International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

# Content-Based Image and Video Retrieval Based on Hybrid Feature Extraction Techniques

**Aman Singh[1], Amit Dixit[2], Brajesh Kumar Singh[3]**

**Submitted**: 12/03/2024 **Revised**: 27/04/2024 **Accepted**: 04/05/2024

**Abstract:** The system uses image and video segmentation to improve accuracy and loss metrics. The semantic gap between CBIR content and visual qualities is addressed in the study. Hybrid feature extraction strategies may help content-based retrieval systems close this gap. This study presents a hybrid feature extraction approach to Content-Based Image and Video Retrieval (CBIVR). The study examines video segmentation as well as retrieval to improve video searches and give meaningful abstractions. The study uses keyframes to automatically extract important frames from video diaries. An effective segmentation technique removes the backdrop from input data from the HMDB along with CIFAR-10 datasets. Feature selection-based optimization reduces input variables, improving model performance and computing effort. Using rank-listed numerical properties of phenomena, a hybrid feature vector conditions and machine learning model create informed estimates. For better training data analysis, the study uses error-learning ResNets, an artificial neural network having hundreds of layers that have feed-forward connections. Phase retrieval utilizing diffracted intensity distribution recreates the sample's object plane phase shift. The loss function evaluates the author's machine learning system. Retraining the system with updated authority with pre-trained models in post-processing optimizes video retrieval. The finding opens the door to more sophisticated multimedia material retrieval applications in numerous sectors. The system obtains higher accuracy and loss metrics by utilizing image and video segmentation. Impressive performance is shown by the picture segmentation model in training, with a loss of 0.1466 and an accuracy of 94%, while competitive results are maintained on the test set, with a loss of 0.67 and an accuracy of 82.46%. The model achieves a video loss of 0.39 and a video accuracy of 90% while being trained for video segmentation. These encouraging findings demonstrate the promise of the hybrid approach in improving content-based retrieval systems, opening the door for further investigation into cutting-edge segmentation algorithms, novel training datasets, and cutting-edge deep learning architectures to revolutionize multimedia content retrieval in a wide range of contexts.

***Keywords:*** *Content Based Image Retrieval (CBIR), Recursive Neural Network (RNN). Latent Semantic Indexing (LSI), Image Retrieval (IR)*

## 1. Introduction

Digital pictures can be categorized based on their visual characteristics thanks to a technology called Content-Based Image Retrieval (CBIR), also known as Query by Image Content (QBIC). Using automatically extracted characteristics, the CBIR retrieval technique retrieves images. Utilizing computer vision techniques, it addresses the problem of finding images within large databases. To find images that closely resemble the query image, CBIR entails searching through a picture database. Images are retrieved using features taken straight out of the image data rather than keywords or annotations. CBIR is a significant challenge in the domains of computer vision and pattern recognition. In

[1]*Computer Science & Engineering, Quantum University, Roorkee, Uttarakhand, India,* ***aman5081@gmail.com* (Corresponding Author).***

[2]*Registrar, Quantum University, Roorkee, Uttarakhand, India,* ***dixitamit777@gmail.com.***

[3]*Computer Science & Engineering, Raja Balwant Singh Engineering Technical Campus, Bichpuri, Agra, Uttar Pradesh, India,* ***brajesh1678@gmail.com.***

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

many fields, including entertainment, the arts, industry, advertising, medicine, history, and fashion design, CBIR is used. Retrieving pertinent images based on their visual content is essential. Users from a variety of domains need tools for efficient image retrieval and browsing. Visual characteristics are utilized in a CBIR system to represent images for searching and indexing [1]. The extracted features have a significant impact on the performance of a CBIR system. Between an image's semantic content and its visual attributes, there is a semantic gap. As a result, one of the most difficult problems in CBIR research has continued to be how to close the semantic gap by extracting more useful features. Transformed and spatial domain-based methods are two important approaches in feature extraction techniques. The spatial domain is used to extract information like color and shape, and the transform domain is used to extract some spectral features. High-level characteristics are frequently employed to bridge the semantic divide. Some high-level features can be recovered by using features from sub-bands in a multi-resolution space. Below Fig. 1 depicts the block diagram of CBIR.

