Iranian Journal of Blood & Cancer

Iran J Blood Cancer, 2026, 18(2), 15-24
https://ijbc.ir

Original Article

# HDLF-CBHIR: A Hierarchical Deep Learning Framework for Content-Based Histopathological Images Retrieval

Amin Heyrani Khameneh[1], Morteza Valizadeh[1*], Mahdi Chehel Amirani[1], Seyed Sadra Kashef [1]

[1]Department of Electrical and Computer Engineering, Urmia University, Urmia, Iran.

Scan and read the article online

**Citation** Heyrani Khameneh A, Valizadeh M, Chehel Amirani M, kashef S. HDLF-CBHIR: A Hierarchical Deep Learning Framework for Content-Based Histopathological Images Retrieval. Iran J Blood Cancer. 2026 June 30;18(2): 15-24.

doi 0

**Article info:**
**Received:** 22 Apr 2026
**Accepted:** 08 Jun 2026
**Published:** 30 Jun 2026

**Keywords:**
Artificial intelligence (AI)
Deep learning
Content-based image retrieval (CBIR)
Digital pathology
Triplet network
Diagnosis
Cosine similarity

## Abstract

**Objective:** To propose a deep learning-based hierarchical framework for content-based histopathology image retrieval (CBHIR) that can serve as a diagnostic aid for pathologists.

**Methodology:** The framework extracts deep features using a pre-trained EfficientNet-B0 model, followed by global average pooling (GAP) for dimensionality reduction. A triplet network with a custom loss function is designed for deep hashing; the loss function leverages cosine similarity and incorporates data engineering by selecting the most challenging samples in the feature space to enhance retrieval performance. The model is evaluated on the Kimia path24c dataset and compared with baseline methods.

**Results:** The proposed framework achieved an accuracy of 98.76% for patch-level retrieval, 98.61% for whole-slide retrieval, and a total accuracy of 97.39% in publicly available Kimia path24c dataset.

**Conclusion:** The experimental results demonstrate the efficiency of the proposed framework for the CBHIR task. Moreover, the output of this work can be used in diagnostic practice to help pathologists analyze digital histopathology samples, as well as in education to support training of pathology professionals.

## 1. INTRODUCTION

Content-based image retrieval (CBIR) is the task of searching for images with similar content to that of a query image. A CBIR task is generally divided into two main parts: feature extraction and finding and defining appropriate criteria for image retrieval [1]. Traditional feature extraction methods use hand-crafted features like color, texture, and shape and combine them with key point-based image descriptions and representations (e.g., the Scale Invariant Feature Transform (SIFT) [2] combined with the Bag of Words (BoW) model [3]). In recent years, deep learning-based methods have shown significant performance in various machine vision tasks, including image retrieval. Feature extraction using pre-trained CNNs on large datasets such as ImageNet [4] that are fine-tuned on images of the target database can extract more meaningful and deeper features than conventional methods. This can directly lead to improved classification accuracy in the classification

* Corresponding Author:
Morteza Valizadeh
Affiliation: Department of Electrical and Computer Engineering, Urmia University, Urmia, West Azerbaijan Province.
E-mail: mo.valizadeh@urmia.ac.ir

problem as well as better image retrieval [5]. In the past, pathologists had to use guidance books as a reference to make a final diagnosis in suspicious cases, which was an inefficient and time-consuming process [6]. Recently, with the advancement of imaging techniques and the emergence of digital pathology, specimen slides are prepared as whole slide images (WSI) for pathological viewing, analysis, and clinical diagnosis [7]. Although these WSI images can be used to improve performance and speed up the diagnosis and treatment process, the lack of well-labeled databases by pathologists makes them a challenging task to use as a reference. Content-based histopathology image retrieval (CBHIR) is a subset of a CBIR system that is used to search and find similar images to a query image given to the system from a large reference database to assist pathologists in automatic diagnosis and analysis of histopathological images [8]. Several supervised and unsupervised methods have attempted to overcome the challenges associated with CBHIR [9]. In a prototype system created by Zheng et al. [10] the dot product of four different types of image features, color histogram, image texture, Fourier coefficients, and wavelet coefficients is used to determine image similarity. Banu et al. [11] used several color feature extraction and indexing strategies to assess the effectiveness of CBHIR. For feature extraction and indexing, the suggested descriptors are based on statistical modeling and wavelet decomposition. An appropriate embedding space and a similarity metric must be used in order to rank the retrieved images [12,13]. Historically, the fine texture and pattern variability seen in hematoxylin and eosin (H&E) images have been difficult for handmade features like SIFT to capture [2]. Convolutional neural networks (CNNs) are currently quite successful at learning meaningful representations that lead to successful H&E image retrieval. For processing the distinct features of histopathological images, the majority of current techniques rely on using pre-trained convolutional neural networks directly for retrieval tasks, which may not be the best option [14]. Since a high-dimensional feature space is utilized to compute similarity and retrieval over very large databases, these approaches can be computationally complex [15] or inefficient [16]. Additionally, the entire process can be very time-consuming [17]. Considering these, a variety of deep learning techniques have been suggested by researchers. A deep ranking hashing method for retrieving and classifying histopathology images based on pairwise comparisons was proposed by Shi et al. [18], aiming to concurrently learn the representations and binary codes within a single response model. To enhance deep metric learning for image retrieval, attention-based models have been proposed [19]. Triplet loss and histogram loss [20] are two examples of triple-based loss functions that are frequently used to learn regressions that preserve the relative order of similarities. In histopathological image analytics, Babaie et al. [21] presented a novel dataset for retrieving and classifying digital pathology scans, thus providing a useful resource for the wider research community. Kieffer et al. [22] explored the use of deep features in the Kimia Path24c dataset [23] by exploiting transfer learning with pre-trained architectures and comparing pre-trained networks against training from scratch. Shafiei et al. [23] used benchmark deep learning models, like VGG16, InceptionV3, and DenseNet-121, as feature extractors compared using the proposed Kimia Path24C dataset, which is a color version of the classical Kimia Path24 aimed at compensating for the lack of staining information in tissue pattern recognition. Another interesting approach [24] uses CNNs trained with multi-similarity loss and a mixed-attention mechanism (spatial and channel attention) to improve feature discrimination. The model is assessed on both a self-established histopathological dataset and Kimia Path24C [23]. Image descriptors are obtained by squeeze-and-excitation (SE) blocks [25] and fully connected layers. This is consistent with the increased emphasis on attention processes and embedding optimization in histopathological image retrieval, enhancing the utility of Kimia Path24C as a color-aware computational pathology benchmark. In this paper, we present a hierarchical framework for feature extraction, and a suitable hashing model that is specifically designed to handle the issues of CBHIR. Our technique combines powerful feature extraction capabilities with appropriate triplet network architecture and loss mechanisms.

