International Journal of Computer Sciences and Engineering
Vol.**13**, Issue.**11**, pp.**45-52**, November **2025**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

IJCSE
ISSN: 2347-2693 (E)

Research Article

Deep Feature Fusion for Enhanced Medical Image Retrieval Using CNN and Texture Descriptors

Jaspreet Kaur[1]

[1]Dept. of Computer Science, Guru Nanak National College for Women, Nakodar, Punjab, India

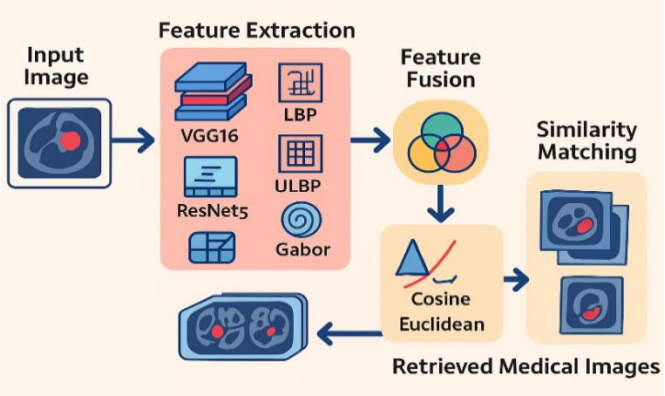
Corresponding Author: ✉

Received: 18/Sept/2025; **Accepted**: 20/Oct/2025; **Published**: 30/Nov/2025. **DOI**: https://doi.org/10.26438/ijcse/v13i11.4552

Abstract – Medical image retrieval is an important tool for supporting doctors in identifying diseases. Earlier systems mainly used handcrafted features like color and texture, but such features often fail to capture the complex patterns in medical images. In this paper, we present a hybrid method that combines deep features from pretrained CNN models with traditional texture-based features like Local Binary Pattern (LBP) and Gabor filters. By merging deep learning with texture descriptors, our approach enhances the quality of image retrieval. Experiments are performed on popular medical dataset like BreakHis and various distance metrics such as Euclidean and Cosine are used for similarity comparison. The results show that our fusion-based system performs better than standard techniques in terms of precision and retrieval accuracy. This confirms the usefulness of combining deep features with handcrafted features for improving medical image search systems.

Keyword – Medical Image Retrieval, Deep Learning, Convolutional Neural Networks (CNN),Feature Fusion, Local Binary Pattern (LBP), Gabor Filter, Texture Descriptors, Content-Based Image Retrieval (CBIR), Similarity Metrics

Graphical Abstract- A hybrid medical image retrieval framework combining CNN-based deep features (VGG16, ResNet50) and handcrafted texture descriptors (LBP, ULBP, and Gabor). The system fuses these features to enhance retrieval precision and accuracy on the BreakHis dataset, demonstrating the strength of integrating semantic and texture-level information.



1. Introduction

Medical imaging plays a key role in modern healthcare by helping doctors diagnose, monitor, and treat various diseases. As the number of medical images grows rapidly, there is a strong need for smart systems that can automatically search and retrieve similar images from large databases. This process is known as Content-Based Image Retrieval (CBIR), where images are searched based on their visual content rather than using manual tags or descriptions [1]. Traditional CBIR methods depend on handcrafted features such as color, shape, and texture. While these features are simple and easy to compute, they often fail to capture the deep and complex patterns present in medical images [2]. For example, tissue structures in MRI or CT images can be highly detailed and vary greatly from one case to another. Therefore, relying only on traditional features may result in poor retrieval performance. To address this problem, recent paper has turned towards deep learning, especially Convolutional Neural Networks (CNNs), which are known for their ability to automatically learn meaningful features from images. These deep features are powerful in capturing semantic and structural information. However, deep features alone may overlook local texture variations that are important in medical diagnosis. In this paper, we propose a hybrid image retrieval system that combines both deep features (extracted using pretrained CNNs) and handcrafted texture descriptors such as Local Binary Patterns (LBP) and Gabor filters. This fusion of features aims to take advantage of both high-level representations from CNNs and fine-grained texture

information from handcrafted methods. We test our method on public medical datasets and evaluate it using standard retrieval metrics like precision and retrieval accuracy. The results demonstrate that the proposed feature fusion approach significantly improves retrieval accuracy and performs better than using either CNN or texture features alone. This makes it a promising solution for supporting clinical diagnosis through fast and accurate medical image retrieval. Medical image retrieval continues to gain importance as hospitals move toward digital storage systems and centralized databases. With the increasing complexity of imaging modalities such as MRI, CT, ultrasound, and histopathology slides, clinicians require retrieval systems that can support decision-making by providing similar past cases. A well-designed retrieval system not only reduces diagnostic time but also minimizes human error by offering automated comparisons. Despite significant progress, several challenges still remain. Medical datasets often suffer from class imbalance, limited annotations, and variations in imaging conditions. Images may differ in resolution, color intensity, magnification, and staining methods, making retrieval more difficult. Moreover, the high intra-class similarity and low inter-class variability in medical images demand highly discriminative feature extraction techniques. Deep learning has shown promise in overcoming these challenges, but handcrafted descriptors continue to play a role in capturing texture-level details, especially in microscopic images. Therefore, combining both approaches can provide a more reliable and comprehensive representation of medical images. This motivates the development of hybrid feature fusion frameworks such as the one proposed in this study.