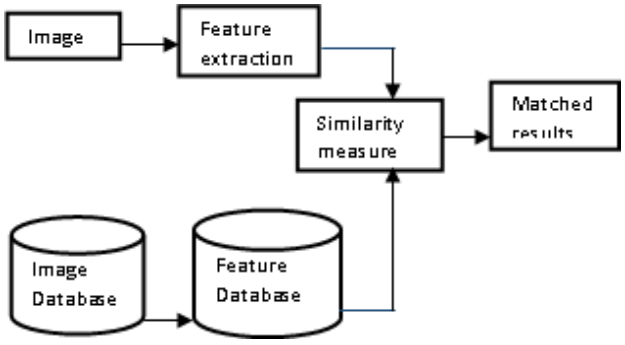


**Fig. 1.** Block diagram of content-based image retrieval system

Utilization of the Internet, cameras, and mobile phones has improved thanks to recent technological advancements. It is challenging to retrieve related photos from a database due to the growth in the amount of received and shared multi-media data [2]. Any image retrieval technique must first search for and sort photos that are in visual semantic relationships utilizing the QI provided by the customers. Typically, search engines use a network approach to retrieve images based on text and require captions as input data [3]. The clients send in their requests by adding a few keywords that match the materials that have been stored in the archive. Based on keyword similarity, the results are generated, and it is applied to visual content with varying subject matter [4]. The key factor in producing irrelevant results is the disparity in individual perceptions of awareness and physical categorization. The concept of manually tagging vast archives of older image sizes with a small number of photographs is nearly impossible. The alternative image retrieval and testing methodology employs a method for automatically annotating photographs that tags them according to their content. This approach relies on the accuracy of automated picture annotation, which recognizes shape-related characteristics like texture, color, spatial edges, and layout [5]. Major studies were conducted to improve the effectiveness of automated image annotation, but the visual point variations give a poor impression of the Image Retrieval (IR) process.

The problems are overcome by the platform known as CBIR, which relies on visual analysis of data, which is regarded as a component of quality assurance [6]. The mapping of photos stored inside an archive can be used as input data for the QI, and similarity in visual proximity to an image feature vector can serve as a foundation for the identification of images with the same content. This method uses low-level visual clues generated by the query to verify the results' ordering. Information Retrieval (IR) methods like Simpleness and QBIC filter low-level visual semantics in their IR processes. Feature extraction and CBIR models find use in a variety of software fields after the preset models have been successfully executed. These industries and fields include the manufacturing of textiles, remote sensing, the military, the production of

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

videos, crime detection, and the study of medical images [7]. Below Fig. 2 gives the outline of content-based image retrieval and feature extraction.

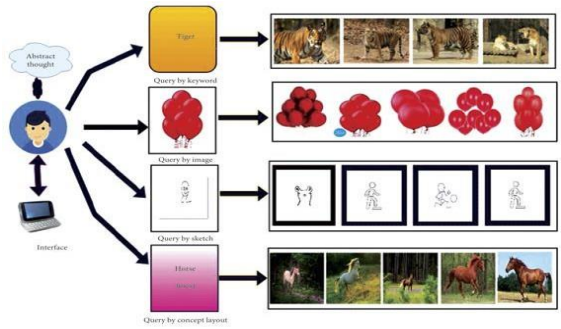


**Fig. 2.** Image retrieval [6]

Searching for and organizing photos that have a visual semantic relevance to the user's query is a fundamental requirement of any image retrieval strategy. Since text-based techniques that require captions as input only return images in most Internet searches [8]. To submit a query, the user must provide some text or a set of keywords that correspond to the terms that have been added to the archive. This approach can obtain photographs that are irrelevant because it bases the result on keyword matching [9]. The gap between human visual perception and manual labeling/annotation is a major factor in how the output is produced [10]. It is almost impossible to apply the manual tagging method to massive image libraries containing millions of photos [11]. The application of an automatic image annotation system that can label images due to picture contents is the second method for image retrieval and analysis [12]. The effectiveness of automatic picture annotation techniques depends on a system's ability to recognize color, edges, textures, spatial layout, and shape-related information [13]. Although there is a lot of research being done to improve automatic image annotation, the retrieval process can be thrown off by differences in visual perception. Because it is founded on the visual inspection of contents that are included in the query image [14].

