SCIENTIFIC CULTURE, Vol. 10, No. 1, (2024), pp 139-148

ADVANCING IMAGE RETRIEVAL: UNITING ATTENTION-POWERED CONVNETS WITH SIFT FEATURES

L. Anish*[1] and S. Thiyagarajan [2]

1. Research Scholar, Department of Computer Science, St Joseph University Chumukedima Nagaland India.
2. Professor, Department of Computer Science, St Joseph University Chumukedima Nagaland India.

Corresponding author: L. Anish

E-mail: anishlazer2013@gmail.com

Abstract:

The increasing need for image retrieval from multimedia databases has made Content-Based Image Retrieval (CBIR) a crucial field in the last ten years. Developing a system for CBIR is not an easy task. This article proposed a new technique for CBIR founded on salient regions and deep learning. This research introduces a novel approach employing theAttention-Enhanced Convolutional Neural Networks (Attention-Enhanced CNN) model. The methodology involves collecting the CBIR image dataset and it was preprocessing through Histogram equalization (HE) to enhance image quality, following preprocessing, feature extraction is performed using Scale-Invariant Feature Transform (SIFT) to capture intricate patterns and textures in the images. The classification step utilizes the Attention-Enhanced CNN. This approach is implemented and tested through simulations, and the results indicate a substantial positive deviation in the performance and retrieval of the images effectively compared to existing methods. The performance metrics are Accuracy, recall, precision, APR, F1-Score, ARR, MAP, FPR, and Error showing the measurements of this proposed model.

Keywords: Scale-Invariant Feature Transform (SIFT), Histogram equalization (HE), Attention-Enhanced CNN, CBIR system

1. Introduction

The recent tremendous rise of digital image storage has directed the expansion of retrieval systems of images. Users may search, retrieve, and discover information from a vast digital collection of images with the aid of retrieval image systems. Metadata addition for images such as descriptions, titles, keywords, and captions is a frequent practice in conventional image search methods [1]. Conventional methods have had the main impact on web-based image annotation applications, although social web apps and the semantic web have emerged. Meta-search of Image, exploration of image collection, and CBIR are some methods for image retrieval search [2]. Rather than using textual descriptions, CBIR uses user-specified image attributes to assess an image's shapes, textures, and colors to the query image (QI). To assess how complex the architecture of the image search system is necessary to ascertain the

amount and kind of image data. Two other factors that impact design are the expected volume of users on a search engine and the variety of the user population [3]. Key pointidentification, orientation assignment and descriptor calculation are some of the stages that go into extracting SIFT features. In an image, key points are specific regions that remain consistent and distinct despite changes. The descriptors' invariance of image rotation is guaranteed via orientation assignment. At last, descriptors are generated by using the local gradient data in every key point, therefore identifying distinct patterns in the image[4]. CBIR is a system for recovering imagesfound on optical information such as color, texture, and structure. Reducing the necessity for human participation in the indexing process while increasing the effectiveness of image indexing and retrieval was the basic purpose of CBIR [5]. The degree to which the computations and images are comparable determines the image retrieval method performs. A distinct, reliable, swift approach needs to be able to determine how similar the two images are in a perfect environment.An effective CBIR for histology could be without supervision, exact, and rapid, it capable of serving an immense collection of data [6]. The goal of this paper propose a new method Attention-Enhanced CNN for CBIR based on salient regions and deep learning.

2. Related work

Appearances were generally divided into two categories: local and global. A hybrid CBIR technique with two search levels is developed as an outcome of the study's use of both global and local information [7]. The experimental results showed the importance of examining the scheme of two-layer in improving accuracy when compared to contemporary approaches. A hybrid CBIR technique with two layers of filtering was produced as a consequence of the research [8] which mandated the use of both global and local characteristics. The returned images andQIfrom the main level were compared at the next level. Comprehensive assessments are performed using the extensively employed and well-regarded (Corel-1 k) dataset. The CBIR system was created by [9] using the support vector machine and genetic algorithm classifier to explain mixed features in image retrieval in a multi-class scenario. They used the first 3 color moments wavelets(Bi-Orthogonal, Daubechies,and Haar) to extract features.The CBIR methods covered in [10] can be divided into three categories: shape, texture, and color aspects. The comparative analysis of these three aspects and integrated characteristics in terms of several factors covered. Response time, recall, and accuracy were the requirements.The extraction of important characteristics from an image database and their storage in feature vectors were shown in [11]. Features such as texture, form and color signature are included in the repository. QI characteristics and database images were subjected to an inventive similarity assessment utilizing a metaheuristic approach (GA with simulated annealing). The primary concept of CBIR was to find the identical images for an image provided as QI by using distance measurements.The two visual feature descriptor extraction techniques investigated in [12] were SIFT and Oriented Fast and Rotated BRIEF (ORB). Whereas SIFT analyzes images according to orientation and size, ORB employs fast key points. The K-means clustering method was used to determine the mean of each cluster, and locality-preserving extension reduction technique lowers the feature vector dimensions.