The key contributions of this paper can be summarized as follows:

- The proposed CBHIR framework trains the feature extraction and deep hashing phases hierarchically, hence decreasing the model complexity and need for large training datasets.
- We proposed a new triplet network to process the extracted features by pre-trained model and generate more discriminative features.
- We devised a custom triplet loss function with the help of cosine similarity, which is a more versatile similarity metric and thus provides a controllable margin than that of Euclidean distance.
- To empower the characterizing capabilities of the extracted features, we utilize the batch-hard approach to the training of our proposed retrieval model and select the input samples sparsely.

## 2. METHODOLOGY

### 2.1. Dataset and Evaluation

The dataset used in this paper is the publicly available Kimia Path24C dataset [23], which is designed for image classification and retrieval in digital pathology. Kimia Path24C consists of 24 WSIs representing different staining methods and tissue textures, among which 1325 test patches of size 1000 × 1000 were manually selected and extracted from the original WSIs as a fixed test set. Then, the training set can be constructed by sampling the rest of the WSIs. The training set used in this paper is pre-built and provided by KIMIA Lab with 22591 patches. Examples of these images randomly selected from different patches are shown in **Fig. 1.**

For each test image $x_j$, where j = 1, 2, ..., $n_{tot}$, the K test images that are visually most similar to the query image are extracted from the remaining test images. As stated in (1), retrieval is considered successful if at least one of the K retrieved samples has the same class label as the query image $x_j$. Then, the correct retrieval for one query image $x_j$ is assigned. The overall retrieval score is calculated as the average of the recall scores for all test data, as shown in (2).

$$(1)\ \text{Recall@}K_j = \begin{cases} 1, & if\,\exists k: y_k = y_j, \text{for } k = 1,2,\dots,K \\ 0, & otherwise \end{cases}$$

$$(2)\ Recall@K = \frac{1}{n_{\text{tot}}}\sum_{j=1}^{n_{\text{tot}}} Recall@K_j$$

The performance of image retrieval is tested for the Kimia Path24c dataset using the patch-level and whole-slide-level accuracy metrics introduced in [21]. This dataset includes test images from 24 different classes. Let $n_{\_tot}$ be the total number of test images and $n_{\_\Gamma s}$ be the number of test images from the *s-th* class, where s = 1, 2, ..., 24. We denote the total number of retrieval images in the CBHIR task as R. The patch-level accuracy $\eta_p$ can be defined as:

$$(3)\ \ \eta_p = \frac{1}{n_{\text{tot}}}\left(\sum_{s=1}^{24} |\mathcal{R} \cap \Gamma_s|\right)$$

The whole-slide-level accuracy $\eta_w$ can be defined as:

$$(4)\ \ \eta_w = \frac{1}{24}\sum_{s=1}^{24} \frac{|\mathcal{R} \cap \Gamma_s|}{n_{\Gamma_s}}$$

The overall accuracy $\eta_{\text{total}}$ is then calculated as the product of the patch-level and whole-slide-level accuracies:

$$(5)\ \ \eta_{\text{total}} = \eta_p \times \eta_w$$

It is important to note that the patch-level accuracy $\eta_p$ is equivalent to the formulation of Recall@1 in the previous explanation.

### 2.2. Preprocessing and Feature Extraction

Before feature extraction of the histopathological image dataset, the preprocessing required for image processing was performed. To make the raw data compatible with the EfficientNet-B0, the images had to be resized. In the case of Kimia Path24c, the images were resized and center cropped to 224 × 224 size.