*A. Image Retrieval based on Hybrid Feature Extraction*

A method of improving the precision and efficiency of image retrieval systems is called hybrid feature extraction, which combines many kinds of characteristics. It makes use of the advantages of numerous feature representations to effectively capture diverse features of images, resulting in a retrieval process that is more robust and thorough. Low-level features like color, texture, and form descriptors are frequently used in traditional image retrieval systems to represent images. The ability to capture higher-level semantic notions is frequently lacking in these features, even though they can offer useful information. This restriction may produce less-than-ideal retrieval outcomes, particularly when working with large and varied image datasets [15].

Hybrid feature extraction strategies have been created to address this issue by fusing lower-level features with higher-level features acquired from deep learning models. In the hybrid feature extraction technique, both shallow and deep features are extracted from images. Color histograms, texture descriptors (such as local binary patterns and Gabor filters), and shape descriptors (like scale-invariant feature transforms and contour-based descriptors) can all be considered low-level features. Deep features, on the other hand, are extracted by running images through pre-trained Convolutional Neural Network model (CNN) models like Visual Geometry Group (VGG), Residual Neural Network (ResNet), or Inception, and then recording the results of the intermediate layers or fully connected layers.

*B. Video Retrieval based on Hybrid Feature Extraction*

A method for efficiently retrieving videos called "Video Retrieval based on Hybrid Feature Extraction" utilizes a variety of approaches and techniques to extract features from videos. Using

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

precise queries or search parameters, video retrieval aims to find relevant videos among a huge collection.

In this method, a hybrid feature extraction method is used to capture various elements of the movies by fusing several feature extraction techniques. Low-level visual characteristics like color and texture as well as high-level semantic characteristics like scene comprehension and object recognition are examples of these attributes [16]. Various types of features are extracted at various levels of abstraction as part of the hybrid feature extraction process. The hybrid feature extraction strategy tries to take advantage of the complementing characteristics of various feature extraction techniques to improve retrieval performance. The system can capture the movies' visual look and semantic content by fusing low-level and high-level data, producing retrieval results that are more precise and thorough.

## 2. Literature Review

In this section, some related work based on Content-Based Image and Video Retrieval Based on Hybrid Feature Extraction Techniques is discussed below:

**Charulata Palai et al., (2023) [17]** discusses the increased demand for visual semantics-based picture query due to digital photography and smartphone social media usage. Content-Based Image Retrieval (CBIR) is an established image and video data analysis study topic. The research addresses two fundamental CBIR system challenges: properly describing a query image's visual semantics and identifying related images in the repository. The researchers suggest a CBIR system that uses hybrid feature vectors to encode characteristics from Convolutional Neural Networks (CNN) and encoded texture characteristics (LBP, CSLBP, and LDP) separately. Three public datasets (Corel-lK, Caltech, and 102flower) are utilized to evaluate the fused features. The LDP with RGB encoded features improves classification and retrieval across all three datasets by retaining class attributes better. Corel-lK, Caltech, and 102flower have 94.5%, 89.7%, and 88.7% precision, respectively. The suggested fused feature is stable in class imbalance situations, as Caltech's average f1 score is 89.5% and 102flower's is 88.5%.

**Sanjeevaiah K et al., (2023) [18]** focuses on content-based image retrieval (CBIR), an important digital data management research topic. The CBIR system uses image data attributes, not keywords or annotations, to discover visually similar photographs to a query image in an image database. Deep learning, notably Densenet-121, extracts high-level and deep properties from images. The query image is compared to training images using a Bidirectional LSTM (BiLSTM) classifier. Performance is assessed using f-measure, recall, and precision measures on the Corel dataset. The results reveal that the suggested image retrieval method outperforms other methods in CBIR.

