INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

OPEN ACCESS

Received: 13-03-2024

Accepted: 03-06-2024

Published: 19-06-2024

Citation: Anish L, Thiyagarajan S (2024) Unveiling Visual Treasures: Harnessing Deep Learning for Content-Based Image Retrieval. Indian Journal of Science and Technology 17(25): 2610-2621. https://doi.org/10.17485/IJST/v17i25.745

* **Corresponding author**.

anishlazer2013@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Unveiling Visual Treasures: Harnessing Deep Learning for Content-Based Image Retrieval

L Anish[1]*, S Thiyagarajan[2]

1 Research Scholar, Department of Computer Science, St Joseph University, Chümoukedima, Nagaland, India

2 Professor, Department of Computer Science, St Joseph University, Chümoukedima, Nagaland, India

Abstract

Objective: An essential aspect of computer vision is content-based image retrieval (CBIR), which enables users to search for images based on their visual content instead of created annotations. Advances in technology have resulted in a significant rise in the complexity of multimedia content and the emergence of new research fields centered on similar multimedia material retrieval. The efficacy of retrieval is impacted by the limits of the present CBIR systems, which result from overlooked algorithms and computing restrictions. **Methods:** This research introduces a novel approach employing the Siamese Edge Attention Layered Convonet (SEAL Convonet) for Image Retrieval. We utilize the CBIR image dataset through Gaussian smoothing to enhance image quality for data preprocessing and the Canny Edge Detector (CED) for edge detection, following pre-processing. The Histogram of Oriented Gradients (HOG) is used for feature extraction to extract complex textures and patterns from the images. **Findings:** This approach is implemented and tested through simulations as well as the results indicate a substantial positive deviation in the performance and retrieval of the images compared to existing methods. The performance metrics are accuracy (97 %), precision (94 %), recall (91 %), F1-Score (97 %), False Positive Rate (FNR) (0.0013), Matthew's correlation coefficient (MCC) (0.85), and False Negative Rate (FPR) (0.0036) show the measurements of this proposed model. **Application:** The state of the art in this work is researching the influence of optimizers on the accuracy process, as indicated by the findings.

Keywords: CBIR; Gaussian smoothing; Canny Edge Detector; Histogram of oriented gradients (HOG); Siamese Edge Attention Layered convonet (SEAL Convonet); Database

1 Introduction

In multiple divisions of machine vision and artificial intelligence (AI), the Content-Based Image Retrieval (CBIR) approach is used to obtain appropriate images from a database using an input image of the object or content that with interest. CBIR systems evaluate visual features of images, such as color, texture, shape, and spatial arrangement, to conduct searches, in contrast to conventional text-based retrieval systems [1]. It is particularly beneficial in circumstances such as digital libraries, medical imaging, surveillance, and e-commerce as verbal explanations of images might be insufficient or difficult. Numerous crucial phases are usually included in CBIR systems. Visual material in images is analyzed and numerically represented using feature extraction methods [2]. The CBIR technique offers an automatic retrieval mechanism that enables the users to do searches again with updated information is a significant advantage over keyword-based alternatives that usually require tedious and time-consuming annotation of database images. One of the key elements that ascertain the amount the CBIR structure performs on the basis of accuracy and speed is image descriptor [3]. The visual aspects of the images were frequently given more weight than written annotation. After text was added to the images, a text-based search method derived from conventional database management systems was employed. The vast amount of visual content that people from many languages and geographical locations have contributed has possibly minimal Metadata at all or Metadata in a number of languages. [4]. Text-based image retrieval systems require user annotation due to the difficulty in generating descriptive phrases for large archives. CBIR represents and classifies images according to visual elements that address the semantic gap between user information requests and image representations in CBIR systems [5]. Integrating human perception throughout the search request that involves the user in assessing the retrieval results is the fundamental notion behind relevance feedback. The restrictions of metadata-based approaches and the numerous applications for effective image retrieval have attracted greater focus from CBIR [6].

The deep metric learning [7] approach's capacity to effectively incorporate in-class variation can be limited when there is an excessive emphasis on categorization resulting in a reduction of latent generative components. Study [8] employed the C-means clustering technique that reliance on maintained feature weighting and grouping can result in increasing issues with big datasets. The resilience and efficacy of the Deep CNN framework under real-world circumstances can be affected by the model [9]. By effectively enhancing routing in wireless sensor networks, our proposed SEAL Convonet approaches provided advantages over the existing ones. CBIR research incorporates human-centered design and user-friendly interfaces to improve search efficiency. However, traditional systems use basic visual traits, leading to mismatches and inaccuracies. Large images also pose scalability issues, resulting in longer retrieval times. This paper proposes a method using SEAL Convonet to extract high-level semantic features from images, improving image retrieval accuracy and exploring CBIR techniques. Figure 1 illustrates the block diagram of fundamental of CBIR system.