3. Methodology

In this section, we have formed the Attention-Enhanced ConvNets model for image retrieval. Initially, CBIR image datasets are gathered and HE is employed for preprocessing the data. The SIFT approach is used to extract the data features. Figure 1 depicts the overall process of methodology.

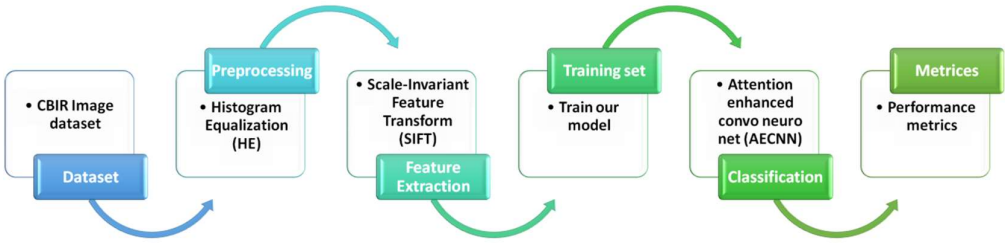

Figure 1: Overview of the study

3.1. Dataset

In this part, we gather a dataset from the open source of the Kaggle website https://www.kaggle.com/datasets/theaayushbajaj/cbir-dataset. It contains 4738 images, and we implemented our research only for 4737 images.

3.2. Preprocessing using histogram equalization (HE)

The following data collection preprocess step was utilized. One method used in image processing to modify contrast is called HE. It accomplishes a even distribution of contrast throughout the histogram, enabling areas with less local contrast to show stronger contrast. Because it highlights the greatest contrast levels, this approach dramatically increases contrast. HE is especially useful for images with a black-and-white focus and history like medical images. In image processing; creating a severity histogram is another histogram-based technique. Different properties such as average, variance, skewness, elongation, entropy, and energy are taken into consideration in this kind of histogram. When the image was dark, the histogram was biased towards the lower end of the grayscale, with the image data condensed in the histogram. It proved feasible to change the shades of gray to be more intense at the shaded end, which could improve the images' visibility and give the histogram's range to more equal distribution. The histogram of a computerized image with different levels of intensity is shown in Equation (1):

$$s_{(p_t)=r_t} \quad (1)$$

In this case, (p_t) stands for the rth intensity value and dr for the number of pixels in the image that have the provided level. Scaling histograms according to the entire amount of pixels in images was standard behavior. Equation (2) shows the correlation between the chance that cr will occur in $L * K$ images and a normalized histogram. The preprocessed image from the original image is shown in Figure 2.

$$A_{(p_t)} = \frac{r_t}{L*K} \qquad (2)$$

Figure 2: Preprocessed Image

In CBIR systems, HE helps with more efficient image preprocessing and retrieval by improving visual quality and contrast.

3.3. Feature extraction is performed by SIFT

SIFT was used for feature extraction after the preprocessing. One common CBIR approach for deriving valuable data from images is called SIFT. An enhanced image retrieval method is called SIFT. Input is an image, and output is a vector representation of the image features created using the SIFT method. The method extracts distinctive invariant properties. It indicates that scale, rotation, and perspective invariance apply to the recovered features. Image correspondence is a common problem in computer vision, including object or scene identification, spatial connection, tracking movement, and searching for a three-dimensional structure from many different images. The risk of obstruction, trash, or noise creating disruption is reduced since they are well-localized in the two domains of space. Several characteristics can be extracted from typical images using efficient methods. Moreover, anonly characteristic preserve precisely coordinated with a high prospect against animmense database of characteristics due to their significant differences, which paves the way for entity and image recognition.An essential element of the SIFT approach is the large number of features when it produces, and densely cover the image at all sizes and places. A (500 * 500) pixel image can typically provide around 2000 stable features. When it comes to object recognition, the quantity of features is particularly important since reliable detection of small things in crowded backgrounds depends on the correct corresponding of at least three qualities from each object.In CBIR systems, the SIFT method matches and finds local features removedin images to provide fast and precise similarity searches.