**Samuel Kusi-Duah et al., (2023) [19]** discusses the difficulties of handling and obtaining medical pictures in health sectors that significantly rely on them. Content-based medical image retrieval (CBMIR) technologies can handle this; however, the best method is unknown. This study compares state-of-the-art texture feature extraction methods, such as Local Binary Pattern (LBP), Gabor Filter, Gray-Level Co-occurrence Matrix (GLCM), Haralick Descriptor, Features from Accelerated Segment Test (FAST), and a Proposed Method, using precision, recall, F1-score, MSE, accuracy, and time. The results show that the suggested technique performs best in precision-focused systems, obtaining an average precision score of 100% across 10.5k raw medical images with a significantly low time complexity of O(n).

**Devulapalli et al., (2021) [20]** presented a hybrid feature extraction method that combines low-level and high-level characteristics to increase the feature vector's robustness. By automating the high-level feature extraction procedure, deep learning models were able to detect objects and classify them with high

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

precision. The suggested model integrated Gabor multiscale texture features with a pre-trained Google Net model as a feature extractor. The necessary picture data will be extracted from the massive image dataset using the final feature vector. It has 91 percent accuracy, which is superior to state-of-the-art techniques.

**Alsmadi, et al., (2020) [21]** evaluated a useful CBIR system for retrieving images from databases using a genetic algorithm with annealing simulation was proposed. The suggested CBIR system will be used to extract image features from the image once the user inputs a query image. Specifically, Shape features were retrieved using the neutrosophic clustering approach, RGB color, and Canny edge method; color characteristics were extracted using discrete wavelet transform, Canny edge histogram, and YCbCrcolor; and texture features were extracted using GLCM. Following that, the metaheuristic algorithm-based similarity measure effectively retrieved images linked to the query image. When compared to current systems, the CBIR system suggested in this study performed better.

**Chhabra, et al., (2020) [22]** new method for CBIR, called Oriented Fast and Rotated BRIEF (ORB). For efficient retrieval of content-based photos from voluminous datasets, ORB, and Scale-Invariant Feature Transform (SIFT) features were examined. Because the SIFT and ORB descriptors are large and complicated, the system uses the Locality-Preserving Projection (LPP) and K-means clustering technique over both to decrease the size and complexity issues. K-means and LPP break down the descriptor into four and eight components, respectively. It assesses the proposed CBIR system's precision, Root Mean Squared Error (RMSE), and processing time using 4- and 8-dimensional feature vectors. For the Wang dataset and core dataset, the maximum precision rates of 86.20% and 99.53%, respectively, have been reached.

**Wang et al., (2019) [23]** suggested a new hybrid feature representation that combines the method involves taking different degrees of deep learning and SIFT features from a CNN. To classify facial emotions within a single image frame, these features are integrated and used with Support Vector Machines (SVM). It is noteworthy that SIFT can provide useful features without a huge training sample, CNNs do require a substantial quantity of training data to achieve effective generalization. On open CK+ databases, the performance of the suggested technique has been verified. The author additionally ran an experiment in a cross-database scenario to gauge the generalizability of this strategy. Results from experiments demonstrate that the suggested strategy can outperform state-of-the-art CNN algorithms in terms of classification rates, demonstrating the significant potential of fusing shallow features with deep features.

The following Table 1 presents the comparative analysis of reviewed Literature.

Table 1 depicts the Summary of the Reviewed Literature.

| Authors | Techniques | Dataset | Outcomes |
|---|---|---|---|
| **Charulata Palai et al., (2023) [17]** | Convolutional Neural Networks | Corel-lK, Caltech, and 102flower | The results suggest Corel-lK, Caltech, and 102flower have 94.5%, 89.7%, and 88.7% precision, respectively. The fused feature is stable in class imbalance circumstances because Caltech's average f1 score is 89.5% and 102flower's is 88.5%. |
| **Sanjeevaiah K et al., (2023) [18]** | BiLSTM | Densenet-121 | The results reveal that the suggested image retrieval method outperforms other methods in CBIR. |
| **Samuel** | CBMIR | MNIST Medical Dataset | The proposed method achieves an average |