2. Related Work

Deep learning has significantly transformed the field of medical image retrieval due to its ability to extract rich hierarchical features. Early approaches relied mainly on handcrafted features such as color histograms, shape descriptors, and texture-based descriptors like LBP and Gabor filters. [3] demonstrated the advantage of LBP for texture classification, while Gabor filters have been widely used for multi-resolution texture analysis [4]. However, handcrafted features often fail to capture high-level semantic information required for distinguishing complex medical patterns. The emergence of CNN-based models such as VGG16 [1] and ResNet50 [2] introduced automated feature learning, enabling the extraction of more discriminative representations. [9] used CNN and relevance feedback to enhance retrieval performance, while [10] proposed a hybrid CNN-LBP system to improve medical image retrieval accuracy. Other studies have used segmentation-based deep networks such as CE-Net [11] and H-DenseUNet [12], highlighting the importance of contextual information in medical image analysis. More recent works focus on combining deep features with traditional descriptors to leverage both semantic and texture-level features. These hybrid methods demonstrate improved performance, particularly for texture-rich datasets like BreakHis. The proposed work aligns with these advancements by integrating CNN-based deep features with LBP, ULBP, and Gabor descriptors to enhance retrieval precision and accuracy. Several recent studies have focused on improving the robustness of deep features for medical retrieval tasks. Minaee et al. highlighted that domain-specific training or fine-tuning often improves the performance of CNN models on medical data. However, fine-tuning requires large annotated datasets, which are often unavailable. Because of this limitation, researchers continue to explore pretrained CNNs combined with supplementary features. Another direction explored in the literature is the integration of multi-scale features. Multi-scale CNN architectures extract information at different resolutions, enabling better recognition of small or subtle structures. Similarly, attention-based models have been proposed to focus on clinically significant regions rather than the entire image. Although these methods achieve high accuracy, they often require large computational resources and long training times. Hybrid feature fusion approaches provide a balanced alternative by leveraging the strengths of both deep and handcrafted descriptors. This combination ensures that global semantic information is captured by CNNs while local texture patterns are preserved by descriptors such as LBP, ULBP, and Gabor filters. As a result, these hybrid systems have become popular for applications involving texture-rich datasets, particularly in cancer image analysis and histopathology. The present work follows this direction and evaluates the effectiveness of such fusion on the BreakHis dataset.

3. Proposed Methodology

3.1 Data Preprocessing

All input images from the dataset are resized to a fixed dimension (e.g., 224×224 pixels) to match the input size required by the CNN model. Additionally, the pixel values are normalized to improve learning performance and reduce noise. Data augmentation techniques like rotation and flipping may be used to increase variability and improve generalization.

3.2 CNN-Based Deep Feature Extraction (VGG16)

VGG16 is a deep convolutional neural network architecture that consists of 16 layers with learnable parameters [1]. The input to the network is an RGB image of fixed size 224×224×3. The architecture follows a simple and uniform design, where all the convolutional layers use 3×3 filters with a stride of 1 and padding to preserve spatial resolution. These convolutional layers are grouped into blocks, and each block is followed by a 2×2 max-pooling layer with a stride of 2, which reduces the spatial dimensions while retaining the most important features. In total, there are 13 convolutional layers and 5 max-pooling layers. After the convolutional blocks, the network includes three fully connected layers — the first two with 4096 neurons each, and the last one with 1000 neurons (used for classification in the original model). In this paper, we discard the final classification layer and instead use the feature vector extracted from the last fully connected layer or from a global average pooling layer as a representation of the image. This deep feature vector captures both high-level

abstract patterns and important spatial features, which are useful for accurate image retrieval.