3.4. Advancing Image Retrieval using Attention-Enhanced CNN

After extracting the features, color, texture, and existing forms of the image are the main descriptors in the CBIR system. In CBIR systems, Convolutional Neural Networks (CNNs) can be employed efficiently. CBIR systems are made to retrieve images from a database without the need for written descriptions or metadata, depending on the illustrationsatisfied of the image. The capacity of CNNs to extract hierarchical characteristics from images has revolutionized the region of computer vision, and this ability makes CNNs ideal for CBIR tasks. Primary descriptors can be used to find and retrieve comparable images

from anenormous image set. Because of the quantity of the collection, manually extracting images from its difficult. To improve the classification accuracy, we suggest using the attention-enhanced CNN.To train and evaluate the Attention Enhanced CNN model, an extensive image dataset has to be gathered in the first stage. Representative samples from the target domain can be included in the dataset. The images follows with a standard size adjustment and their pixel values are normalized as part of pre-processing.Image saliencyis one of the features recognized and that are identified using the attention network component of the proposed methodology. The saliency values in this research are obtained without the need for further training or weighting. The image preprocessing layer performs some of that processing. Equation (3) represents the saliency map that the suggested framework would produce if this section is dubbed Attent. Equation (4) is used to compute the hashing network entry, to proceed.

$$saliencyregions = Attent(x_{previous}) \quad (3)$$

$$x_{current} = saliencypoints \odot x_{previous} \quad (4)$$

The lowest layers of CNN are the max pooling and convolution layers, whereas the top levels, fully linked layers, are similar to the classic MLP (Multi-layer Perceptron). MLP combines logistic regression with hidden layers. The collection of 4D features that the bottom layer operates the input to the first completely linked higher layer. These features are flattened into a 2D matrix of resized feature maps. Figure 3 depicts the attention component of the suggested end-to-end structure in detail. The encoder component consists of two layers for max-pooling, five elu layers, five sequential normalization layers, and 5 layers for convolution.

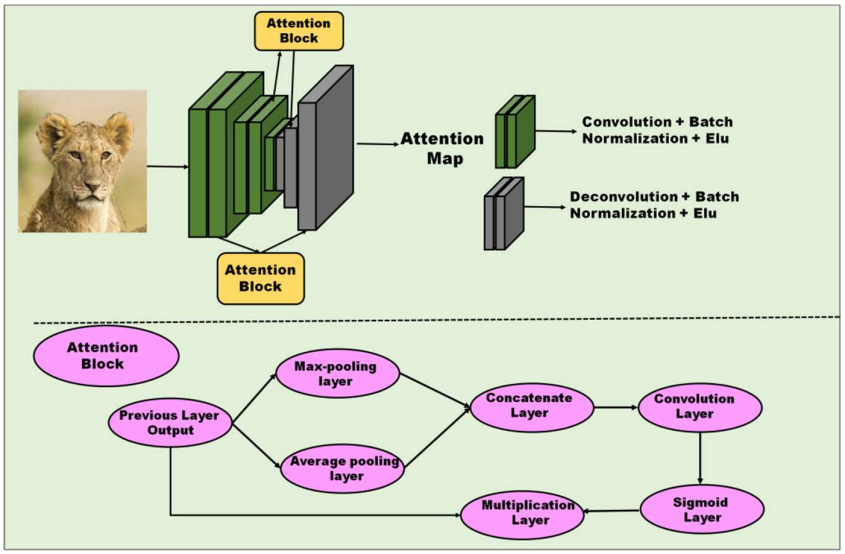

Figure 3: Structure of Attention Enhanced CNN

Each convolutional layer has a pixel size of 5 × 5 and a depth that ranges from 32 to 128. Three deconvolution levels, two batches of normalization layers, and two elu layers make up the decoding part. Two attention blocks in the attention section share the identical layers. It

can safeguard standard low-frequency information due to its unique mode of communication.To increase retrieval efficiency, an attention-enhanced CNN model for CBIR systems that are improved for attention dynamically focuses on salient regions in images.

4. Result

In this part, the following metrics are considered to assess models' performances for advancing image retrieval.The test image is compared with the expected categories. To examine the proposed attention-enhanced CNN-based model's categorization and retrieval performance, we conducted experiments using metrics.