The authors of this paper implement a pre-trained EfficientNet-B0 model, which is essentially a combination of convolutional, residual, and fully connected layers. The last fully connected layer that provides the 1000th-dimensional output, corresponding to the number of object classes in the ImageNet dataset, is removed. The remaining layers are frozen as feature extractors that cannot be modified. In **Fig. 2(a)**, the authors show how these layers are connected to a classification layer. The new layer is used to classify the Kimia Path24c dataset, which contains 24 different classes. For example, rotation as a data augmentation method, which aims to reduce the sensitivity of the model to image variations, is used in this phase of processing. To achieve high-level features, considering the complexity of histopathological images, the last convolutional layer was fine-tuned. These layers are being trained in a way that contributes to the discrimination power of the challenging images in the hidden layer for better retrieval. Finally, the features from the model that had the best classification accuracy on the Kimia Path24 dataset were extracted and stored for use in the next phase of the system.

### 2.3. The Proposed Deep Hashing Model

In this paper, a new deep hashing model is presented that combines feature engineering techniques with a specially designed triplet network architecture, as illustrated in **Fig. 2**. The suggested method uses a triplet network architecture to learn image retrieval semantic-preserving embeddings. This triplet network consists of sequential layers that share

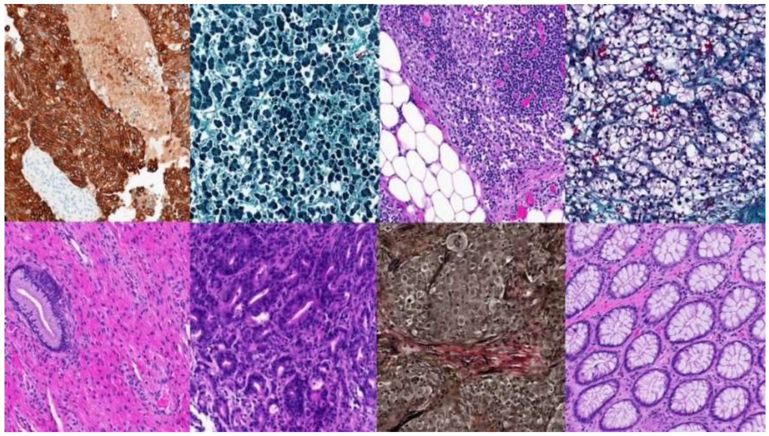

**Figure 1.** Training patches from every Kimia Path24c WSI.

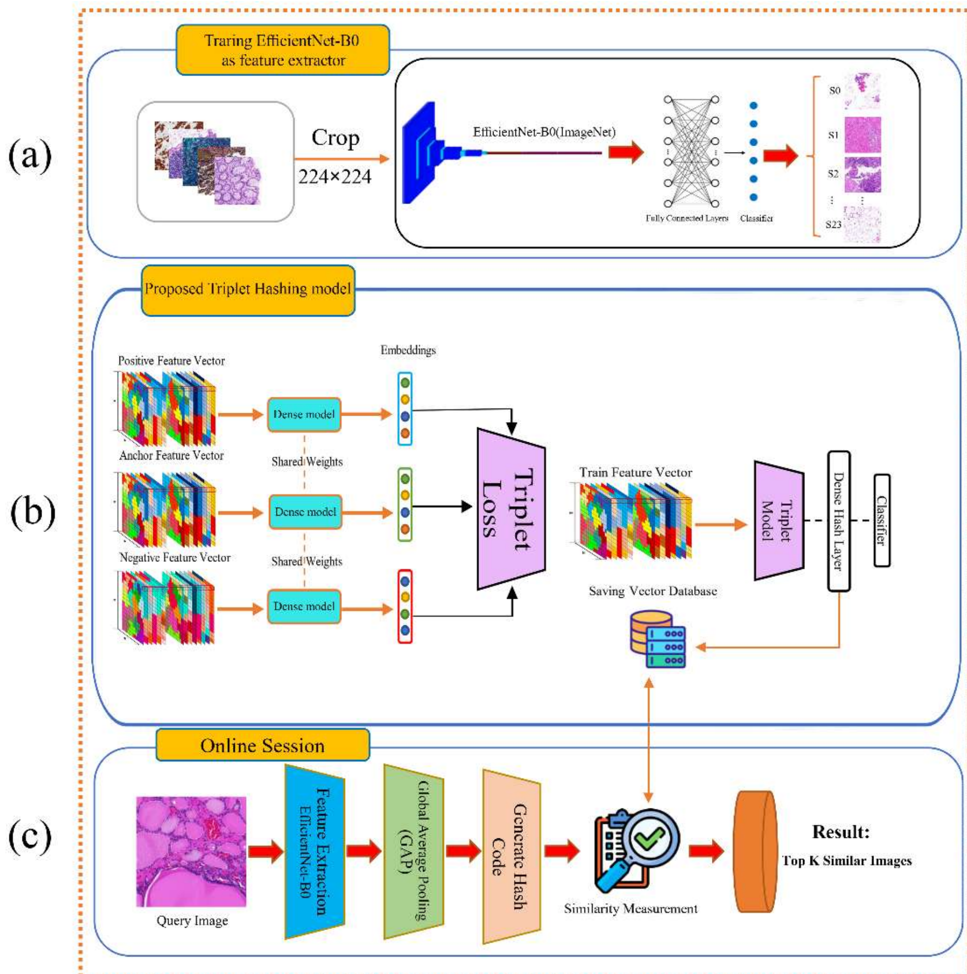


**Figure 2.** Overview of the training and querying of our hierarchical system. (a) Extracting histopathological image features using pre-trained EfficientNet-B0 model. (b) Illustration of the proposed hashing model based on a triplet network architecture. (c) Online session f the proposed retrieval system.