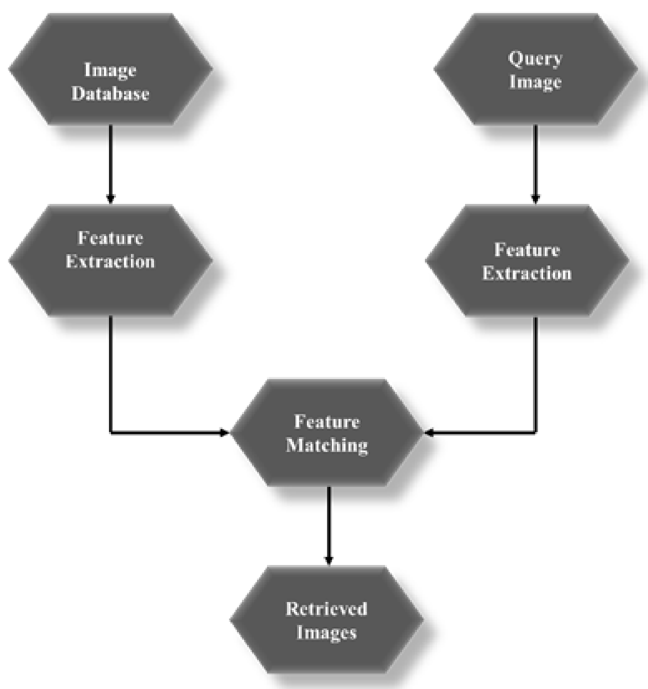

Fig 1. CBIR system

1.1 Contribution of the study

The approach employs Gaussian smoothing, Canny Edge Detector, HOG for feature extraction, and SEAL Convonet for categorization, improving image quality and capturing complex patterns and surfaces. Models are employed to test the SEAL convonet model, and the findings demonstrate that it is both additional efficient and better to different techniques now in implementation for image retrieval uses. The investigation demonstrates significant improvements in recovery efficiency whenever contrasted to other different techniques, as exhibited through a comprehensive collection of effectiveness measures.

- The approach utilizes Gaussian smoothing for pre-processing, HOG for feature retrieval, and SEAL Convonet for classification, improving image quality and obtaining complex configurations and textures.
- The SEAL convonet Framework is evaluated via models that substantially exceed different Current techniques in image retrieval tasks, suggesting its outstanding effectiveness and Supremacy.
- The Investigation demonstrates significant advancements in retrieval effectiveness when contrasted to other current techniques, as shown through a set of capability metrics.
- Through utilizing the Siamese Edge Attention Layered Convonet (SEAL Convonet), Unveiling Visual Treasures shows a substantial advancement in CBIR.
- In a diversity of industries, such as online commerce, medical care, and safety, enhanced image retrieval dependability and efficiency are crucial. It Investigates essential image recognition problems through using deep learning methods, offering feasible results for application.

Paper[10] presented a novel approach inefficient to sparse complementary features for dynamic image representation, locality-preserving extension for optimal feature selection, and fuzzy c-means clustering. Study[7] introduced a CBIR system in this work that utilizes transfer learning techniques from a convolutional neural network (CNN) educated on an extensive image database, allowing for the use of generalized image description as it was originally acquired. Paper[8] proposed a regularized discriminative deep metric learning approach that overcomes this limitation by learning a representation that permits discrimination between classes but encodes the suppressed factor independently in every class. Research[9] proposed a distinctive feature-weighting-based learning technique to enhance the effectiveness of systems for classifying or retrieving images in a multi-label setting. Study[11] used chest X-ray images that suggest a Deep CNN(DCNN) method for the detection of lung pneumonia infection. Utilizing the content-based image retrieval technique, the collection images were tagged with additional content and information. Paper[12] examined dictionary learning (DL) tasks that involve CBIR and are suitable for using sparse vectors. The CNN architecture yields a vector for each image, which was utilized as the DL input. Paper[13] suggested a secure cloud-based CBIR architecture that retrieves images without requiring user input. Utilizing a standard DNN algorithm that has been satisfactorily educated the user extracts the image vectors of features.

To efficiently retrieve the images from the databases, the study introduced an innovative CBIR simulation utilizing the Deep Learning-based Inception v3 Model (DLIM)[14]. The simulation outcomes demonstrated the effectiveness of the DLIM approach during retrieval. Utilizing CBIR[15], users can search an image database for images with comparable visual content to a given query image. The article discusses the use of machine learning and image processing in evaluating and categorizing images using CBIR algorithms[16]. Paper[17] explored the deep learning developments in content-based image retrieval over the past decade, categorizing modern techniques and analyzing performance using traditional methods. The research[18] provides a comprehensive overview of CDIR techniques and applications, including their use in clothing, infrared, remote sensing, and sketching. The use of DL[19] in a CBIR technique for chest CT scans significantly improved diagnosis reliability and inter-reader consistency among readers with varying knowledge levels. Paper[20] proposed an innovative approach for image retrieval using color and texture features, using an extended version of local neighborhood difference patterns (ELNDP) and optimized color histogram features. The approach outperformed existing approaches regarding of recall and precision scores. Article[21] proposed an efficient query-sensitive co-attention mechanism for large-scale Content-based Image Retrieval (CBIR) task. The method used clustering of local feature to reduce computation cost and improved retrieval results under challenging situations. Author[22] advanced in digital data collection and storage have made large image datasets easier. Image processing was a rapidly growing technology in computer science. Content-based image retrieval reduced ambiguity and time, with a model displaying image with minimum vector distance. Research[23] explored Content-based Image Retrieval (CBIR), a method for efficiently retrieving images from large collections. It examined feature selection, extraction, and representation, incorporating deep learning techniques and researched 215 articles and highlights future research directions for CBIR's evolution. Author[24] introduced a content-based image retrieval model that combines low-level and mid-level features. The model's performance on various datasets was evaluated, with the best-evaluated results showing 99.4% precision. The model's explainability and extracted features' value were also interpreted using the Shapley value.

1.2 Research Gap

CBIR retrieves images from databases based on visual content, overcoming challenges like feature extraction, image appearance variations, and efficient search algorithms for large datasets. It includes semantic gaps, feature extraction complexity among high-level semantics low-level features scalability problems with large datasets, and subjectivity in relevance assessment. To overcome this problem, we proposed SEAL Convonet to enhance image quality and capture intricate patterns and textures.