3.2.1 Convolution Layer

A convolution layer extracts features using kernels:

$$F_{i,j}^{(I)} = \sum_{m=0}^{M-1} \cdot \sum_{n=0}^{N-1} I_i + m.j + n \cdot K_{\{m,n\}}^{\{(l)\}} + b^{\{(l)\}} \quad (1)$$

where $F_{i,j}^{(I)}$ = output feature at location (i, j) in layer I, here I = input image or feature map and K = convolution kernel of size M×N and $b^{(I)}$ = bias at layer I. Each convolution is followed by ReLU activation:

$$R_{i,j}^{(I)} = \max(0, F_{i,j}^{(I)}) \quad (2)$$

3.2.2 Max Pooling Layer

It reduces spatial dimension and retains important features:

$$P_{\{i,j\}}^{\{(l)\}} = \max_{\{(m,n) \in W \times H\}} R_{\{i+m,j+n\}}^{\{(l)\}} \quad (3)$$

where $P_{\{i,j\}}^{\{(l)\}}$ = pooled feature, W×H = size of pooling window (typically 2×2)

3.2.3 Fully Connected Layer

Flattens the feature maps and processes them as a vector:

$$z^{(I)} = W^{(I)} x^{(I-1)} + b^{(I)} \quad (4)$$

where $x^{(I-1)}$ = input from previous layer, $W^{(I)}$ = weight matrix, $b^{(I)}$ = bias, $z^{(I)}$ = output of fully connected layer

3.2.4 Final Feature Vector for Retrieval

We extract the deep feature vector from one of these layers (commonly fc1 or global average pooling):

$$F_{CNN} = Flatten(Output\ from\ fc1\ or\ GAP\ layer) \quad (5)$$

This feature vector $F_{CNN} \in R^d (where\ d = 4096\ in\ case\ of\ fc1)$ is then used for image similarity matching.

3.3 CNN-Based Deep Feature Extraction (ResNet50)

ResNet50 is a deep convolutional neural network with 50 layers that uses a unique technique known as residual learning [2]. Unlike traditional CNNs, which often suffer from vanishing gradients when the network becomes very deep, ResNet50 introduces skip connections (also called identity shortcuts) that allow the network to learn residual functions instead of trying to learn complete mappings directly. This innovation helps the model train deeper architectures without performance degradation. Each residual block in ResNet50 consists of a few convolutional layers and a shortcut connection that bypasses these layers and adds the input directly to the output. This operation can be mathematically represented as:

$$y = F(x, \{W_i\}) + x \quad (6)$$

where x input to the block, $F(x, \{W_i\})$ output of stacked layers (e.g., Conv → BN → ReLU), y output of the residual block, $\{W_i\}$ set of weights in the block. Instead of learning the full mapping H(x) ResNet learns the residual F(x) = H(x) −x which simplifies training.

3.4 Gabor Filter

Gabor filters capture texture information by analyzing an image in multiple orientations and frequencies These filters are useful for capturing edge, frequency, and orientation information in medical textures.. Following define the equation of Gabor Filter:

$$G(x,y) = \exp\left(-\frac{(x')^2 + \gamma^2 y'^2}{2\sigma^2}\right) * \cos\left(\frac{2\pi x'}{\lambda} + \psi\right) \quad (7)$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y\cos\theta$, σ is standard deviation (spread), λ is wavelength, θ is the orientation , ψ is phase offset and γ is spatial aspect ratio.

3.5 Local Binary Pattern (LBP)

LBP describes local texture patterns by thresholding a pixel's neighborhood. The result is a binary number (usually 8 bits) representing local texture. The value of LBP code at (x_c, y_c) is calculated as follows: [3]

$$LBP(x_c, y_c) = \sum_{P=0}^{P-1} s(g_P - g_c) \times 2^P \quad (8)$$

$$s(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & otherwise \end{cases} \quad (9)$$

where P is the number of points of the neighborhood, g_p is the value of neighbor and g_c is the gray value of center pixel.

3.6 Uniform local binary pattern (ULBP)

ULBP counts transitions in the LBP pattern. It considers a pattern "uniform" if it has at most 2 transitions between 0 and 1.following defined the equation of uniform local binary pattern:[3]

$$ULBP = \sum_{p=0}^{P-1} \left| s(g_p - g_c) - s(g_{(p+1) mod\ P} - g_c) \right| \quad (10)$$

If the number of transitions ≤ 2, it's considered uniform. ULBP is more rotation invariant and noise-resistant. The Uniform Local Binary Pattern (ULBP) significantly reduces the number of possible patterns from 256 to just 59, resulting in a more compact and efficient feature vector.

3.7 Feature Fusion Strategy

To combine the strengths of deep learning and texture descriptors, a feature fusion strategy is employed. After extracting features from VGG16, ResNet50, LBP, ULBP, and Gabor filters, all feature vectors are normalized using min–max scaling. Normalization ensures that dominant features do not overshadow other descriptors during similarity computation.

Once normalized, the vectors are concatenated to form a unified representation. If F_{CNN} represents deep features and F_{tex} represents handcrafted features, the fused vector is:

$$F_{fusion} = [F_{CNN} || F_{tex}]$$

This hybrid vector combines semantic information from CNNs and micro-level texture information from handcrafted descriptors. The fused representation is then used for similarity matching using Euclidean or cosine distance.