weights across the three input data streams, as shown in **Fig. 2(b)**. The components of this data consist of an anchor (the target sample), a positive sample (a feature vector that shares the anchor's class label), and a negative sample (a feature vector that has a different class label). During training, the network is guided by a neural attention mechanism that focuses on the hardest triplet samples introduced in [26], as identified by changing the similarity measures to cosine similarity in the training dataset. Essentially, the hardest negative sample is the most similar example to the anchor but from a different class, while the hardest positive sample is the least similar example to the anchor yet from the same class. This forced the network to learn more discriminative embeddings by prioritizing hard triplets. It allowed finer distinctions between similar images of different classes while clustering images from the same class despite visual differences. After training the triplet network as the base model, a new model consisting of a hash code layer and a classification layer is built, as shown in **Fig. 2(b)**. By keeping the weights of the base network frozen, the extended network can generate compact and meaningful hash codes that can be used for efficient image retrieval. Such hash codes may preserve the underlying meaning of images so that visually and semantically similar images can be retrieved quickly and accurately. Meanwhile, the classification layer provides additional supervisory signals to enhance the quality of the learned representation further. As mentioned above, the proposed model uses a custom batch-hard triplet loss function that is based on cosine similarity. This loss aims to maximize the cosine similarity between anchor and positive samples while minimizing the cosine similarity between anchor and negative samples. The custom batch-hard loss function can be formulated as follows:

$$(6)\quad L = \max\{0, \alpha - \min_{p \in P} \cos(a, p) + \max_{n \in N} \cos(a, n)\}$$

where a represents the embedding of the anchor sample, p is the set of positive samples, n is the set of negative samples, and $\alpha$ is the margin parameter. The function $\cos(.,.)$ denotes the cosine similarity computed as:

$$(7)\quad \cos(x, y) = \frac{x^{\top}\boldsymbol{y}}{|x||y|}$$

In the case of Euclidean distance, the optimal margin distance between anchor-positive and anchor-negative is less meaningful; hence, cosine similarity should be used, as it provides a normalized and interpretable measure of similarity. Setting the right margin becomes less important as the cosine similarity itself becomes a measure of relative angles between the embeddings.

### 2.4. Search and Retrieval

In the domain of CBIR, the whole search process can generally be achieved in three major steps: calculation of similarity, ranking and retrieval, and finally, visualization of retrieved images [27]. For the calculation of similarity, a wide variety of metrics, such as Euclidean distance, cosine similarity, and other metrics, can be used to quantify how much the query image is similar to every image available in the database. Images will then be ranked based on their similarity to the query image in the hierarchy; the highest-ranked images are retrieved for the user's viewing. In our study, we directly investigated the efficiency of cosine similarity as a robust measure. It considers the magnitude and direction of vectors; hence, it can capture subtle semantic similarities between images. With cosine similarity in our experiments, we could achieve improved precision in retrieving medically relevant images.

### 2.5. Implementation Details

As mentioned before, we used a pre-trained EfficientNet-B0 model as a feature extractor. The model was trained using Keras, a deep learning library, to learn the features of histopathology images by fine-tuning the last layers of the pre-trained model. To reduce the dimensionality of the extracted features, global average pooling (GAP) was performed after the last convolutional layer. Thus, each image is mapped to a feature vector of dimension 1280.

We implemented a baseline model based on a traditional autoencoder structure [28] to compare the performance of the proposed triplet network-based approach. Our baseline model uses a dense autoencoder structure, where two dense layers containing 128 and 64 neurons make up the encoder portion of this model. Then, we used a batch normalization layer and a dropout layer to prevent overfitting. Moreover, in this network, a classification layer is added to the encoder section so that the model can be optimized for reconstruction loss and classification loss simultaneously. Since MAE is more robust to outliers compared with MSE, the reconstruction loss is defined by MAE instead of the commonly used MSE. Moreover, the model makes use of the Adamax optimizer, which has better stability and convergence properties, especially for low-gradient problems. The hash codes of the proposed autoencoder-based baseline model are generated directly from the dense layers present in the encoder part of the network. More precisely, the compact hash codes are obtained from the activations of the last dense layer in the encoder, containing 64 neurons The proposed deep hashing model uses a batch size of 200 with dropout regularization. The triple loss is equipped with a margin parameter of 0.15, which limits the distance between divergent embeddings. Batch-hard mining selects the most challenging negative and positive examples in each batch, thus tuning the training process for stronger and more distinctive embeddings. The Adadelta optimizer is chosen because of its ability to efficiently handle sparse gradients and the resulting fast convergence speed. After several experiments, the learning rate was set to 1.0. To investigate the correlation between the hash code length and the retrieval accuracy, the researchers in this paper experimented with different lengths, and the results were optimal for a length of 64. The activation function "tanh" was used in the hash code layer because it can encode values in the range [-1, 1] and preserve the order of the distance between hash codes.

All experiments were conducted using a single workstation with an NVIDIA® GeForce RTX 4090 GPU, a 13th-generation Intel® Core™ i7-13700K CPU, and 128 GB of RAM.

