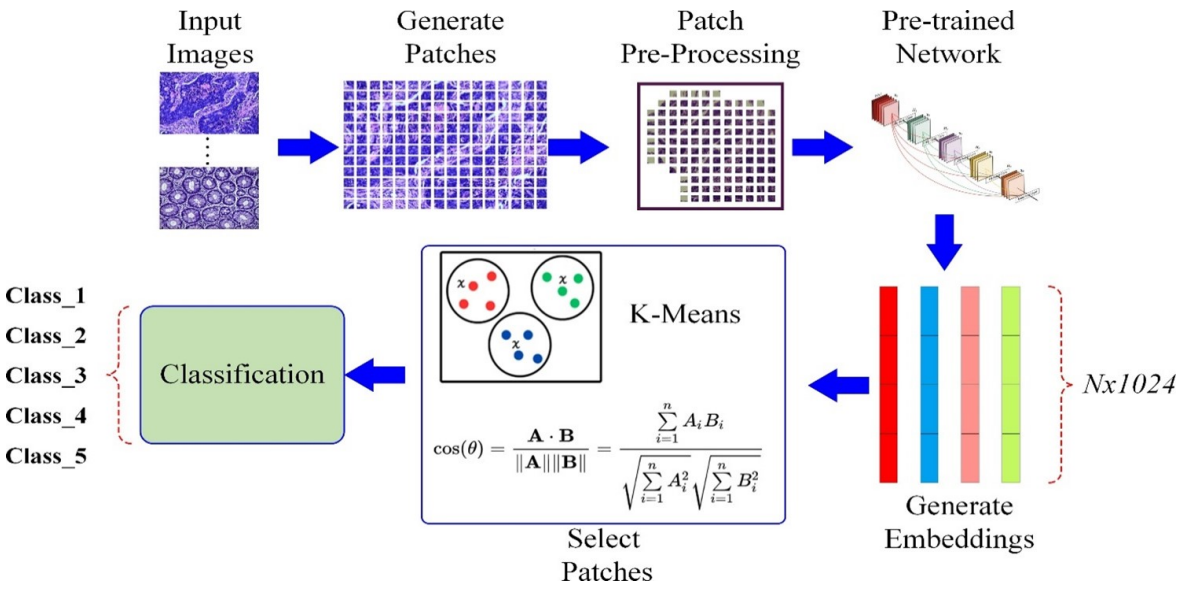


Fig. 1. Flow diagram of the study. The process begins with whole-slide input images, divided into smaller patches with 128×128 sized. Pre-processing is performed before they are fed into a pre-trained network, and two sets are prepared: discarded patches and not-discarded patches. The pre-trained network generates embeddings from these patches. The K-Means algorithm or cosine similarity is then used to select the most similar embeddings. These selected patches are classified into one of five classes using XGBoost, RandomForest, and MLP based on the generated embeddings.

The resulting numbers of training and test images per class are: Normal (48/13), Polyp (203/51), Low-grade IN (482/121), High-grade IN (104/26), and Adenocarcinoma (632/158).

B. Pre-processing

The EBHI images have a resolution of 2048×1536 pixels and cannot be fed directly into the networks due to memory constraints. Therefore, each 200× magnification image is divided into 128×128 patches (Figure 2), which are saved in a new folder corresponding to the original image.

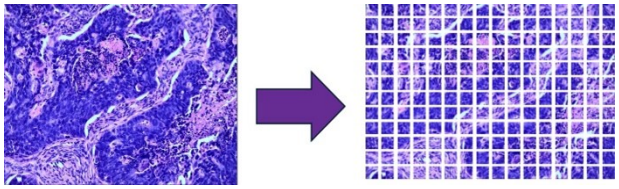


Fig. 2. Example of an image divided into patches. Each WSI is divided into smaller patches with 128×128 sized

We apply an additional cleaning step to remove tissue-free (non-informative) patches (Figure 3). Each patch is converted to grayscale, and its mean intensity, standard deviation, and Laplacian variance are computed to assess information content. Based on experimentally determined thresholds, patches with mean intensity $>$ 160, Laplacian variance $>$ 650, and standard deviation $<$ 41 are considered meaningless and discarded. All experiments are conducted on both the cleaned and raw (uncleaned) patch sets.