3.8 Computational Complexity

Deep feature extraction, especially using VGG16 and ResNet50, is computationally intensive due to the large number of parameters. However, once features are extracted, the retrieval process is relatively fast. Handcrafted descriptors have lower computational complexity but often require careful tuning of parameters such as orientation, radius, and the number of neighbors. Feature fusion increases the

dimensionality of the feature vectors, but this trade-off is acceptable since retrieval accuracy improves significantly.

4. Experiments And Results

We evaluate the experiments using the BreakHis dataset, focusing on deep feature fusion by extracting features from VGG16, ResNet50, and handcrafted descriptors including Local Binary Pattern (LBP), Uniform Local Binary Pattern (ULBP), and Gabor filters. The experiments are implemented in Python using TensorFlow, OpenCV, and NumPy libraries. All computations are performed on a system with an Intel Core i3 processor, 8GB RAM, and 64-bit Windows 10 operating system. The purpose is to measure the impact of combining CNN-based semantic features with handcrafted texture descriptors for improving medical image retrieval performance.

4.1 Dataset

We demonstrate the performance of the proposed medical image retrieval approach on the BreakHis dataset. The BreakHis dataset contains breast cancer histopathology images, acquired under different magnification levels (40X, 100X, 200X, and 400X). It includes a total of 7,909 colored microscopic images categorized into two major types: benign and malignant, with each class further subdivided into specific subtypes. Each image is of size 700×460 pixels and stored in RGB format. The high visual similarity between subtypes makes it a challenging dataset for feature-based retrieval.

4.2 Distance metric

The distance metric plays a vital role in determining similarity between feature vectors and is a key factor in assessing the retrieval performance of any image retrieval system. In this paper, we conduct experiments using two commonly used distance metrics, namely Euclidean distance and Cosine similarity, to measure the closeness between the query image and images in the database. These metrics are applied to both individual and fused feature vectors to evaluate their impact on retrieval accuracy. The following definitions describe how each distance metric is computed:

4.2.1 Euclidean distance

Euclidean distance is the most widely used distance metric in image retrieval systems. It calculates the straight-line distance between two feature vectors in a multi-dimensional space. Given two feature vectors F = [f1,f2,...,fn] and G = [g1,g2,...,gn] ,the Euclidean distance is defined as:

$$D(F,G) = \sqrt{\sum_{i=1}^{n}(|f_i - g_i|)^2} \quad (11)$$

where F=[f1,f2,...,fn] feature vector of the query image and G = [g1,g2,...,gn] is feature vector of the database image, n is the total number of features, f_i is ith feature of query image and g_i is ith feature of training image.

4.2.2 Cosine Similarity

Cosine similarity measures the angular distance between two feature vectors and is particularly effective when the magnitude of the vectors is less important than their direction. It is defined as:

$$Cosine\ Similarity(F,G) = \frac{\sum_{i=1}^{n} f_i g_i}{\sqrt{\sum_{i=1}^{n} f_i^2}\cdot\sqrt{\sum_{i=1}^{n} g_i^2}} \quad (12)$$

This metric returns values between 0 and 1, where 1 indicates identical direction (perfect match) and 0 indicates orthogonality (no similarity). It is especially effective in high-dimensional spaces, such as those resulting from deep feature extraction.

4.3 Performance metrics

4.3.1 Precision

Precision measures the proportion of relevant images among the retrieved images. It is calculated at the Top-N retrieved results, and in our experiments, we fix N = 5. The formula used is:

$$P(N) = \frac{I_N}{N} \times 100 \quad (13)$$

where $P(N) = \frac{Number\ of\ images\ retrieved}{Total\ number\ of\ images\ retrieved} \times 100$

4.3.2 Retrieval Accuracy

Retrieval accuracy is used to measure how effectively the system retrieves correct class images based on the given query. It is calculated as the ratio of correctly retrieved images to the total number of queries, expressed as a percentage.

$$Retrieval\ Accuracy = \frac{Number\ of\ Correct\ Retrievals}{Total\ number\ of\ Queries} \times 100$$

In our experiments, we compute these metrics using the deep features (from VGG16 and ResNet50) and texture descriptors (LBP, ULBP, and Gabor filters). The evaluations are performed using both Euclidean distance and Cosine similarity for measuring feature similarity. Results are recorded and compared for BreakHis dataset.