2 Methodology

The Deep learning (DL) based CBIR system that has been suggested comprises. For every image in the database, the system uses Histogram of Oriented Gradients (HOG) techniques to extract feature vectors. The dataset consists of 4738 images from each class that are labeled, and a DL classifier called SEAL convonet has been taught to anticipate every attribute of vector class designation. Figure 2displays the flow of the suggested method.

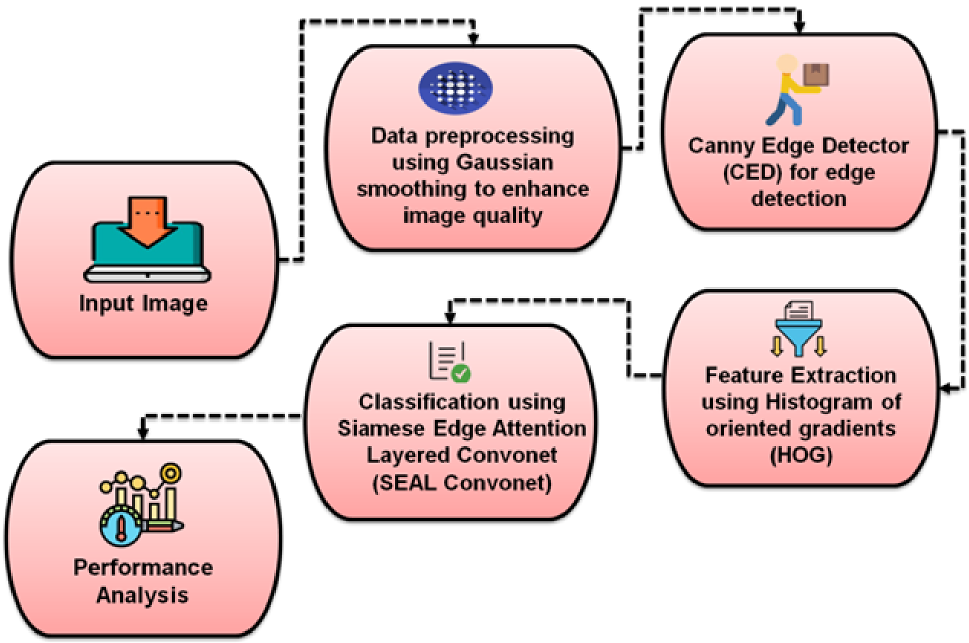

Fig 2. Block diagram for the suggested approach

2.1 Dataset

In this part, using 4738 image content features, CBIR searches and retrieves digital images from a sizable database. We gathered the dataset from (https://www.kaggle.com/datasets/theaayushbajaj/cbir-dataset). Figure 3 displays images of an animal database.

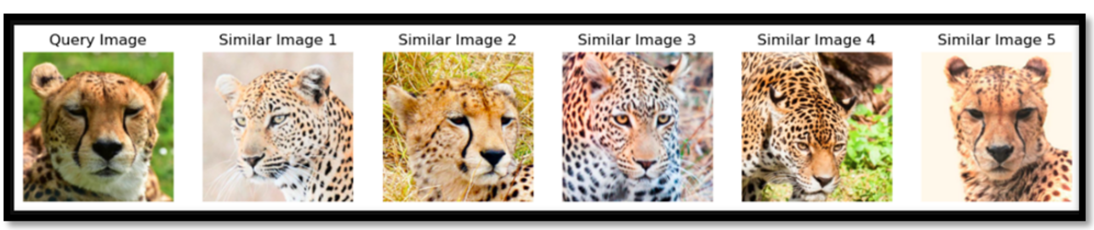

Fig 3. Retrieval of animal images

2.2 Pre-Processing using Gaussian Smoothing

Gaussian smoothing is a widely used method in image processing and computer vision to reduce noise and emphasize key aspects by using a 2D kernel to convolve an image. The process increases surrounding pixels by the kernel resulting in a weighted average. Gaussian smoothing is a low-pass filter that removes high-frequency noise from images, preserving the structure and aiding in edge detection and feature extraction. It reduces noise and noise-affecting gradients.

2.3 Edge detection using Canny Edge Detector (CED)

CED uses Gaussian kernel linear filtering to compute edge strength and direction for each pixel, smoothing noise. Smoothing the image is the initial stage of the canny algorithm. The ideal edge detection operator may be approximated to the best degree by Canny, who also calculated the first derivative of Gaussian functions. Select the suitable one-dimensional Gaussian function to smooth the image based on the row and column, respectively. This involves performing a convolution operation to create an image matrix.

$$H(w,z) = exp\left[-\frac{w^2+z^2}{2\sigma^2}\right]/2\pi\sigma^2 \tag{1}$$

Where $\sigma$is the Gauss filter parameter, which governs the smoothing image extension. The pixel that wasn't considered to be a part of an edge is eliminated in this manner.