### 2.6. Highlights

- A hierarchical content-based histopathological image retrieval framework is proposed based on deep learning.
- Deep features are extracted from input images using transfer learning techniques.
- A custom triplet loss function and cosine similarity are used for learning the semantic similarities.
- Compact hash codes are generated for fast and effective retrieval.
- Practical Application in Digital Pathology.
- State-of-the-Art Performance on Kimia Path24 dataset.

Pathologists examine microscopic samples of tissues to diagnose diseases such as cancer. Finding a matching sample among thousands of images in reference books can be time-consuming and challenging. In this research, we have designed an intelligent system that can quickly find similar images from a large database by simply viewing a tissue image. We used a method called "deep learning" (a branch of artificial intelligence).

We showed the computer numerous examples of healthy and diseased tissue images to learn to identify important features of each image - such as cell shape or arrangement. Then, we created a "hash code" for each image, which acts like a digital fingerprint. When a user provides a new image to the system, it quickly compares its code with the codes of other images and finds the most similar ones. This technology can greatly assist in two areas:

1. Clinical Diagnosis: Pathologists, upon seeing an unclear or suspicious sample, can immediately view confirmed similar samples to make a more accurate decision.
2. Education: Medical students can observe dozens of similar examples by searching for a single image, which aids in their practice.

## 3. RESULTS AND DISCUSSION

The experimental results on the Kimia Path24c dataset certified the superior nature of our suggested methods with respect to the existing baselines. For example, in **Table 1**, the proposed method I (autoencoder-based hashing with EfficientNet-B0 feature extraction) shows very fine retrieval performance with an $\eta_{\_total}$ of 88.82%. It outperforms other existing methods, such as LBP (41.33%), basic CNNs (41.80%), and Inception-V3 (56.98%) [21,22]. Also, the performance surpasses that of VGG16 (83.59%) and Inception-V3 (84.5%) according to [23] because of the deeper, residual learning-based architecture of EfficientNet-B0. The performance superiority over these architectures further suggests that EfficientNet-B0's ability to deal with the vanishing gradient problem and preserve fine-grained spatial information makes it especially suitable for medical image retrieval tasks. On the other hand, the autoencoder-based hashing method is less accurate in comparison with the latest advanced methods, for example, DenseNet-121 [23] (91.62%) and MA + MS-loss (94.95%) [24], thus indicating that although EfficientNet-B0 can process features with very high precision, our proposed method does not extract features as discriminatively as end-to-end supervised learning or metric learning approaches. In comparison with this, the triplet-network-based hashing method has achieved that level of performance generally considered to be state-of-the-art, with an $\eta_{\_total}$ of 97.39% (meaning that it significantly outperforms all existing methods, such as DenseNet-121 and MA + MS-loss). Such huge progress highlights the importance of deep metric learning in CBHIR remark, where the triplet network is capable of explicitly optimizing the feature distances to achieve intra-class compactness and inter-class separability, as shown by the experimental results. It clearly shows that the combination of EfficientNet-B0 with the triplet loss framework would allow the embedding space to be more finely tuned, leading to almost perfect retrieval accuracy.

The fact that EfficientNet-B0 works better than these architectures also implies that its capability to handle the vanishing gradient problem and maintain the details of spatial

**Table 1.** Performance of baseline methods on the Kimia Path24c dataset.

| Paper | Method | $\eta_p$ | $\eta_w$ | $\eta_{total}$ |
|---|---|---|---|---|
| Babaie et al. 21] | LBP24, L1 | 66.11 | 62.52 | 41.33 |
| Babaie et al. [21] | CNN | 64.98 | 64.75 | 41.80 |
| Babaie et al. [21] | DOG | 84.45 | 81.21 | 68.58 |
| Kieffer et al. [22] | Inception-V3 | 74.87 | 76.10 | 56.98 |
| Shafiei et al. [23] | VGG16 | 92.00 | 90.86 | 83.59 |
| Shafiei et al. [23] | Inception-V3 | 92.45 | 91.40 | 84.5 |
| Shafiei et al. [23] | DenseNet 121 | 95.92 | 95.51 | 91.62 |
| Yang et al. [24] | MA + MS-loss | 97.89 | 97.00 | 94.95 |
| The proposed hashing method I [ours] | AE | 94.41 | 94.08 | 88.82 |
| The proposed hashing method II [ours] | Triplet-network | 98.76 | 98.61 | 97.39 |