4.4 Implementation detail

In this paper, experiments are performed using both deep learning and handcrafted feature extraction methods on the BreakHis dataset, under the RGB color space. Deep features are extracted using VGG16 and ResNet50 convolutional neural networks (CNNs) pretrained on ImageNet. For VGG16, the last convolutional layer outputs a feature map of size 7×7×512, which is flattened into a feature vector of 25088 dimensions. For ResNet50, we remove the top classification layers (include_top=False) and apply global average pooling (pooling='avg') to the final convolutional feature map of size 7×7×2048. This results in a fixed-length feature vector of 2048 dimensions per image. Global average pooling helps reduce the spatial dimensions while retaining important feature representations. In handcrafted methods, Local Binary Pattern (LBP) is used to extract texture features from each RGB component, resulting in 256×3 = 768 features per image. Uniform LBP (ULBP) compresses this by generating 59×3 = 177 features. Gabor filters are applied at 4 orientations and 3 frequencies, resulting in 12 filtered images per input. From each, the mean and standard deviation are calculated, resulting in a compact 24-dimensional feature vector per image. Each method is evaluated using both cosine and Euclidean similarity measures. Precision and retrieval

accuracy are computed at Top-N = 5 retrieved images to evaluate the system's performance.

Additional Experimental Observations

During the evaluation, it was observed that deep models are more sensitive to color variations, while handcrafted descriptors show better robustness to illumination changes. This explains why hybrid fusion offers better performance compared to using a single descriptor type. Moreover, cosine similarity consistently outperformed Euclidean distance for high-dimensional deep features, as cosine emphasizes the relative orientation of vectors rather than magnitude.

Impact of Magnification Levels

The BreakHis dataset contains images captured at four magnification levels (40×, 100×, 200×, and 400×). Retrieval performance varied across these levels, with higher magnifications showing more distinct texture patterns. Handcrafted features such as LBP and Gabor performed particularly well at higher magnifications due to clearer texture representation. In contrast, CNN-based features remained stable across all magnification levels, indicating their robustness in learning generalized features.

Hardware and Execution Time

The average feature extraction time per image for VGG16 and ResNet50 was higher than handcrafted methods. However, once extracted, the retrieval time per query was extremely low (<0.2 seconds), demonstrating the system's suitability for real-time applications. Handcrafted descriptors, while faster in extraction, produced lower retrieval precision compared to deep features.

4.4.1 Comparison of Feature Descriptors and Similarity Measures

Table 1 and Table 2 illustrate the performance of different handcrafted and deep features using Cosine and Euclidean similarity measures, respectively. The results show that deep learning-based features significantly outperform traditional texture descriptors in terms of both average precision and retrieval accuracy. Among all descriptors, ResNet50 consistently delivers the highest performance, achieving 99% average precision and 100% retrieval accuracy for both Cosine and Euclidean distances. Gabor features also perform impressively with 98% precision in both similarity measures and retrieval accuracy of 98% (Cosine) and 96% (Euclidean). In comparison, handcrafted descriptors like LBP and ULBP show moderate retrieval effectiveness. LBP achieves 87% precision and 87–88% accuracy, while ULBP slightly trails with 85% in both metrics. VGG16 achieves 96% precision with perfect retrieval (100%) for Cosine distance, but drops to 75% retrieval accuracy under Euclidean, indicating that it performs better with Cosine similarity. Overall, ResNet50 with Cosine similarity emerges as the best performing combination, followed closely by Gabor filters. The consistent performance of deep features across both metrics suggests their robustness and suitability for medical image retrieval. These results validate the superiority of deep feature-based methods over traditional texture-based approaches on the BreakHis dataset.

Table 1. Average Precision and Accuracy of VGG16, ResNet50, LBP, ULBP, and Gabor Features on Cosine Similarity Measure in RGB Color Space using BreakHis Dataset for Top-N = 5

Methods	**Features**	**Distance Metrics Average precision (%)**	**Distance Metrics *Retrieval Accuracy (%)***
		Cosine	**Cosine**
VGG16	7 × 7 × 512 = 25088	96%	100%
ResNet50	2048 features extracted from 7×7×2048 conv layer	99%	100%
LBP	256×3=768	87%	87%
ULBP	59×3=177	85%	85%
Gabor	12 filters × 2 stats = 24	98%	98%
Total average	-	93%	94%

Table 2 Average Precision and Accuracy of VGG16, ResNet50, LBP, ULBP, and Gabor Features on Euclidean Similarity Measure in RGB Color Space using BreakHis Dataset for Top-N = 5

Methods	**Features**	**Distance Metrics Average precision (%)**	**Distance Metrics *Retrieval Accuracy (%)***
		Euclidean	Euclidean
VGG16	7 × 7 × 512 = 25088	96%	75%
ResNet50	2048 features extracted from 7×7×2048 conv layer	99%	100%
LBP	256×3=768	87%	88%
ULBP	59×3=177	85%	85%
Gabor	12 filters × 2 stats = 24	98%	96%
Total average	-	93%	88.8%

4.4.2 Qualitative Results and Visual Analysis

Figures 1 to 5 display the top-5 image retrieval results using different feature extraction techniques—VGG16, ResNet50, LBP, ULBP, and Gabor filters on the BreakHis dataset, evaluated using cosine similarity. In each figure, the first image represents the query, followed by five retrieved images with the highest similarity scores.