2.4 Feature extraction using Histogram of oriented gradients (HOG)

To identify and locate objects, computer vision professionals frequently utilize the feature extraction approach known as HOG.HOG extracts local gradient information from images, transforming it into a feature array. It accurately represents the local gradient allocation structure and aids in target recognition. HOG's use of gradients and edge information in animations suggests data mining techniques that could influence content comprehension and motion tracking. The knowledge acquired from this research might be applied to create or impact animations by dynamically changing certain elements of the animation in response to CBIR that have been gathered. Individuals can calculate the intensity gradients in the A and B orientations using the following equations.

$$G_u = m(U+1,x) - I(u-1,x) \tag{2}$$

$$g_x = u(u,x+1) - I(u,x-1) \tag{3}$$

Where $G_u$and $g_x$is the gradients that are vertical and horizontal, respectively, the amplitude of the gradient$m(u,x)$ shows the difference in gray levels size. The gradient's size and direction are determined using Equations (4) and (5):

$$m(u,x) = \sqrt{G_u^2 + G_u^2} \tag{4}$$

$$\theta(u,x) = arctan\left(\frac{G_x}{G_x}\right) \tag{5}$$

2.5 Siamese Edge Attention Layered Convonet (SEAL Convonet)

An innovative architecture for content-based image retrieval tasks is the SEAL Convonet. SEAL Convonet optimizes feature extraction and matching capabilities by utilizing Siamese networks, edge attention techniques, and convolutional layers. This enables effective comparisons of image similarity. It can accurately retrieve appropriate images from extensive data sets, which improves a variety of image search and recognition applications. The SEAL Convonet is novel in that it combines a layered convolutional structure, a cutting-edge consideration process, and the Siamese building design. The network is suitable for tasks like determining the similarity or dissimilarity of images because of its capacity to evaluate and compare different input samples. Figure 4 depicts the architecture for SEAL convonet for content-based image retrieval tasks.

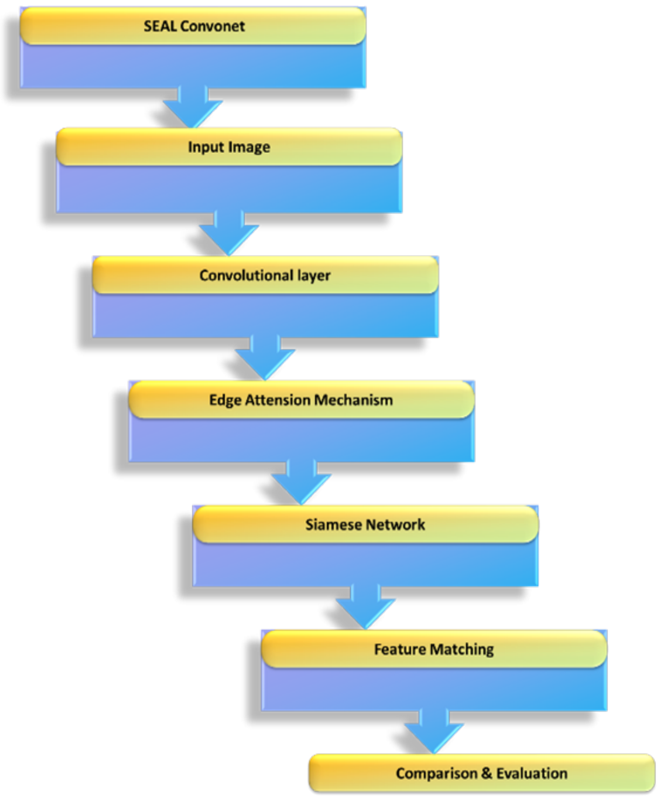

Fig 4. SEAL convonet for content-based image retrieval tasks

Siamese Networks

Siamese Networks belong to a category of networks that employ two identical sub-networks for one-shot classification. These sub-networks share the same setup, parameters, and weights while accommodating different inputs. Ranking losses are frequently applied to Siamese network topologies. Neural networks that operate with shared variables, or weights, are called Siamese networks. A single neural network analyses a variety of inputs while optimizing it throughout training. Two attribute vectors and an image make up each trio in the features retrieval sample of appearance images. As attribute vectors are processed by Siamese networks, both will be sent through the same network.

Edge Attention Mechanism

The term edge attention refers to the network's capacity to identify and focus on edges and contours in the input image. As edge characteristics are frequently essential for tasks like object identification, segmentation, and recognition, this attention mechanism may comprise specialized layers or processes designed to identify and emphasize edge features. The network's capacity to extract pertinent features and distinguish between various groups or categories in the input data can be improved by edge attention methods.

CNNs are layered networks with stacked layers for higher-level representations from input data. They mimic neural activity that is ideal for image problems. CNNs use biological concepts and CNN designs, with structure-based topologies and pooling layers. They often require a mini-batch technique for dealing with large training data. Let $G = [g_1, \ldots .g_N]$ consists of activation features from a set of N training images, where the dimensionality of the feature vector g_j is$R$, and $[g_1, \ldots .g_N], \in \{-1, 1\}$ saves the ground truth labels in a vector. A powerful classifier determines the forecast using the boosting technique. $Z(\cdot)$ is the total of all the weak classifiers weighted together $z(\cdot)$ as follows in Equations (6) and (7)

$$Z(g_1) = \sum_{i=1}^{R} \alpha_i z(g_{ji}, \lambda_i); \tag{6}$$

$$z\left(g_{ji},\lambda_i\right)=\frac{l\left(g_{ji},\lambda_i\right)}{\sqrt{l\left(g_{ji},\lambda_i\right)^2+\eta^2}} \tag{7}$$

where $y_{ji}\epsilon y_j$ is the i^{th} activation feature of j^{th}image. A potential poor classifier is associated with each attribute $z\left(g_{ji},\lambda_i\right)$ with output in the range of $(-1,1)$. $\frac{l(.)}{l(.)^2+\eta^2}$is used to simulate a sign $(\cdot)$method for optimizing gradient descent via derivative computation has been developed. The parameter η to manage the function's ramp $\frac{l(.)}{\sqrt{l(.)^2+\eta^2}}$ as well as is adjustable by the distribution of $l(\cdot)$ as$=\frac{\sigma}{v}$.