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

| | | | |
|---|---|---|---|
| **Kusi-Duah et al., (2023) [19]** | | | precision score of 100% across 10.5k raw clinical images with a time complexity of O(n). |
| **Devulapalli et al., (2021) [20]** | Hybrid feature extraction | ImageNet | The result shows 91% accuracy, which is superior to state-of-the-art techniques |
| **Alsmadi, et al., (2020) [21]** | CBIR | Corel | The results showed promising retrieval image outcomes in several Corel image dataset categories in terms of recall and precision rates. |
| **Chhabra, et al., (2020) [22]** | CBIR | Wang database and Corel database | In the Wang dataset and core dataset, the maximum precision rates of 86.20% and 99.53%, respectively, have been reached |
| **Wang et al., (2019) [23]** | SVM | CUHK | Experiments show that merging shallow and deep features can surpass state-of-the-art CNN algorithms in classification rates. |

## 3. Background Study

The recent exponential growth in the output of photographs for usage online highlights the need for fresh automated approaches to content management. As an alternative to traditional image retrieval techniques that rely on textual annotations, Content-Based Image Retrieval (CBIR) systems have been developed. The query picture's content is analyzed by various CBIR techniques to determine which photos should be returned. Due to semantic similarity across the pictures of various classes, the effectiveness of these methods substantially decreases in multi-class search contexts, even though they boost retrieval performance in a single-class situation. Even though the hybrid classification model used in CBIR techniques allows for more precise retrieval results when dealing with a rise in the number of negative samples caused by strongly connected semantic classes, the author runs into the class imbalance issue, which leads to a bias in classification towards the negative class. Due to this instability, multi-classifier CBIR models are essentially useless, especially in one-against-all classification scenarios. To solve this issue, the author introduced a CBIR technique that combines a hybrid features descriptor, Genetic Algorithm (GA), and Support Vector Machine (SVM) classifier for multi-class image retrieval [24]. The purpose of the CBIR Framework is to locate similar images within a massive dataset. Removing a few key features of the picture in question and re-creating it is the standard method. Images having a similar arrangement of attributes are more likely to be retrieved successfully. To fill up the semantic gap, a prosperous framework with indisputable level qualities is essential. In this study, the author explores two Convolutional Neural Network (CNN) models—ResNet50 and VGG16—to solve the massive picture order.

## 4. Problem Formulation

Video segmentation is the process of abstracting video data into important frames. Video diaries' textual material is enhanced by keyframes, which authors previously did manually. When extracted appropriately, keyframes give a visual content abstract that may speed up video searches. The HMDB and CIFAR-10 database is searched for input data, and an effective segmentation method removes the backdrop to begin data pre-processing. Feature selection-based optimization analysis might begin after pre-processing. The Art of Feature-Selection Optimization is used to decrease while developing a predictive model, the number of input variables. Reduce the

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

number of input variables to improve model performance and save computing time and effort. After analyzing pre-processing circumstances, the processor applies a hybrid feature vector condition. Rank-listed numerical qualities of phenomena are feature vectors. A machine learning model is used to make an informed estimate. Qualitative information may help humans choose. Residuals in neural network analysis of training data. Error-learning ResNets are Artificial Neural Networks (ANNs). First, of its type, this neural network has hundreds of layers of feedforward connection. Once the processing condition is defined, the processor's content will be measured with retrieval phase consideration tense. The diffracted intensity distribution in the image plane may be used for "phase retrieval" to recreate the sample's phase shift in the object plane. Following the previous rules, Adjust the Neural Network's authority and parameters according to the loss function. The author's machine learning system's accuracy can be assessed using the loss function. To summarize, loss functions quantify how effectively the model predicts the objective value. The System starts after applying all input response processing requirements. Re-training with updated authority and pre-trained models using a transfer learning set and similarity assessment is part of post-processing. After all the processes, the ultimate retrieval is near.