The ratio of properly predicted images to the entire amount of input images is used to calculate the classification accuracy, and it is provided by Equation (5).The precision of a classifier can be evaluated by separating the entire amount of properly categorized predictions by the amount of improperly classed predictions, using Equation (6). The recall is calculated using Equation (7) and is distinct as the quantity of accurate constructive outcomes divided by the entire amount of relevant samples.According to Equation (8), the F1 score is a consistent indicator of a classifier's accuracy (the number of cases it properly identifies) and robustness (not ignoring an important amount of images).When compared to other conventional distance measures Figure 4 and Table 1 show the outcomes. When compared to the suggested Attention Enhanced CNN method with other existing methods, it achieved higher values such as accuracy (98%), F1-Score (97%), precision (97%), and recall (96%).

$$Accuracy = \frac{TruePositives + TrueNegatives}{TruePositives + FalsePositives + TrueNegatives + FalseNegatives} \quad (5)$$

$$Precision = \frac{severalamountsofretrievedimages}{wholenoofretrievedimag} \quad (6)$$

$$Recall = \frac{severalamountsofretrievedimagesand}{wholerelatedimagesinthedatabas} \quad (7)$$

$$F1\ score = \frac{2*Precision*Recall}{Precision+R} \quad (8)$$

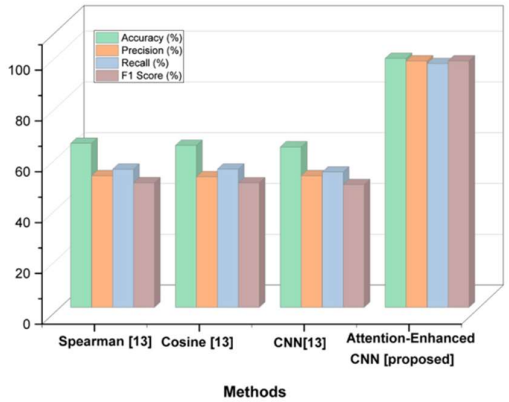

Figure 4: Comparisons of the methods

Table 1: Numerical outcomes

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Spearman [13]	64.56	51.84	54.37	49.04
Cosine [13]	63.75	51.4	54.42	49.01
CNN [13]	63.18	51.83	53.26	48.33
Attention-Enhanced CNN [Proposed]	98	97	96	97

The false positive rate (FPR), as expressed in Equation (9) is the part of negative sample images that are erroneously perceived as affirmative, compared to all unconstructive image samples. When compared to other conventional methods our suggested Attention Enhanced CNN achieved (0.001) lower features. Figure 5 and Table 2 comparisons of the FPR.

$$\text{FPR} = \frac{\text{FP}}{\text{FP+TN}} \qquad (9)$$

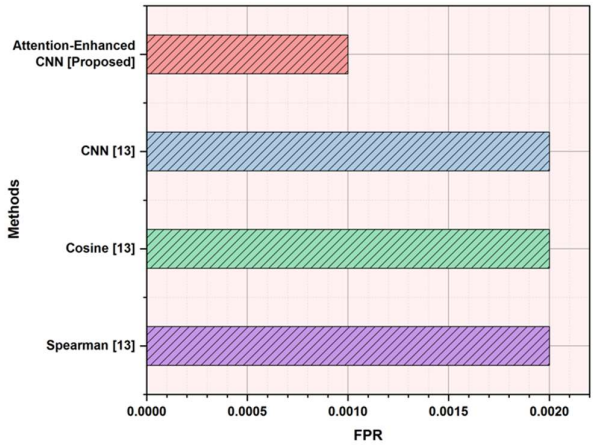

Figure 5: Comparisons of FPR

Table 2: Outcomes of FPR

Methods	FPR
Spearman [13]	0.002
Cosine [13]	0.002
CNN [13]	0.002

Attention-Enhanced CNN [Proposed]	0.001

Table 3 and Figure 6 present the Attention-Enhanced CNN model's performance as compared to several other studies using the estimated error score. The results demonstrate that the suggested model performed lower than the most advanced methods.