C. Extraction of the Embedding Vectors

Before applying patch selection, each patch is represented by a real-valued embedding vector extracted from a CNN.



Fig. 3. Discarding Non-informative Patches from the WSI. The image is converted to a grayscale, and then the average intensity, standard deviation, and Laplacian variance of each patch are calculated. The patches with high mean intensity, high Laplacian variance, and low standard deviation are considered meaningless and discarded.

We use KimiaNet [18], a CNN pre-trained on large-scale histopathology images, as our feature extractor. For each patch, we take the 1024-dimensional feature vector from the layer before the fully connected layer. All patch images are passed through KimiaNet, and their embeddings are saved as NumPy arrays in folders corresponding to each WSI. Consequently, for each WSI, we obtain 192 image patches and 192 associated embedding vectors.

D. Patch Selection

Patch selection is a crucial step in WSI analysis. After a WSI is divided into many patches, not all patches are equally informative for diagnosis, and evaluating every patch with deep networks is computationally expensive. By selecting only the most meaningful patches, we reduce computational complexity and focus the models on the most relevant image regions, which can improve classification performance and lead to more accurate and efficient diagnostic results.

In this study, we use two patch selection methods—cosine similarity and k-means clustering—to choose informative

Authorized licensed use limited to: FATIH SULTAN MEHMET UNIVERSITY. Downloaded on November 11,2026 at 07:58:40 UTC from IEEE Xplore. Restrictions apply.

patches for each image. For each WSI, cosine similarity is computed between all 192 patch embeddings to form a 192×192 similarity matrix. For each patch, the most similar patch is recorded, and the similarity values are sorted to select the top m patches (with $m \in$ 10,20,30,40,50). The selected patches and their embeddings are then copied into new folders for further analysis. A similar strategy is applied with k-means clustering on the patch embeddings, using k = 5 and k = 10. From each cluster, we randomly select 10, 20, 30, 40, or 50 patches in total (e.g., for k = 5, we pick 2, 4, 6, 8, or 10 patches per cluster). Because k-means can produce clusters of different sizes, clusters are first sorted by size; if a cluster has fewer patches than required, all of its patches are used and any remaining patches are taken from the largest cluster. The chosen patches and their embedding vectors are then stored in dedicated folders, ensuring a consistent and organized input for the subsequent classification experiments.

E. Training and evaluation

We use a parametric dataloader for training and testing that loads pre-computed embeddings according to the desired patch selection method (cosine similarity or k-means), number of patches per WSI (10–50), and cleaning condition (discarded vs. not-discarded). Each embedding is associated with its class label, and all embeddings from the corresponding WSI in the test set are loaded to enable slide-level majority voting. **Random Forest and XGBoost.** For RF and XGBoost, the training set consists of patch embeddings and their labels. Each model is trained on the same embeddings and evaluated on test-set embeddings, producing patch-level predictions. WSI-level labels are then obtained via majority voting over all patch predictions per slide. These slide-level predictions are compared with ground truth to compute confusion matrices and evaluation metrics (accuracy, precision, recall, F1). **MLP Model.** The MLP with architecture 1024–512–512–1024–512–5 uses ReLU activations in the hidden layers and softmax at the output. It is trained for 200 epochs using stochastic gradient descent (SGD) on the dataloader outputs, with a validation set used for model selection to improve generalization and mitigate overfitting. During testing, the MLP produces patch-level predictions that are aggregated by majority voting to obtain WSI-level labels, analogous to RF and XGBoost. All experiments are repeated for 10, 20, 30, 40, and 50 selected patches per WSI, for both cosine similarity- and k-means-based selection, and for both cleaned and raw patch sets. This design enables a systematic comparison of RF, XGBoost, and MLP under varying patch counts, selection strategies, and preprocessing conditions.

IV. Results and Discussion

We evaluate Random Forest, XGBoost, and MLP on the EBHI dataset, focusing on the impact of (i) the number of selected patches, (ii) patch selection strategy (cosine similarity vs. k-means), and (iii) discarding vs. keeping non-informative patches. All models operate on 1024-dimensional KimiaNet embeddings, which provide a compact, discriminative representation of each patch.