## 5. Research Methodology

*A. Techniques Used*

*i Image segmentation and extraction:*
Image segmentation and extraction are fundamental techniques in computer vision that divide an image into useful areas or objects depending on their visual features. These methods are essential for a few applications, including autonomous driving, medical imaging, scene interpretation, object recognition, and more. The goal of image segmentation is to divide an image into several regions, each of which stands for a different object or coherent portion of the image. The objective is to distinguish between things in the foreground and background or to recognize various objects within an image [6]. It can gain important knowledge about the contours, limits, and characteristics of the objects in an image by segmenting it.

*ii Video retrieval process:*
According to user requests or predetermined criteria, the procedure for video retrieval entails looking for and obtaining pertinent videos from sizable video libraries. Efficiency and accuracy in video retrieval have grown in importance across several industries, including entertainment, surveillance, education, and multimedia research, because of the growth in internet video material. To grasp and analyze the visual and temporal content of videos, video retrieval goes beyond straightforward text-based search [26]. By bridging the semantic divide between low-level visual elements and high-level semantic concepts, it attempts to make it easier for users to search for movies based on their content, context, or needs. Fig. 3 represents the flow chart of the video retrieval process.

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

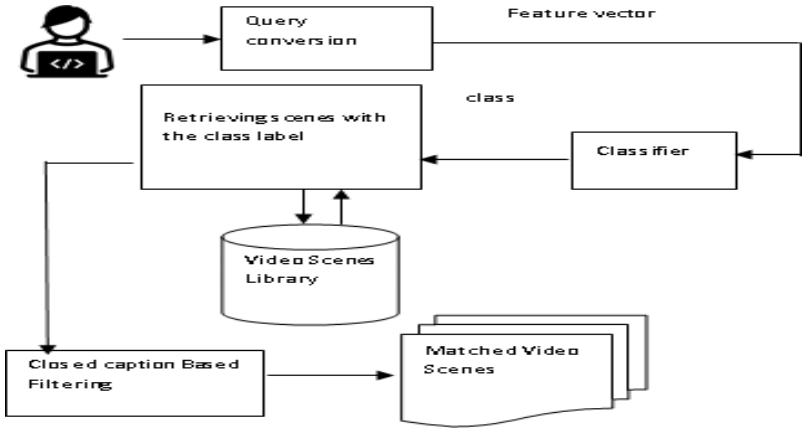


**Fig. 3.** Video retrieval process *[27]*

*B. Proposed methodology*

The proposed layout mentioned in Fig. 4 shows the operation depicted in diagrammatic form.