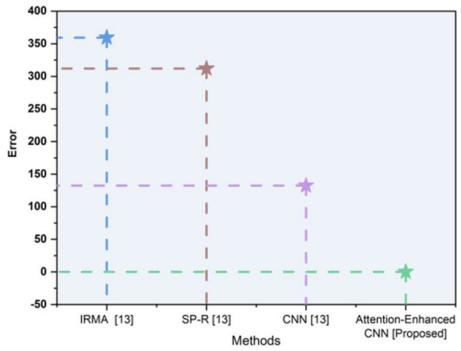

Figure 6: Comparisons of error rate

Table 3: Outcomes of error rate

Methods	Error
IRMA [13]	359.29
SP-R [13]	311.8
CNN [13]	132.45
Attention-Enhanced CNN [Proposed]	0.013

The percentage of pertinent items that are recovered from all of the relevant items that are accessible in the dataset is measured by the Average Rate of Recall (ARR). The recall values acquired at various cut-off points are averaged to compute it. The proportion of relevant components recovered to all relevant data in the dataset is known as recall. Average Rate of Precision (ARP) calculates the percentage of relevant images that are recovered out of all the objects that are retrieved up to a predetermined cut-off point. It is computed as the mean of the precision values acquired at various thresholds. The ratio of pertinent things recovered to all objects recovered is known as precision.Table 4 and Figure 7 depict the comparisons of the Attention-Enhanced CNN model's performance with Existing, when it compared our Attention-Enhanced CNN method achieved high values of APR (99%) and ARR (99%).

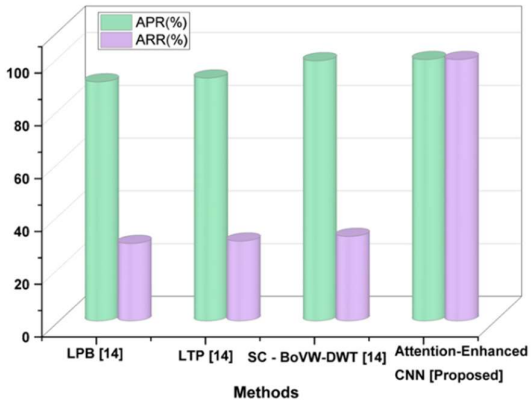

Figure 7: Comparisons of APR and ARR

Table 4: Outcomes of APR and ARR

Methods	APR (%)	ARR (%)
LBP [14]	90.55	29.33
LTP [14]	92.0	30.23
SC-BoVW-DWT [14]	98.54	32.01
Attention-Enhanced CNN [Proposed]	99.0	99.0

Mean Average accuracy (MAP) is the most used statistic for assessing retrieval systems' performance. The metric computes the order of the precisely selected outcomes, which is determined by the following Equation (8).We examined images in the dataset to assess the retrieval performance. It was found that while the test images weren't the most accurate match for some classes, the retrieval outcomes for the remainder were extremely high quality. This is due to the dataset's notable class imbalance, which makes feature learning challenging. Table 5 and Figure 8 present the findings. When comparing the suggested method Attention Enhanced CNN achieved (99%) with other existing methods DVB achieved (59.5%), SPQ achieved (78.5%) and Autoret achieved (80.1%), it shows that Attention Enhanced CNN achieved a higher MAP value.

$$MAP = \frac{\sum_{q=1}^{Q} avgP(q)}{Q} \qquad (10)$$

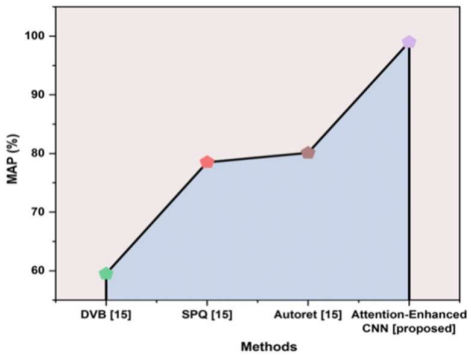

Figure 8: Comparisons of MAP

Table 5: Outcomes of MAP

Methods	MAP (%)
DVB [15]	59.5
SPQ [15]	78.5
Autoret [15]	80.1
Attention-Enhanced CNN [Proposed]	99.0

5. Conclusion

This research uses deep learning algorithms and salient regions to provide a novel strategy for CBIR. To extract features and extract complex patterns and textures, the technique first preprocesses the CBIR image dataset using HE to improve image quality. To improve retrieval performance even the classification stage makes use of Attention-Enhanced CNN. The suggested methodology integrates the Attention-Enhanced CNN model, exhibiting notable improvements in image retrieval accuracy in contrast to current techniques. After conducting rigorous testing and simulations, the findings demonstratesignificant improvements in Attention Enhanced CNN values in several performance indicators, including accuracy (98%), recall (96%), MAP (99%), precision (97%), APR (99%), F1-Score (97%), ARR (99%), FPR (0.001), and Error (0.013). This shows that the suggested model efficiently and accurately retrieves images, providing an idea for the development of CBIR systems.Further development of deep learning models to integrate multimodal data and contextual knowledge will enable more accurate and effective image retrieval in the future.