Table IV summarizes the quantitative results. Increasing the number of patches generally improves performance for all models and selection methods, reflecting the benefit of more comprehensive sampling of each WSI. XGBoost consistently achieves the highest performance, with a maximum accuracy of 0.84 and strong precision, recall, and F1 scores. Random Forest reaches up to 0.74 accuracy, while the MLP trails behind but improves as the number of patches increases, indicating that it is more data-hungry.

Effect of discarding patches. The impact of discarding non-informative patches differs by model and configuration. For Random Forest, not discarding patches often yields slightly better accuracy and F1, particularly at higher patch counts and with k-means selection. For XGBoost, the effect is mixed: discarded patches can yield marginally higher accuracy in some settings (e.g., larger patch counts), whereas keeping all patches sometimes improves precision, recall, and F1 (notably for 20–30 patches with cosine similarity). The MLP shows the largest variability: for 10 patches, keeping all patches improves all metrics across selection methods; at higher patch counts, the advantage alternates between discarded and not-discarded settings depending on the combination of patches and selector. Overall, the decision to discard patches should be adapted to the classifier and experimental setup rather than fixed a priori.

Cosine similarity vs. k-means. When using k-means selection with $k = 5$ or $k = 10$, Random Forest and XGBoost exhibit almost identical performance across all patch counts, indicating that these models are relatively insensitive to the choice of k and to modest variations in cluster structure. For MLP, k-means with $k = 5$ tends to yield slightly better accuracy in some settings (e.g., 10, 40, and 50 patches), but the gains are small and configuration-dependent. Comparing cosine similarity with k-means, we observe broadly similar performance trends across models and patch counts. Cosine similarity with not discarded patches often provides small but consistent gains for MLP and modest improvements for XGBoost at specific patch numbers, while Random Forest results are largely comparable across methods.

Class-wise performance. To better understand behavior in this five-class setting, we inspect confusion matrices for the best configurations of each model (Figure 4). The best Random Forest setup (40 cosine-selected patches, discarded data) achieves 0.74 accuracy and excellent performance on Adenocarcinoma (156 true positives, 2 false negatives). However, it fails to correctly classify any Normal or High-grade IN samples and shows only moderate performance on Polyp and Low-grade IN. Thus, overall accuracy masks substantial weakness on specific classes.

The best XGBoost configuration (40 cosine-selected patches, discarded data) attains 0.84 accuracy and improves recognition of Normal and Polyp compared to Random Forest, albeit with residual confusion between these classes and Adenocarcinoma. Low-grade IN is reasonably well identified (112 true positives, 24 misclassifications), while High-grade

Authorized licensed use limited to: FATIH SULTAN MEHMET UNIVERSITY. Downloaded on November 11,2026 at 07:58:40 UTC from IEEE Xplore. Restrictions apply.

TABLE II
COMPARISON OF DIFFERENT CLASSIFICATION METHODS. THE TABLE SHOWS THE EFFECT OF THE NUMBER OF PATCHES, PATCH SELECTION AND CLASSIFICATION METHODS, AND DISCARDED AND NOT DISCARDED PATCHES.