$$\sum_{i=1}^{R}\alpha_i=1 \tag{8}$$

When $\alpha_i=0$, because it is dormant, the associated neuron will not engage in feed-forward or backpropagation.

$$\varepsilon^P=\beta\varepsilon_{storng}^P+(1-\beta)\varepsilon_{weak} \tag{9}$$

where $\beta\in\ [0,\ 1]$ strikes a balance between the losses experienced by strong and weak classifiers.

$$\varepsilon_{storng}^P=\frac{1}{M}\sum_{j=1}^{M}\left(Z\left(y_j\right)-y_j\right)^2 \tag{10}$$

$$\varepsilon_{weak}=\frac{1}{NM}\sum_{j=1}^{M}\sum_{\alpha\_i=0}^{1\le i\le R}\left[z(g_{ji},\ \lambda_i)-x_j\right]^2 \tag{11}$$

The neural network is well-suited for tasks like measuring image similarities or differences since it has a Siamese design, which allows it to compare and contrast various input samples efficiently. This suggests that several convolutional layers form the foundation of network design.

3 Results and Discussion

The examination of the experimental outcome is done using the following parameters: F1-measure, recall, accuracy, precision, and FNR, MCC, and FPR. These variables are contrasted using two cutting-edge techniques; including the suggested SEAL Convonet, and other well-known methods such as "Rotation-Invariant Uniform Local Binary Patterns (RULBP), "Improved Local Leader-based Spider Monkey Optimization (ILL-SMO)", and Gray Level Co-occurrence Matrix with DL-based Enhanced Convolution Neural Network (GLCM-DLECNN)".

3.1 Accuracy and precision

A machine learning model's accuracy may be measured by looking at the ratio of correctly projected negative consequences to correctly expected good outcomes, which represents the model's conclusion. True Positives (TP) are circumstances that the model correctly categorizes as positive. False Negatives (FN) are results that the algorithm displays as positive even if they are wrongly classified as negative. The Accuracy of the suggested approach, which has a SEAL convonet value of 97%, is greater than that of RULBP's 85.33% accuracy, ILL-SMO's 95% accuracy, and GLCM-DLECNN's 90.23% accuracy. By calculating the ratio of properly recovered images towards the entirety quantity of images in the data collection, the accuracy of a core metric is used to assess the overall correctness of the CBIR system. It gives a sense of how well the system can reliably obtain pertinent images. Figure 5 displays the corresponding outcomes for the accuracy measurements. Compared to other current methods, the proposed procedure offers a high level of accuracy.

The precision quantifies the degree of accuracy within the system outputs. This criterion demonstrates that an increase in the proportion of accurate outcomes relative to incorrect findings increases the operation's precision. The precision of the suggested approach, which has a SEAL convonet value of 94%, is greater than that of RULBP's 81% precision, ILL-SMO's 84% accuracy, and GLCM-DLECNN's 89.9% precision. By calculating the ratio of properly recovered images towards the entirety quantity of images in the data collection, the precision of a core metric is used to assess the overall correctness of the CBIR system. It gives a sense of how well the system can reliably obtain pertinent images. Figure 6 displays the corresponding outcomes for the precision measurements. Compared to other current methods, the proposed procedure offers a high level of Precision.

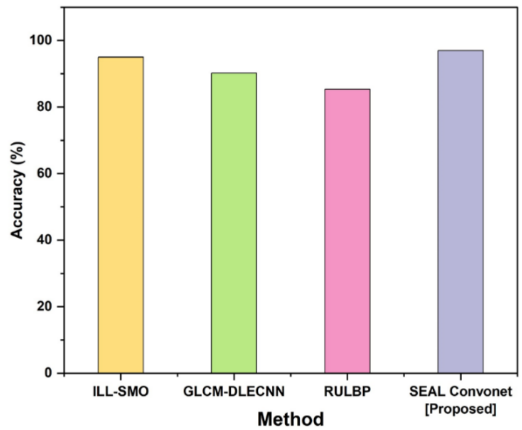

Fig 5. Comparison of accuracy

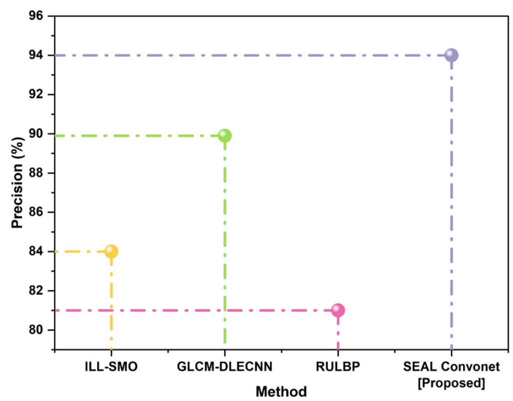

Fig 6. Comparison of precision

3.2 Recall and F1-score

The ratio of correctly assigned results to the base of all outcomes is referred to as recall. The incorrect diagnoses among the right results, an operation is said to have a greater recall value based on these characteristics. The SEAL Convonet significance of the recommended approach, which is 91%, advanced than other methods of ILL-SPO, GLCM-DLECNN, and RULBP, which attains 89%, 85%, and 79%, respectively, for recall. The system's capacity to retrieve all pertinent images from the database is evaluated using recall. It determines the proportion of appropriately retrieved pertinent images to all relevant images that are accessible as shown in Figure 7. The system's capacity to retrieve all pertinent images from the database is evaluated using recall. It determines the proportion of appropriately retrieved pertinent images to all relevant images that are accessible.