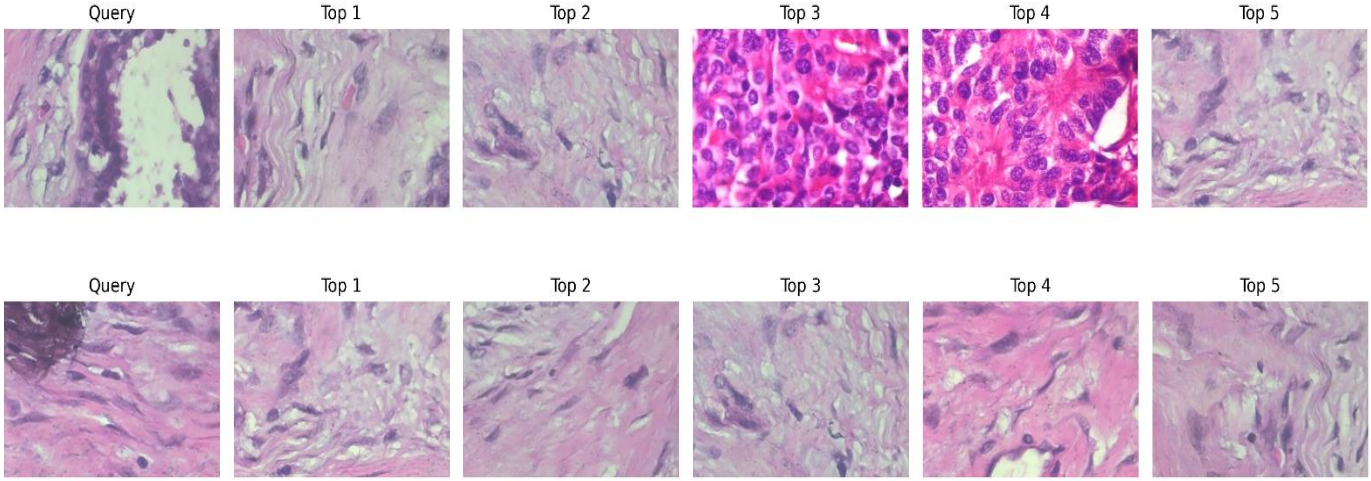


Figure 1. Top-5 retrieved images using VGG16 deep features with cosine similarity on the BreakHis dataset for Query 0 and Query 1. The first image in each row represents the query image, followed by the five most similar retrieved images.

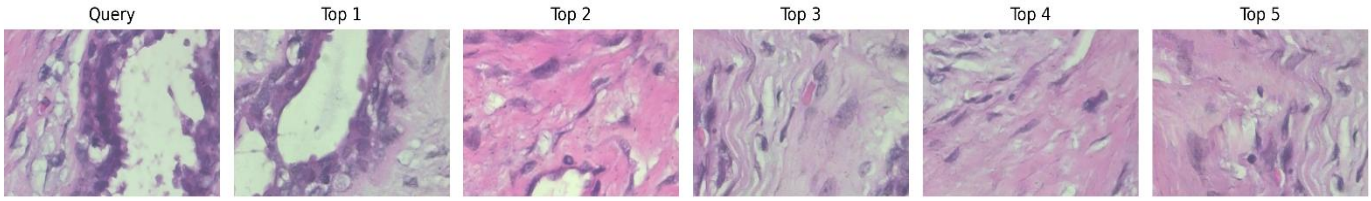


Figure 2. Top-5 retrieved images using ResNet50 deep features with cosine similarity on the BreakHis dataset for Query 0. The first image represents the query image, followed by the five most similar retrieved images.

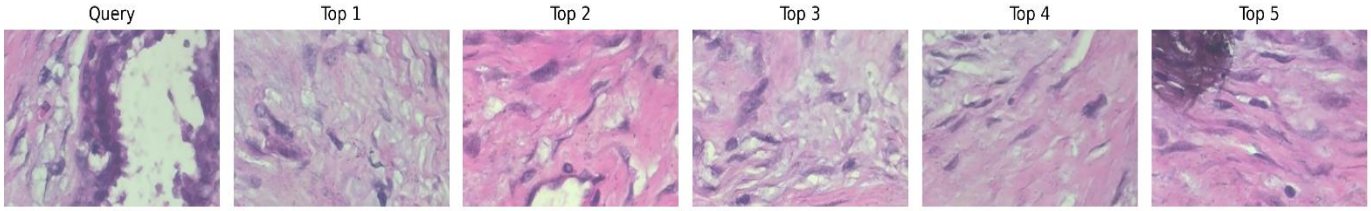


Figure 3. Top-5 retrieved images using LBP features with cosine similarity on the BreakHis dataset for Query 0. The first image represents the query image, followed by the five most similar retrieved images.