Models	Patches	Method	Discarded				Not Discarded			
			Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
Random Forest	10	Cosine Similarity	0.70	0.45	0.42	0.41	0.72	0.45	0.43	0.42
		K-means (k=5)	0.69	0.40	0.44	0.39	0.71	0.41	0.45	0.41
		K-means (k=10)	0.68	0.38	0.43	0.38	0.71	0.41	0.45	0.40
	20	Cosine Similarity	0.72	0.45	0.42	0.42	0.72	0.45	0.43	0.43
		K-means (k=5)	0.72	0.42	0.45	0.41	0.71	0.41	0.45	0.42
		K-means (k=10)	0.71	0.41	0.46	0.41	0.73	0.43	0.46	0.42
	30	Cosine Similarity	0.72	0.46	0.44	0.43	0.73	0.45	0.43	0.43
		K-means (k=5)	0.72	0.42	0.46	0.42	0.73	0.43	0.46	0.42
		K-means (k=10)	0.71	0.41	0.45	0.40	0.73	0.43	0.46	0.42
	40	Cosine Similarity	0.74	0.46	0.44	0.44	0.73	0.46	0.44	0.43
		K-means (k=5)	0.73	0.43	0.46	0.42	0.73	0.43	0.46	0.42
		K-means (k=10)	0.71	0.41	0.45	0.41	0.73	0.43	0.46	0.42
	50	Cosine Similarity	0.73	0.46	0.44	0.43	0.74	0.46	0.44	0.43
		K-means (k=5)	0.71	0.42	0.45	0.41	0.73	0.43	0.46	0.42
		K-means (k=10)	0.73	0.43	0.46	0.42	0.73	0.43	0.46	0.42
XGBoost	10	Cosine Similarity	0.79	0.48	0.49	0.48	0.78	0.47	0.49	0.48
		K-means (k=5)	0.79	0.51	0.67	0.51	0.78	0.50	0.67	0.50
		K-means (k=10)	0.78	0.50	0.67	0.50	0.78	0.50	0.67	0.50
	20	Cosine Similarity	0.80	0.48	0.51	0.49	0.81	0.68	0.53	0.52
		K-means (k=5)	0.80	0.53	0.69	0.54	0.80	0.53	0.68	0.54
		K-means (k=10)	0.80	0.53	0.68	0.54	0.80	0.53	0.68	0.54
	30	Cosine Similarity	0.82	0.50	0.52	0.50	0.81	0.69	0.52	0.52
		K-means (k=5)	0.82	0.57	0.70	0.59	0.82	0.57	0.69	0.59
		K-means (k=10)	0.82	0.57	0.69	0.59	0.82	0.57	0.69	0.59
	40	Cosine Similarity	0.84	0.70	0.59	0.60	0.82	0.69	0.56	0.57
		K-means (k=5)	0.82	0.57	0.69	0.59	0.82	0.57	0.69	0.59
		K-means (k=10)	0.82	0.56	0.69	0.59	0.82	0.57	0.69	0.59
	50	Cosine Similarity	0.83	0.70	0.58	0.59	0.83	0.70	0.58	0.60
		K-means (k=5)	0.82	0.57	0.70	0.59	0.82	0.57	0.69	0.59
		K-means (k=10)	0.82	0.57	0.69	0.59	0.82	0.57	0.69	0.59
MLP	10	Cosine Similarity	0.49	0.27	0.31	0.28	0.51	0.28	0.31	0.29
		K-means (k=5)	0.59	0.36	0.33	0.35	0.62	0.38	0.34	0.36
		K-means (k=10)	0.58	0.36	0.33	0.34	0.59	0.34	0.34	0.33
	20	Cosine Similarity	0.54	0.31	0.35	0.32	0.54	0.30	0.33	0.31
		K-means (k=5)	0.62	0.37	0.35	0.36	0.61	0.38	0.36	0.37
		K-means (k=10)	0.64	0.39	0.36	0.37	0.62	0.39	0.36	0.37
	30	Cosine Similarity	0.59	0.34	0.38	0.36	0.55	0.38	0.40	0.38
		K-means (k=5)	0.63	0.38	0.35	0.36	0.61	0.37	0.36	0.36
		K-means (k=10)	0.63	0.37	0.36	0.36	0.63	0.39	0.38	0.38
	40	Cosine Similarity	0.57	0.32	0.34	0.33	0.61	0.35	0.37	0.36
		K-means (k=5)	0.64	0.38	0.37	0.37	0.64	0.40	0.36	0.38
		K-means (k=10)	0.63	0.38	0.36	0.37	0.62	0.38	0.35	0.36
	50	Cosine Similarity	0.61	0.35	0.38	0.36	0.60	0.34	0.37	0.35
		K-means (k=5)	0.61	0.36	0.34	0.36	0.63	0.39	0.36	0.37
		K-means (k=10)	0.63	0.38	0.36	0.36	0.59	0.35	0.34	0.34

IN remains particularly challenging with no correctly predicted samples. Adenocarcinoma is classified most reliably, with 158 true positives and few confusions. These results confirm XGBoost as the strongest overall model, while highlighting persistent difficulty separating High-grade IN and, to a lesser extent, Normal from other classes.