The F1 score is a machine learning metric that assesses a model's predictive power by evaluating its performance in each class separately, rather than as a whole. Precision and recall are two competing criteria that are used to get the F1 score. The F1-score findings are displayed in Figure 8. Compared to existing methods, our suggested approach is greater and F1-score. The SEAL Convonet significance of the recommended approach, which is 97%, higher than other methods of ILL-SPO, GLCM-DLECNN, and RULBP, which attains 79%, 87.8% and 80%, respectively, for F1-score. The system's capacity to retrieve all pertinent images from the database is evaluated using f1-score. It determines the proportion of appropriately retrieved pertinent images to all relevant images that are accessible.

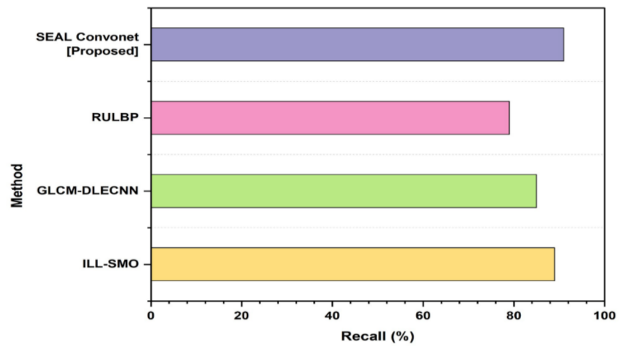

Fig 7. Comparison of Recall

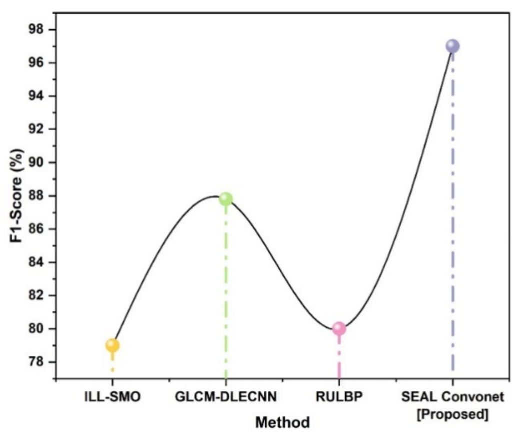

Fig 8. Comparison of F1 score

3.3 Matthew's correlation coefficient (MCC)

The MCC is a more reliable empirical value that always provides an excellent rating where the forecast accurately predicts the TN, FN, TP, and FP confusion matrix categories, proportionate to corresponding sizes. Retrieval in CBIR may be viewed as a binary classification of employment. The system divides every image in the database that corresponds to a query image into categories that are either relevant or irrelevant. The resultant MCC is displayed in Figure 9.

The SEAL Convonet value of the proposed method, which is 0.85, is higher than other methods of ILL-SPO, GLCM-DLECNN, and RULBP, which attains 0.77, 0.69 and 0.6, respectively, for MCC.

3.4 False Negative Rate (FNR) and False Positive Rate (FPR)

The fraction of true positive cases that the model mispredicted as negative is measured by FNR. A false negative, also known as a false negative mistake, is a test result that incorrectly suggests that a condition is incorrect. A high FNR suggests that a large number of pertinent images are missing from the system, which may cause user annoyance and decrease the retrieval system's usefulness. Divide the entire number of FP by the amount of TN and FP to get the FPR. FPR stands for false positive rate, or the percentage of real negative cases that the model mispredicted as positive. Using the number of FP and TN, the FP rate is computed. It is the likelihood of a false alarm happening, which means that a positive result will be reported even in the event of a negative real value.

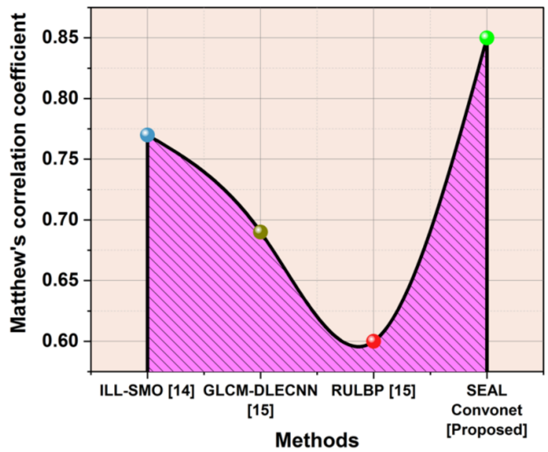

Fig 9. Comparison of MCC

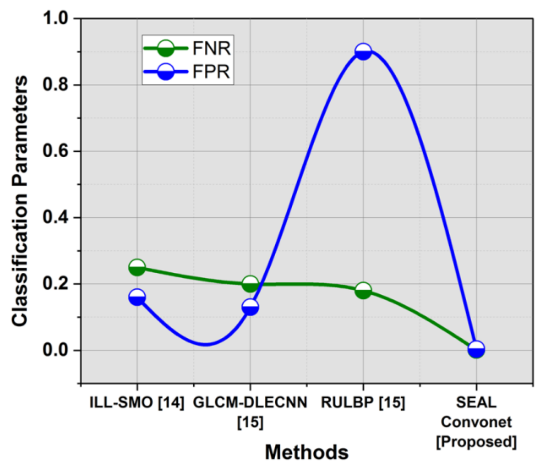

Fig 10. Outcome of FNR and FPR

Figure 10 shows the outcome of FPR and FNR. The SEAL Convonet value of the proposed method has achieved FNR (0.0013) and FPR (0.0036), lower than other methods of ILL-SPO, GLCM-DLECNN, and RULBP, which attains FNR0.25, 0.16, 0.2 and for FPR0.13, 0.18 and 0.9, respectively. Positive findings and appropriate performance were obtained by the proposed SEAL Convonet when compared with other current models. Table 1 displays the total results of the suggested approach.