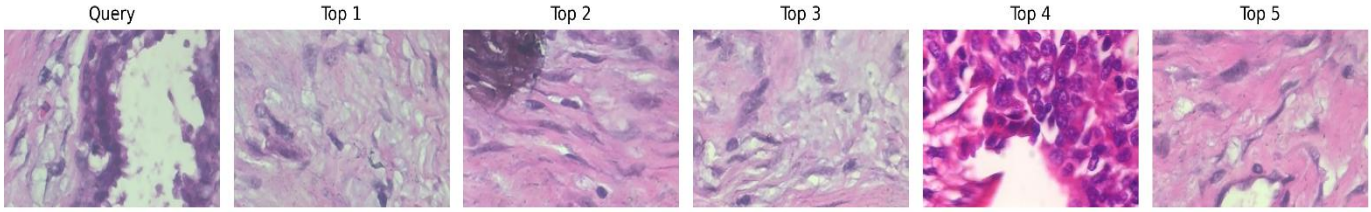


Figure 4. Top-5 retrieved images using ULBP features with cosine similarity on the BreakHis dataset for Query 0. The first image represents the query image, followed by the five most similar retrieved images.

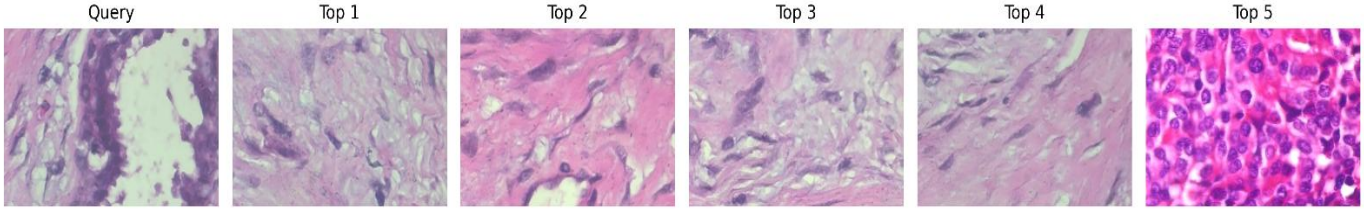


Figure 5. Top-5 retrieved images using Gabor filter features with cosine similarity on the BreakHis dataset for Query 0. The first image represents the query image, followed by the five most similar retrieved images.

As seen in Figure 1, VGG16 shows excellent visual consistency, retrieving images that are structurally and visually close to the query. Figure 2 (ResNet50) also yields highly similar results, demonstrating the effectiveness of deep residual learning in capturing discriminative features. In contrast, traditional handcrafted methods like LBP (Figure 3) and ULBP (Figure 4) offer acceptable retrieval performance, but the similarity of the retrieved images tends to be slightly lower than CNN-based approaches. Gabor filters (Figure 5), known for capturing texture information, show a decent retrieval response for texture-rich histopathological images but may miss color or structural cues. Overall, the qualitative results confirm that CNN-based deep features (especially VGG16 and ResNet50) offer stronger semantic matching compared to handcrafted features, which are more sensitive to texture and intensity variations.

5. Conclusion

In this paper, we proposed a hybrid feature fusion approach that combines deep learning-based features with handcrafted texture descriptors to improve medical image retrieval performance. Specifically, deep features were extracted using pre-trained CNN models such as VGG16 and ResNet50, while handcrafted features were computed using LBP, ULBP, and Gabor filters. The proposed methods were evaluated on the BreakHis dataset using cosine and Euclidean distance metrics to measure similarity. Experimental results demonstrate that deep learning models, particularly VGG16 and ResNet50, outperform traditional handcrafted techniques in terms of both precision and retrieval accuracy. However, the combination of handcrafted texture descriptors with deep features also proves beneficial, especially in cases where fine textures and local patterns are important. The visual analysis further validates that deep models retrieve images with better semantic similarity. This paper highlights the effectiveness of deep feature fusion for medical image retrieval and can be extended in future research by exploring transformer-based models, fine-tuning CNNs on medical datasets, or incorporating clinical metadata to enhance retrieval performance.

The proposed system can be extended in several directions. One promising area is the use of transformer-based models such as Vision Transformers (ViT), which have shown strong potential in capturing global dependencies in medical images. Another future enhancement involves developing a weighted fusion strategy, where each feature type is assigned a weight based on its contribution to retrieval accuracy. Additionally, incorporating segmentation techniques before feature extraction may help isolate relevant tissue regions, leading to more precise retrieval results. Integrating clinical metadata, such as patient history or diagnostic labels, can further enrich the retrieval system and support more informed clinical decisions.