The MLP configuration with 20 patches selected by k-means ($k = 10$) on discarded data reaches 0.64 accuracy. It performs well on Adenocarcinoma (132 true positives) and Low-grade IN (80 true positives) and moderately on Polyp (23 true positives), but completely fails on Normal and High-grade IN (zero true positives for both), frequently misclassifying them as other classes. This suggests that the MLP is biased toward more prevalent or visually distinctive classes and requires further regularization, architectural tuning, or class rebalancing to perform competitively in this multi-class setting.

Summary. Across all experiments, XGBoost emerges as the most effective classifier for patch-based WSI analysis on EBHI, combining high accuracy with relatively stable behavior across patch counts and selection methods. Random Forest provides a strong baseline but is more limited on challenging classes, while the MLP benefits from increased patch counts yet struggles with class imbalance and subtle distinctions. Both cosine similarity and k-means clustering are viable patch selection strategies, and their relative advantages are modest compared to the choice of classifier and the number of patches used. Class-wise results underline the importance of addressing difficult categories such as Normal and High-

Authorized licensed use limited to: FATIH SULTAN MEHMET UNIVERSITY. Downloaded on November 11,2026 at 07:58:40 UTC from IEEE Xplore. Restrictions apply.

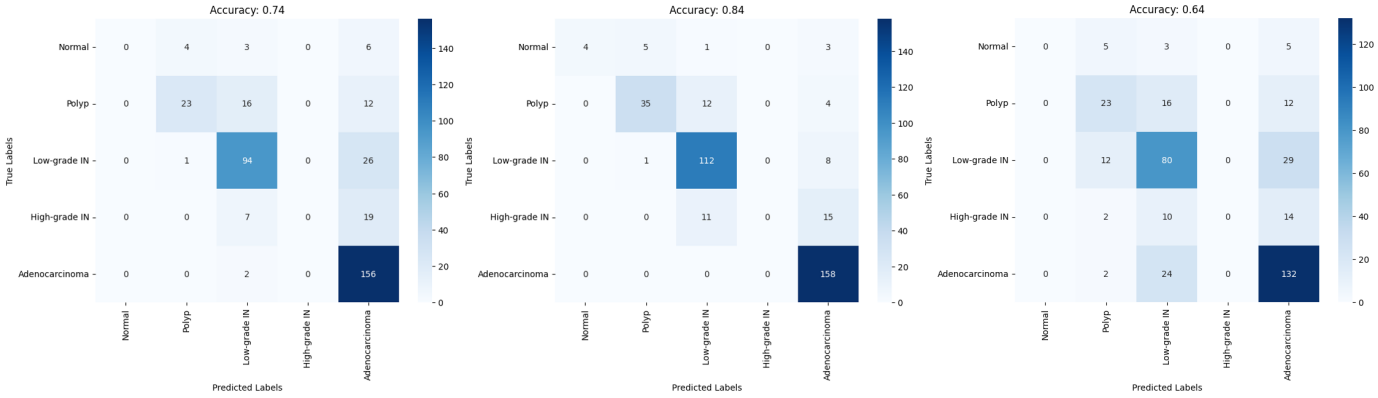


Fig. 4. Confusion Matrix Results of Random Forest (left),XGBoost (middle) and MLP (right) Trained with 40 Patches Selected using Cosine Similarity from Cleaned Data.

grade IN via more balanced data, refined sampling, or tailored loss functions. Overall, the findings underscore the key role of patch selection, preprocessing, and model choice in building reliable and efficient histopathology classification systems.