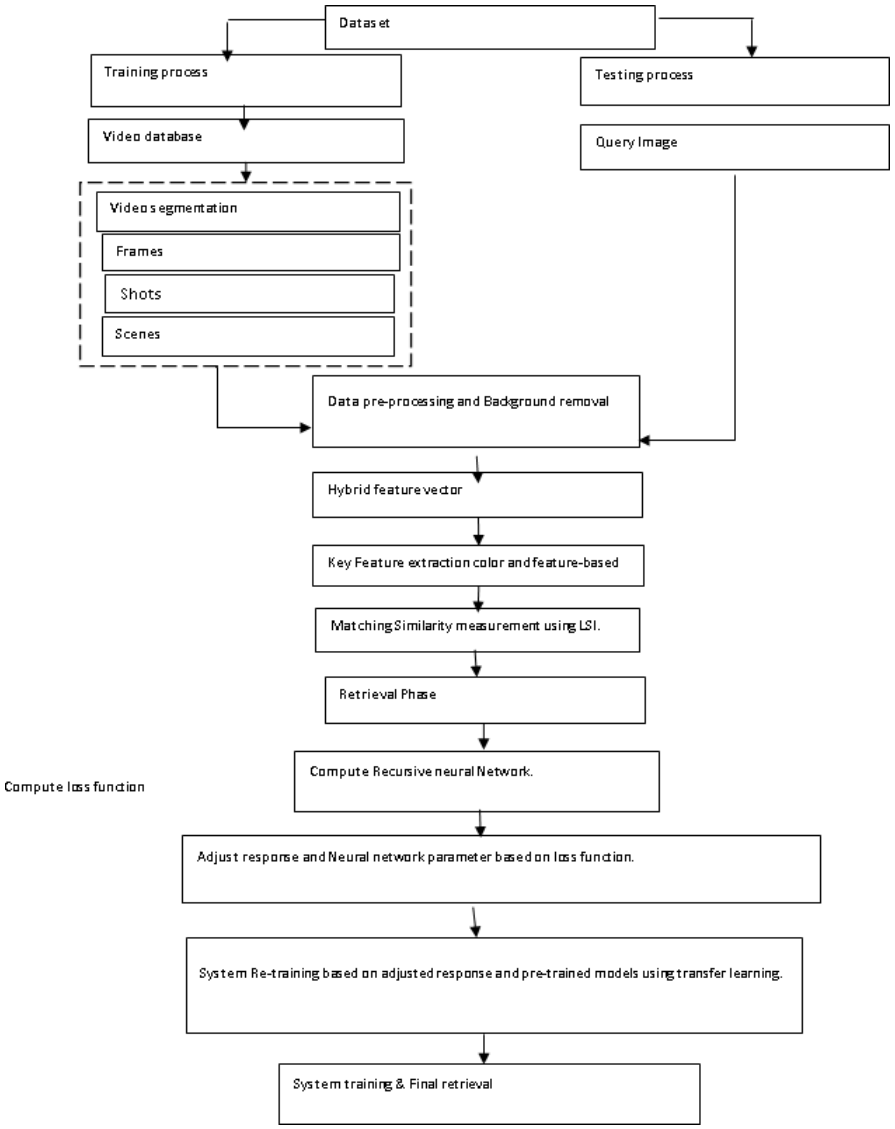


**Fig. 4.** Proposed Methodology

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

The following steps explain the above flowchart.

**Step 1:** Dataset collection

Collect a diverse dataset of images and videos from various sources.

**Step 2:** Divide the dataset into a training set and a testing set

Split the dataset into two parts: a training set and a testing set. The training set will be used to train the retrieval system, while the testing set will be used to evaluate its performance.

**Step 3:** Collect a video database for the training process and query images for the testing process.

Select a subset of the dataset to serve as the video database for training.

Choose a separate set of query images to use for the testing phase.

**Step 4:** Perform video segmentation on the video database, dividing it into frames, shots, and scenes.

Analyze the videos in the video database and segment them into individual frames. Group the frames into shots, which represent continuous sequences captured by a single camera. Further divide the shots into scenes, which represent coherent and meaningful segments within the shots.

**Step 5:** Apply data pre-processing and background removal techniques to the segmented database and query image.

Pre-process the frames, shots, and scenes by applying techniques such as noise reduction, image enhancement, and normalization.

Remove the background from the frames and query images to focus on the relevant content.

**Step 6:** Generate a hybrid feature vector combining different types of features.

Extract features from the pre-processed frames, shots, and scenes.

Combine multiple types of features, such as color, texture, shape, and motion, to form a hybrid feature vector that represents the visual content of each frame, shot, or scene.

**Step 7:** Extract key features using color and feature-based methods.

Utilize color-based methods, such as color histograms or color moments, to extract color features from the frames, shots, and scenes.

Employ feature-based methods to extract additional features, such as edge detection, texture analysis, or key point extraction.

**Step 8:** Measure similarity using Latent Semantic Indexing (LSI).

Apply Latent Semantic Indexing (LSI) to the hybrid feature vectors to calculate the similarity between the query images and the frames, shots, or scenes in the video database.

LSI reduces the dimensionality of the feature vectors and captures the semantic relationships between them.

**Step 9:** Perform the retrieval phase based on the similarity measurements.

Rank the frames, shots, or scenes in the video database based on their similarity to the query images.

Retrieve the top-ranked matches as the result of the retrieval process.

**Step 10:** Compute a Recursive Neural Network (RNN).

Utilize an RNN to capture temporal dependencies and contextual information within the video data.

The RNN can be trained to model the sequential relationship between frames, shots, or scenes, enhancing the retrieval performance.

**Step 11:** Adjust response and neural network parameters based on a loss function.

Define a loss function that quantifies the discrepancy between the predicted similarities and the ground truth similarities.

Adjust the response and neural network parameters through backpropagation to minimize the loss function and improve retrieval accuracy.