Abbreviation

IRMA- Image retrieval in medical applications

SP-R - Spatial Pyramid Representation

LBP -Local Binary Pattern

LTP - Local Ternary Pattern

SC-BoVW-DWT- Scattering Coefficients - Bag of Visual Words – Discrete Wavelet Transform

DVB – Deep Variational Binaries

SPQ - Self-supervised Product Quantization

Reference

1. Li, X., Yang, J., & Ma, J. (2021). Recent developments of content-based image retrieval (CBIR). *Neurocomputing*, *452*, 675-689.https://doi.org/10.1016/j.neucom.2020.07.139
2. Madduri, A. (2021). Content-based Image Retrieval System using Local Feature Extraction Techniques. *International Journal of Computer Applications*, *183*(20), 16-20.
3. Choe, J., Hwang, H. J., Seo, J. B., Lee, S. M., Yun, J., Kim, M. J., ... & Kim, B. (2022). Content-based image retrieval by using deep learning for interstitial lung disease diagnosis with chest CT. *Radiology*, *302*(1), 187-197.
4. Mohagheghi, S., Alizadeh, M., Safavi, S. M., Foruzan, A. H., & Chen, Y. W. (2021). Integration of CNN, CBMIR, and visualization techniques for diagnosis and quantification of COVID-19 disease. *IEEE Journal of Biomedical and Health Informatics*, *25*(6), 1873-1880. https://doi.org/10.1109/JBHI.2021.3067333
5. Garg, M., & Dhiman, G. (2021). A novel content-based image retrieval approach for classification using GLCM features and texture-fused LBP variants. *Neural Computing and Applications*, *33*(4), 1311-1328. https://doi.org/10.1007/s00521-020-05017-z
6. Dubey, S. R. (2021). A decade survey of content-based image retrieval using deep learning. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(5), 2687-2704. https://doi.org/10.1109/TCSVT.2021.3080920
7. Salih, S. F., & Abdulla, A. A. (2023). An effective bi-layer content-based image retrieval technique. *The Journal of Supercomputing*, *79*(2), 2308-2331.https://doi.org/10.1007/s11227-022-04748-1
8. Salih, F. A. A., & Abdulla, A. A. (2023). Two-layer content-based image retrieval technique for improving effectiveness. *Multimedia Tools and Applications*, *82*(20), 31423-31444.https://doi.org/10.1007/s11042-023-14678-6
9. Khan, U. A., Javed, A., & Ashraf, R. (2021). An effective hybrid framework for content-based image retrieval (CBIR). *Multimedia Tools and Applications*, *80*(17), 26911-26937.https://doi.org/10.1007/s11042-021-10530-x
10. Shukran, M. A. M., Malaysia, U. P. N., Malaysia, U. P. N., & Malaysia, U. P. N. (2021). A new approach to the techniques of content-based image retrieval (CBIR) using color, texture, and shape features. *Journal of Materials Science and Chemical Engineering*, *9*(01), 51. https://doi.org/10.4236/msce.2021.91005

11. Alsmadi, M. K. (2020). Content-based image retrieval using color, shape, and texture descriptors and features. *Arabian Journal for Science and Engineering*, *45*(4), 3317-3330.https://doi.org/10.1007/s13369-020-04384-y
12. Chhabra, P., Garg, N. K., & Kumar, M. (2020). Content-based image retrieval system using ORB and SIFT features. *Neural Computing and Applications*, *32*(7), 2725-2733. https://doi.org/10.1007/s00521-018-3677-9
13. Karthik, K., & Kamath, S. S. (2021). A deep neural network model for content-based medical image retrieval with multi-view classification. *The Visual Computer*, *37*(7), 1837-1850.https://doi.org/10.1007/s00371-020-01941-2
14. Rao, R. V., & Prasad, T. J. C. (2021). Content-based medical image retrieval using a novel hybrid scattering coefficients-bag of visual words-DWT relevance fusion. *Multimedia Tools and Applications*, *80*(8), 11815-11841.https://doi.org/10.1007/s11042-020-10415-5
15. Monowar, M. M., Hamid, M. A., Ohi, A. Q., Alassafi, M. O., & Mridha, M. F. (2022). AutoRet: A self-supervised spatial recurrent network for content-based image retrieval. *Sensors*, *22*(6), 2188.https://doi.org/10.3390/s22062188