V. Conclusion

In conclusion, this study provides insights into optimizing a patch-based system for whole-slide histopathology image classification using the EBHI dataset of 5532 images across five classes. By addressing the computational challenges of large WSIs with a patch-based approach, we systematically evaluated the effects of patch selection methods, patch numbers, and data cleaning on the performance of Random Forest, XGBoost, and MLP models. XGBoost consistently achieved the best results, with a maximum accuracy of 0.84 and strong robustness to complex histopathological patterns. Both cosine similarity and k-means clustering are effective for patch selection, with model-dependent impacts. Increasing the number of patches generally improved performance, especially for XGBoost and Random Forest, indicating the benefit of more comprehensive sampling. The influence of discarding non-informative patches varied across models, underscoring the importance of preprocessing choices. Class-wise analysis showed that all models performed well on some classes (e.g., Adenocarcinoma) but struggled with others, particularly Normal and High-grade IN, highlighting the need for more balanced data and targeted strategies for difficult categories. Using a pre-trained network (KimiaNet) for feature embedding enabled efficient representation of complex images. Overall, the study emphasizes the critical roles of patch selection, data quality, and model choice in building high-performance histopathology classifiers and points to future work on advanced preprocessing, improved handling of challenging classes, validation on additional datasets, and ensemble-based approaches.

References

[1] E. Jangam, "Automatic detection of covid-19 from chest ct scan and chest x-rays images using deep learning, transfer learning and stacking," *Applied Intelligence*, vol. 52, no. 2, p. 2243–2259.

[2] J. S. Kim, "Deep learning-based diagnosis of alzheimer's disease using brain magnetic resonance images: An empirical study," *Scientific Reports*, vol. 12, no. 1.

[3] S.-C. Huang *et al.*, "Self-supervised learning for medical image classification: a systematic review and implementation guidelines," *Digital Medicine*, vol. 6, no. 1.

[4] C. McGenity, "Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy," *Digital Medicine*, vol. 7, no. 1, p. 114.

[5] X. Li *et al.*, "Multi-level feature fusion network for nuclei segmentation in digital histopathological images," *The Visual Computer*, Mar. 2022.

[6] J. M. Lee and J. S. Bae, "Enhancing diagnostic precision in liver lesion analysis using a deep learning-based system: opportunities and challenges," *Nature Reviews Clinical Oncology*, p. 1–2.

[7] D. Tellez *et al.*, "Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology," *Medical Image Analysis*, vol. 58, pp. 101 544,.

[8] X. Wang, "Weakly supervised deep learning for whole slide lung cancer image analysis," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, p. 3950–3962.

[9] S. Poudel and S.-W. Lee, *Whole Slide Image Classification and Segmentation Using Deep Learning*. ACM.

[10] M. Lu *et al.*, "Data-efficient and weakly supervised computational pathology on whole-slide images," *Nature Biomedical Engineering*, vol. 5, no. 6, pp. 555–570,.

[11] H. Xiang, "Multi-scale representation attention based deep multiple instance learning for gigapixel whole slide image analysis," *Medical Image Analysis*, vol. 89, p. 102890.

[12] H. Xu, "A whole-slide foundation model for digital pathology from real-world data," *Nature*, vol. 630, no. 8015, p. 181–188.

[13] Z. Fu, "Whole slide images classification model based on self-learning sampling," *Biomedical Signal Processing and Control*, vol. 90, p. 105826.

[14] C. L. Srinidhi, "Deep neural network models for computational histopathology: A survey," *Medical Image Analysis*, vol. 67, p. 101813.

[15] M. Abbasniya *et al.*, "Classification of breast tumors based on histopathology images using deep features and ensemble of gradient boosting methods," *Computers and Electrical Engineering*, vol. 103, pp. 108 382,.

[16] B. Loef *et al.*, "Using random forest to identify longitudinal predictors of health in a 30-year cohort study," *Scientific Reports*, vol. 12, pp. 10 372,.

[17] A. Rezaeipanah *et al.*, "Design of ensemble classifier model based on mlp neural network for breast cancer diagnosis," *Inteligencia Artificial*, vol. 24, no. 67, pp. 147–156,.

[18] A. Riasatian, "Fine-tuning and training of densenet for histopathology image representation using tcga diagnostic slides," *Medical Image Analysis*, vol. 70, p. 102032.

[19] W. Hu *et al.*, "Ebhi: A new enteroscope biopsy histopathological h&e image dataset for image classification evaluation," *Physica Medica*, vol. 107, pp. 102 534,.

Authorized licensed use limited to: FATIH SULTAN MEHMET UNIVERSITY. Downloaded on November 11,2026 at 07:58:40 UTC from IEEE Xplore. Restrictions apply.