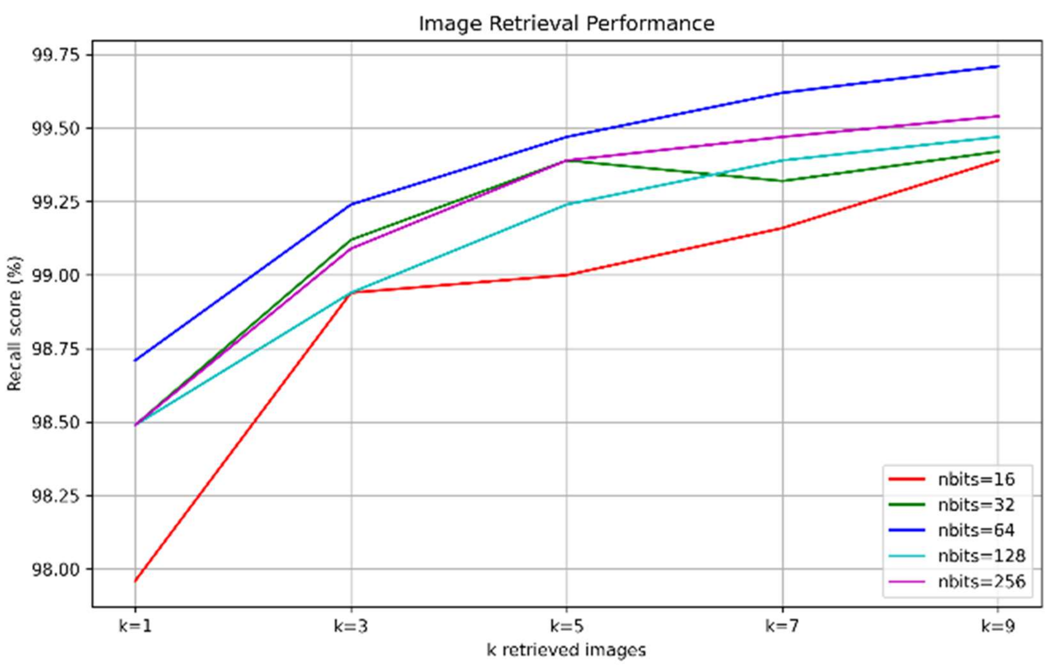


Figure 3. Retrieval performance with respect to the number of k retrieved images.

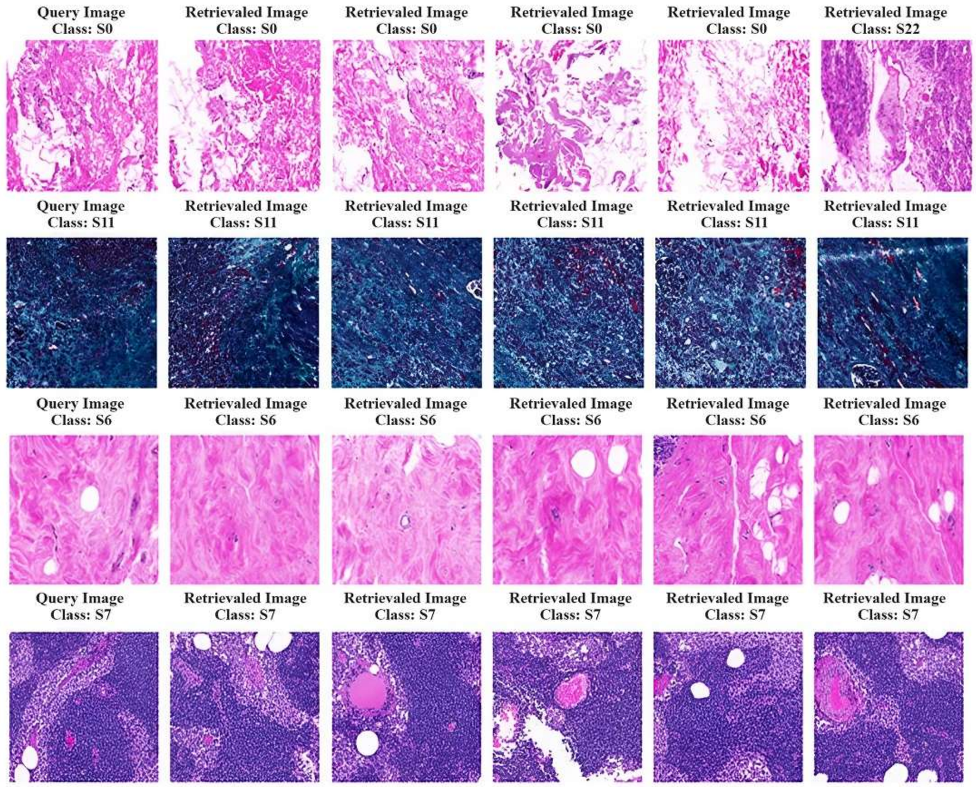


**Figure 4**. Top 5 retrieved images from the Kimia Path24c dataset for five randomly selected query images. The class label for each retrieved image is shown at the top of the image, indicating whether it is a true or false retrieval result.