Positive findings and appropriate performance were obtained by the proposed SEAL Convonet when compared with other current models. Table 1 displays the total results of the suggested approach.

Table 1. Overall outcome of the proposed method

Classification Parameters	Equation	SEAL Convonet [Proposed]
Precision (%)	$Precision = \frac{TP}{TP+FP}$	94 %
F1-score (%)	$F1-score = \frac{(precision)\times(recall)\times 2}{recall+precision}$	97 %
Accuracy (%)	$Accuracy = \frac{TN+TP}{TN+TP+FN+FP}$	97 %
Recall (%)	$Recall = \frac{TP}{TP+FN}$	91 %

Continued on next page

Table 1 continued

FNR	$FNR = \frac{FN}{TP+FN}$	0.0013
MCC	$MCC = \frac{(TP-TN)-(FNXFN)}{\sqrt{(TP+FN)(TN+FN)(TP+FP)(TN+FP)}}$	0.85
FPR	$FPR = \frac{FP}{TP+FP}$	0.0036

The ILL-SMO[25] issue, despite its complexity, faces challenges like low search accuracy, slow convergence time, and local optimality, requiring complex and time-consuming parameter selection. RULBP[26] promotes rotation invariance; its accuracy may not increase with increasing computing complexity due to binary pattern clarity, potentially causing inaccurate representations of fundamental visual data. GLCM-DLECNN[26], despite its computational complexity and accuracy issues, has the potential to enhance texture-based image retrieval, especially for color images with larger dimensionality. Its efficiency can be impacted by preprocessing biases. A loss of latent producing components caused by an excessive emphasis on classification can limit the ability of the deep metric learning[7] technique to process in-class variance. The C-means clustering approach that relies on maintained weighting of features and classification, could lead to greater issues with large datasets, as demonstrated by the study[8]. The model can have an impact on the Deep CNN framework's robustness and effectiveness in practical situations[9]. Compared to existing methods, our proposed SEAL Convonet techniques offered advantages by improving routing in wireless sensor networks.

4 Conclusion

Image retrieval offered promising avenues for enhancing retrieval accuracy and efficiency, CBIR systems frequently face drawbacks from neglected methods and computing limits that impair retrieval efficiency. CBIR is a method for quickly and effectively obtaining comparable images from a database, focusing on content rather than keywords. Through simulations, this SEAL convonet can be implemented into practice and evaluated; the findings show a significant improvement in accuracy and efficient image retrieval when compared to previous approaches. The measures of this suggested model are displayed by the performance metrics, which include accuracy (97 %), precision (94 %), recall (91 %), F1-Score (97 %), MCC (0.85), FNR (0.0013), and FPR (0.0036). The state of the art in this work is researching the influence of optimizers on the accuracy process, as indicated by the findings. SEAL Convonet is a prominent model that is used to check similarity. SEAL Convonet enhances Image Retrieval by leveraging attention mechanisms and deep convolutional layers, facilitating more accurate feature extraction and similarity assessment in image datasets, leading to improved retrieval performance. The drawback of SEAL Convonet is computational complexity, particularly during training, which demands significant resources and time due to the intricate attention mechanisms and multiple convolutional layers. Enhancement of the study presents improved performance metrics, more verification on various datasets and real-life situations could improve the stability and applicability of the proposed SEAL Convonet technique. Resolving issues with computing efficiency gained by the model's complexity can make it easier to utilize in real-world situations. To increase performance and retrieval accuracy on current approaches, the research presents SEAL Convonet, a revolutionary methodology that uses advanced methods. The dependence on sophisticated methods such as SEAL Convonet can result in computing complexity and possible difficulties when working with huge datasets or in real-time applications. Therefore, more research into efficient optimization methodologies is required. Future directions in cross-modal retrieval techniques will be investigated in image retrieval research, allowing for the integration of textual and visual data. Enhancing the interpretability of acquired models and exploring techniques for effectively managing big datasets will be beneficial. It is essential that image retrieval systems investigate adversarial resilience and domain adaptability.

References

1) Li X, Yang J, Ma J. Recent developments of content-based image retrieval (CBIR). *Neurocomputing*. 2021;452:675–689. Available from: https://doi.org/10.1016/j.neucom.2020.07.139.
2) Alsmadi MK. Content-Based Image Retrieval Using Color, Shape and Texture Descriptors and Features. *Arabian Journal for Science and Engineering*. 2020;45(4):3317–3330. Available from: https://doi.org/10.1007/s13369-020-04384-y.
3) Madduri A. Content based Image Retrieval System using Local Feature Extraction Techniques. *International Journal of Computer Applications*. 2021;183(20):16–20. Available from: https://www.ijcaonline.org/archives/volume183/number20/32039-2021921549/.
4) Singh S, Batra S. An efficient bi-layer content based image retrieval system. *Multimedia Tools and Applications*. 2020;79:17731–17759. Available from: https://doi.org/10.1007/s11042-019-08401-7.
5) Keisham N, Neelima A. Efficient content-based image retrieval using deep search and rescue algorithm. *Soft Computing*. 2022;26(4):1597–1616. Available from: https://doi.org/10.1007/s00500-021-06660-x.
6) Kashif M, Raja G, Shaukat F. An efficient content-based image retrieval system for the diagnosis of lung diseases. *Journal of digital imaging*. 2020;33(4):971–987. Available from: https://doi.org/10.1007/s10278-020-00338-w.