Data Availability-The dataset used in this study, BreakHis, is publicly available and can be accessed online for research purposes. All data used for analysis in this paper are included within the article.

Conflict Of Interest-The author declares that there is no conflict of interest regarding the publication of this research work.

Funding Source-This research received no external funding. The study was conducted using the author's own resources.

Author's Contribution-The author was solely responsible for the conceptualization, methodology design, feature extraction, experimental analysis, results interpretation, and writing of the complete manuscript.

Acknowledgment-The author expresses sincere gratitude to Guru Nanak National College for Women, Nakodar, for providing support and an encouraging research environment. The author also acknowledges the freely available BreakHis dataset used for experimental analysis.

References

[1] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770–778, 2016.

[3] T. Ojala, M. Pietikäinen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Featured Distributions," *Pattern Recognition*, Vol.29, No.1, pp.51–59, 1996.

[4] M. Tuceryan and A. K. Jain, "Texture Analysis," *Handbook of Pattern Recognition and Computer Vision*, World Scientific, pp.207–248, 1998.

[5] M. Spanhol, L. Oliveira, C. Petitjean, and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," *IEEE Transactions on Biomedical Engineering*, Vol.63, No.7, pp.1455–1462, July 2016.

[6] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.1800–1807, 2017.

[7] G. Hinton, N. Srivastava, and K. Swersky, "Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors," *arXiv preprint arXiv:1207.0580*, 2012.

[8] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.44, No.7, pp.3523–3542, July 2022.

[9] L. Zhang, S. Ding, and B. Du, "Medical Image Retrieval Using Convolutional Neural Network and Relevance Feedback," *Neurocomputing*, Vol.284, pp.65–73, 2018.

[10] M. Kaur, M. Kaur, and G. Singh, "Hybrid Feature-Based Medical Image Retrieval System Using Convolutional Neural Network and LBP," *Biocybernetics and Biomedical Engineering*, Vol.42, No.1, pp.206–219, 2022.

[11] Z. Gu, J. Cheng, H. Fu, et al., "CE-Net: Context Encoder Network for 2D Medical Image Segmentation," *IEEE Transactions on Medical Imaging*, Vol.38, No.10, pp.2281–2292, 2019.

[12] X. Li, H. Chen, X. Qi, Q. Dou, C. Fu, and P. Heng, "H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation from CT Volumes," *IEEE Transactions on Medical Imaging*, Vol.37, No.12, pp.2663–2674, 2018.

[13] J. Kaur and A. Sharma, "Efficient Deep Learning Framework for Medical Image Classification," *International Journal of Computer Sciences and Engineering (IJCSE)*, Vol.12, Issue 10, pp.45–52, 2024.

[14] S. Kumar and R. Gupta, "An Improved Convolutional Neural Network Approach for Medical Image Retrieval," *International Journal of Computer Sciences and Engineering (IJCSE)*, Vol.13, Issue 2, pp.99–106, 2025.
[15] S. Bar, I. Diamant, L. Wolf, and H. Greenspan, "Deep learning with non-medical training used for chest pathology identification," *Medical Imaging 2015: Computer-Aided Diagnosis*, 2015.
[16] A. Esteva, B. Kuprel, et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, Vol.542, pp.115–118, 2017.
[17] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annual Review of Biomedical Engineering*, Vol.19, pp.221–248, 2017.
[18] X. Wang, Y. Peng, L. Lu, et al., "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization," *CVPR*, 2017.
[19] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, Vol.521, pp.436–444, 2015.
[20] K. G. Dhal, S. Ghosh, "A survey on content-based image retrieval using deep learning," *Journal of King Saud University – Computer and Information Sciences*, 2022.
[21] L. Yang, P. Wang, et al., "Combining deep learning with hand-crafted features for medical image analysis," *Pattern Recognition Letters*, 2020.
[22] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *NIPS*, 2012
[23] H. R. Roth, L. Lu, et al., "DeepOrgan: Multi-level deep convolutional networks for automated pancreas segmentation," *MICCAI*, pp.556–564, 2015.
[24] M. Anthimopoulos, S. Christodoulidis, et al., "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, 2016.

AUTHOR PROFILE

Er. Jaspreet Kaur received her Diploma in Information Technology in 2012, B.Tech. in Information Technology in 2015, and M.Tech. in Computer Science in 2017. She is currently a Faculty Member in the Department of Computer Science at Guru Nanak National College for Women, Nakodar, with over seven years of teaching experience. Her research interests include Image Processing, Machine Learning, Data Mining, and Artificial Intelligence. She has published several research papers in reputed international journals and conferences, including IEEE.