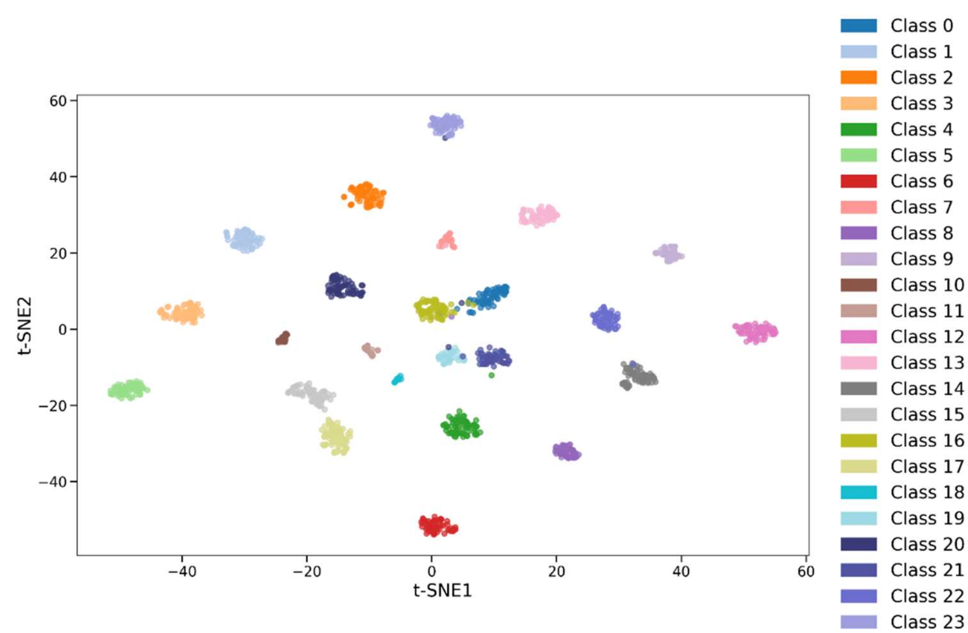


**Figure 5**. t-SNE visualization of embeddings from the Kimia Patch24C test dataset using proposed method II.

information for image retrieval makes it an excellent choice for medical image retrieval tasks. In spite of this, the autoencoder-based hashing method is less accurate than the recent advanced methods, for example, DenseNet-121 [23] (91.62%) and MA + MS-loss (94.95%) [24]. This highlights that EfficientNet-B0 is a backbone that provides strong and robust feature extraction; the proposed method may not be able to leverage as much of its discriminative potential as that of the end-to-end supervised or metric learning approaches. On the other hand, the triplet-network-based hashing method records the best performance thus far, with an $\eta_{total}$ of 97.39% obtained, which is significantly better than all previous works, including DenseNet-121 and MA + MS-loss. Such a drastic increase emphasizes the role of deep metric learning in histopathological image retrieval, where the triplet network's facilitation of feature distance optimization directly leads to the establishment of intra-class compactness and inter-class separability. The findings indicate that although EfficientNet-B0 may be ideal for the initial extraction of features, the integration of a triplet loss framework further constrains the latent space, thereby making retrieval effectiveness almost perfect.

In this research, we assess how well our method works on the Kimia Path24c dataset by varying the number of images retrieved and the length of the hash code. As shown in **Figure 3**, 64-bit hash codes outperform not only shorter (16, 32-bit) hash codes but also longer (128, 256-bit) hash codes. This is due to 64-bit codes offering nearly sufficient representation capacity while still maintaining a low computational cost. Shorter codes possess less discrimination strength, resulting in higher collision rates and diminished retrieval accuracy, whereas longer hash codes may contain redundancy without appreciable performance advantages, potentially leading to overfitting and increased computational costs. As a result, 64-bit hash codes are the best option, as they provide enough discriminability with the 24-class dataset to acquire sufficient data without incurring excessive generalization and processing costs.

To illustrate the visual performance of the suggested technique in CBHIR, **Fig. 4** presents the top 5 retrieved images for four randomly selected classes. From the returned images, it is clear that the performance of our method is effective in retrieving histopathological images that are visually similar to the query images. Furthermore, to confirm the effectiveness of our triplet network-based hashing method, the design of the learned embedding space is scrutinized using t-distributed stochastic neighbor embedding (t-SNE) [29]. In fact, the 2D view of test dataset embeddings, as shown in **Fig. 5**, exhibits several clusters of separation that correspond to different tissue classes, where semantically similar samples have become neighboring groups in the embedding space. As it turns out, a major point of the visualization is that the method employed in this work retains many topological relationships intact; thus, dissimilar cases remain far apart, while closely positioned are those images with similar histopathological features. By allowing these binary representations to maintain the semantic structure of the original embedding space, this spatial arrangement demonstrates that the triplet network is proficient in encoding the discriminative features into the short hash codes.

## 4. CONCLUSION

This research paper introduces a new hierarchical structure for Content-Based Image Retrieval (CBIR) methods that utilize deep learning techniques. The method includes EfficientNet-B0 for feature extraction, and a triplet network trained with a custom triplet loss function. The proposed loss function is a combination of cosine similarity and embedding space learning, which can maintain the semantic proximity of histopathological images. Two hashing models were proposed: one using a dense autoencoder and the other using a triplet network with a custom triplet loss.

The experimental results showed that the triplet network-based model is able to generate concise hash codes as a result of embeddings and binarization using cosine similarity. This method allows for fast retrieval of visually similar images, which also contributes to the efficiency of the image retrieval application. The model was trained on the free Kimia Path24c dataset and achieved very impressive results. The results show that the proposed method can be an excellent tool in the future regarding the effectiveness and accuracy of CBIR for digital pathology.

### Acknowledgment

Not applicable.

### Conflict of interest

The author declares that there are no competing financial interests or personal relationships that could have influenced the research reported in this study.

### Funding

This research received no internal or external funding.

### Ethical statement

Not applicable.

## References

1. Chen W, Liu Y, Wang W, Bakker EM, Georgiou T, Fieguth P, Liu L, Lew MS. Deep learning for instance retrieval: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022 Nov 1;45(6):7270-92.

2. Lowe DG. Distinctive image features from scale-invariant keypoints. International journal of computer vision. 2004 Nov;60(2):91-110.

3. Sivic, Zisserman. Video Google: A text retrieval approach to object matching in videos. InProceedings ninth IEEE international conference on computer vision 2003 Oct 13 (pp. 1470-1477). IEEE.

4. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In2009 IEEE conference on computer vision and pattern recognition 2009 Jun 20 (pp. 248-255). Ieee.

5. Pang S, Ma J, Xue J, Zhu J, Ordonez V. Deep feature aggregation and image re-ranking with heat diffusion for image retrieval. IEEE Transactions on Multimedia. 2018 Oct 18;21(6):1513-23.

6. Chew EJ, Tan PH. Evolutionary changes in pathology and our understanding of disease. Pathobiology. 2023 Jun 6;90(3):209-18.

7. Masjoodi S, Anbardar MH, Shokripour M, Omidifar N. Whole Slide Imaging (WSI) in Pathology: Emerging Trends and Future Applications in Clinical Diagnostics, Medical Education, and Pathology. Iranian Journal of Pathology. 2025 Jul 1;20(3):257.

8. Zheng Y, Jiang Z, Ma Y, Zhang H, Xie F, Shi H, Zhao Y. Content-based histopathological image retrieval for whole slide image database using binary codes. InMedical Imaging 2017: Digital Pathology 2017 Mar 1 (Vol. 10140, pp. 266-271). SPIE.

9. Adem AM, Kant R, Kumar K, Mittal V, Jain P. Exploring self-supervised learning for disease detection and classification in digital pathology: A review. Biomedical and Pharmacology Journal. 2025 Feb 20;18(March Spl Edition):59-72.

10. Zheng L, Wetzel AW, Gilbertson J, Becich MJ. Design and analysis of a content-based pathology image retrieval system. IEEE transactions on information technology in biomedicine. 2003 Dec 31;7(4):249-55.

11. Banu MS, Nallaperumal K. Analysis of color feature extraction techniques for pathology image retrieval system. In2010 IEEE International Conference on Computational Intelligence and Computing Research 2010 Dec 28 (pp. 1-7). IEEE.

12. Calderaro S, Bosco GL, Rizzo R, Vella F. Deep metric learning for histopathological image classification. In2022 IEEE Eighth International Conference on Multimedia Big Data (BigMM) 2022 Dec 5 (pp. 57-64). IEEE.

13. Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. Medical image analysis. 2017 Dec 1;42:60-88.

14. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. nature. 2017 Feb 2;542(7639):115-8.

15. Spanhol FA, Oliveira LS, Cavalin PR, Petitjean C, Heutte L. Deep features for breast cancer histopathological image classification. In2017 IEEE international conference on systems, man, and cybernetics (SMC) 2017 Oct 5 (pp. 1868-1873). IEEE.

16. Sharma H, Zerbe N, Klempert I, Hellwich O, Hufnagl P. Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology. Computerized Medical Imaging and Graphics. 2017 Nov 1;61:2-13.

17. Zheng Y, Jiang Z, Xie F, Zhang H, Ma Y, Shi H, Zhao Y. Feature extraction from histopathological images based on nucleus-guided convolutional neural network for breast lesion classification. Pattern Recognition. 2017 Nov 1;71:14-25.

18. Shi X, Sapkota M, Xing F, Liu F, Cui L, Yang L. Pairwise based deep ranking hashing for histopathology image classification and retrieval. Pattern Recognition. 2018 Sep 1;81:14-22.

19. Wang Z, Zheng L, Liu Y, Li Y, Wang S. Towards real-time multi-object tracking. InEuropean conference on computer vision 2020 Aug 23 (pp. 107-122). Cham: Springer International Publishing.

20. Ustinova E, Lempitsky V. Learning deep embeddings with histogram loss. Advances in neural information processing systems. 2016;29.

21. Babaie M, Kalra S, Sriram A, Mitcheltree C, Zhu S, Khatami A, Rahnamayan S, Tizhoosh HR. Classification and retrieval of digital pathology scans: A new dataset. InProceedings of the IEEE conference on computer vision and pattern recognition workshops 2017 (pp. 8-16).

22. Kieffer B, Babaie M, Kalra S, Tizhoosh HR. Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. In2017 seventh international conference on image processing theory, tools and applications (IPTA) 2017 Nov 28 (pp. 1-6). IEEE.

23. Shafiei S, Babaie M, Kalra S, Tizhoosh HR. Colored Kimia Path24 dataset: configurations and benchmarks with deep embeddings. arXiv preprint arXiv:2102.07611. 2021 Feb 15.

24. Yang P, Zhai Y, Li L, Lv H, Wang J, Zhu C, Jiang R. A deep metric learning approach for histopathological image retrieval. Methods. 2020 Jul 1;179:14-25.

25. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. InProceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 7132-7141).

26. Kertész G. Combining negative selection techniques for triplet mining in deep metric learning. In2022 IEEE 10th Jubilee International Conference on Computational Cybernetics and Cyber-Medical Systems (ICCC) 2022 Jul 6 (pp. 000155-000160). IEEE.

27. Kapoor R, Sharma D, Gulati T. State of the art content-based image retrieval techniques using deep learning: a survey. Multimedia Tools and Applications. 2021 Aug;80(19):29561-83.

28. Choi H, Kim M, Lee G, Kim W. Unsupervised learning approach for network intrusion detection system using autoencoders: H. Choi et al. The Journal of Supercomputing. 2019 Sep;75(9):5597-621.

29. Van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of machine learning research. 2008 Nov 1;9(11).