7) Passalis N, Iosifidis A, Gabbouj M, Tefas A. Variance-preserving deep metric learning for content-based image retrieval. *Pattern Recognition Letters*. 2020;131:8–14. Available from: https://doi.org/10.1016/j.patrec.2019.11.041.
8) Ghodratnama S, Moghaddam HA. Content-based image retrieval using feature weighting and C-means clustering in a multi-label classification framework. *Pattern Analysis and Applications*. 2021;24:1–10. Available from: https://doi.org/10.1007/s10044-020-00887-4.
9) Rajasenbagam T, Jeyanthi S, Pandian JA. Detection of pneumonia infection in lungs from chest X-ray images using deep convolutional neural network and content-based image retrieval techniques. *Journal of Ambient Intelligence and Humanized Computing*. 2021;p. 1–8. Available from: https://doi.org/10.1007/s12652-021-03075-2.
10) Bibi R, Mehmood Z, Yousaf RM, Saba T, Sardaraz M, Rehman A. Query-by-visual-search: multimodal framework for content-based image retrieval. *Journal of Ambient Intelligence and Humanized Computing*. 2020;11:5629–5648. Available from: https://doi.org/10.1007/s12652-020-01923-1.
11) Karthik K, Kamath SS. A deep neural network model for content-based medical image retrieval with multi-view classification. *A deep neural network model for content-based medical image retrieval with multi-view classification The Visual Computer*. 2021;37:1837–1850. Available from: https://doi.org/10.1007/s00371-020-01941-2.
12) Kumar RB, Marikkannu P. An Efficient Content Based Image Retrieval using an Optimized Neural Network for Medical Application. *Multimedia Tools and Applications*. 2020;79:22277–22292. Available from: https://doi.org/10.1007/s11042-020-08953-z.
13) Hassan A, Liu F, Wang F, Wang Y. Secure content based image retrieval for mobile users with deep neural networks in the cloud. *Journal of Systems Architecture*. 2021;116. Available from: https://doi.org/10.1016/j.sysarc.2021.102043.
14) Ranjith E, Parthiban L, Latchoumi TP, Kumar SA, Perera DG, Ramaswamy S. An Effective Content Based Image Retrieval System Using Deep Learning Based Inception Model. *Wireless Personal Communications* . 2023;133:811–829. Available from: https://doi.org/10.1007/s11277-023-10792-8.
15) Hameed IM, Abdulhussain SH, Mahmmod BM, Pham DT. Content-based image retrieval: A review of recent trends. *Cogent Engineering* . 2021;8(1):1–37. Available from: https://doi.org/10.1080/23311916.2021.1927469.
16) Jordan T, Elgazzar H. Content-Based Image Retrieval Using Deep Learning. In: Advances in Data Science and Information Engineering . Transactions on Computational Science and Computational Intelligence;Springer, Cham. 2021;p. 771–785. Available from: https://doi.org/10.1007/978-3-030-71704-9_56.
17) Dubey SR. A decade survey of content based image retrieval using deep learning. *IEEE Transactions on Circuits and Systems for Video Technology*. 2021;32(5):2687–2704. Available from: https://doi.org/10.1109/TCSVT.2021.3080920.
18) Zhou X, Han X, Li H, Wang J, Liang X. Cross-domain image retrieval: methods and applications. *International Journal of Multimedia Information Retrieval*. 2022;11(3):199–218. Available from: https://doi.org/10.1007/s13735-022-00244-7.
19) Choe J, Hwang HJ, Seo JB, Lee SM, Yun J, Kim MJ, et al. Content-based Image Retrieval by Using Deep Learning for Interstitial Lung Disease Diagnosis with Chest CT. *Radiology*. 2021;302(1):187–197. Available from: https://dx.doi.org/10.1148/radiol.2021204164.
20) Kelishadrokhi MK, Ghattaei M, Fekri-Ershad S. Innovative local texture descriptor in joint of human-based color features for content-based image retrieval. *Signal, Image and Video Processing*. 2023;17(8):4009–4017. Available from: https://dx.doi.org/10.1007/s11760-023-02631-x.
21) Hu Z, Bors AG. Co-attention enabled content-based image retrieval. *Neural Networks*. 2023;164:245–263. Available from: https://dx.doi.org/10.1016/j.neunet.2023.04.009.
22) Choudhary C, Nagra LS, Das P, Singh J, Jamwal SS. Optimized Ensemble Machine Learning Model for Chronic Kidney Disease Prediction. In: 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). IEEE. 2024;p. 292–297. Available from: https://doi.org/10.1109/ICCCIS60361.2023.10425073.
23) Srivastava D, Singh SS, Rajitha B, Verma M, Kaur M, Lee HN. Content-Based Image Retrieval: A Survey on Local and Global Features Selection, Extraction, Representation, and Evaluation Parameters. *IEEE Access*. 2023;11:95410–95431. Available from: https://dx.doi.org/10.1109/access.2023.3308911.
24) Taheri F, Rahbar K, Salimi P. Effective features in content-based image retrieval from a combination of low-level features and deep Boltzmann machine. *Multimedia Tools and Applications*. 2023;82(24):37959–37982. Available from: https://dx.doi.org/10.1007/s11042-022-13670-w.
25) Darapureddy N, Karatapu N, Battula TK. Optimal weighted hybrid pattern for content based medical image retrieval using modified spider monkey optimization. *International Journal of Imaging Systems and Technology*. 2021;31(2):828–853. Available from: https://doi.org/10.1002/ima.22475.
26) Sivakumar M, Kumar NM, Karthikeyan N. An Efficient Deep Learning-based Content-based Image Retrieval Framework. *Computer Systems Science & Engineering*. 2022;43(2):683–700. Available from: https://doi.org/10.32604/csse.2022.021459.