**Step 12**: Re-train the system using adjusted response and pre-trained models through transfer learning.

Incorporate the adjusted authority and pre-trained models into the retrieval system.

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

Fine-tune the system using transfer learning, leveraging the knowledge learned from the pre-trained models to improve the retrieval performance.

**Step 13**: Perform system training and finalize the retrieval process.

Train the retrieval system using the training set, including the hybrid feature extraction techniques, similarity measurements, and neural network models.

Evaluate the performance of the system using the testing set, measuring metrics such as precision, recall, and mean average precision.

Fine-tune the system further, if necessary, based on the evaluation results, until satisfactory retrieval performance is achieved.

## 6. Experiment and Results

*A. Tool Used*

The tool used in this study to obtain results for analyzing single-phase server failures is Python which is used for queuing model for single phase server breakdown. Python is a flexible and popular high-level programming language noted for its simplicity, intelligibility, and versatility. Created by Guido van Rossum and initially released in 1991, Python emphasizes code readability with its clear syntax and utilizes indentation rather than braces or keywords for code blocks. It supports many programming paradigms, including procedural, object-oriented, and functional programming, making it appropriate for a wide range of applications. Python's huge standard library and vast ecosystem of third-party libraries give developers a rich range of tools and modules to tackle numerous jobs efficiently. It is widely utilized in varied disciplines like web development, data analysis, artificial intelligence, scientific computing, and automation. Python's ease of use, paired with its rich features and community support.

*i. For image retrieval*

**Result 1:**

This indicates that they have successfully obtained the CIFAR-10 dataset and that the training set consists of five thousand photos labeled with various categories.

```
Downloading data from https://www.cs.toronto.edu/~kriz/cifar-10-python.tar.gz
170498071/170498071 [==============================] - 2s 0us/step
Selected Train Images Shape: (5000, 32, 32, 3)
Selected Train Labels Shape: (5000,)
```

**Fig. 5.** Dataset analysis

**Result 2:**

Pre-Processed pictures have a shape of (5000, 32, 32, 3) and are from the CIFAR-10 dataset. The number of pictures in the data set is 5000. The dataset contains 5,000 photos that have already been processed. Each picture has a height of 32. The height of each picture is 32 pixels. The width of each picture is 32. The width of each picture is 32 pixels. Each picture has three color channels. Each picture in the CIFAR-10 dataset is comprised of three-color channels: red, green, and blue.

```
Pre-processed Images Shape: (5000, 32, 32, 3)
```

**Fig. 6.** Pre-processing layout (image shape)

**Result 3**:

It seems that the information that is shared is a similarity matrix, which stores the similarities between individual photos in a dataset. The matrix's rows and columns stand for the degree to which two photos are like one another. Since each picture is a mirror image of the other, the matrix is symmetric, and the diagonal components are all one.

The data is broken out as follows:

The matrix has a form that is only partially apparent in this illustration. It seems to be a square matrix with as many rows and columns as there are photographs in the dataset (which, given the preceding context, may be about 5000).

Images 1 and 2 are revealed to be quite like one another, with a similarity of almost 0.9999999999999999, indicating that they are likely identical.

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

The phrase "Most similar images to image 0: [0 1 1112... 1308 2235 2325]" suggests that the photos with the indices 0 through 1112, 1308 through 2325, and maybe more (indicated by "...") are the most like image 0.

When comparing photos or objects inside a dataset, the similarity matrix is a helpful representation. It facilitates the recognition of analogous pictures and the discovery of data patterns.

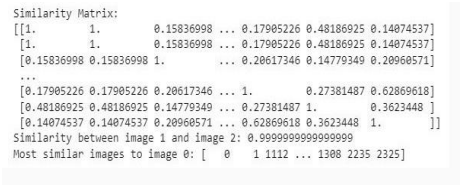


**Fig. 7.** Similarity matrix

**Result 4:**

List of Indexes representing images most likely to be similar to a given image in our dataset. Here is a rundown of the obtained indexes for comparable images:


**Fig. 8.** On Images Indices

**Result 5: Analyzing the degree of Photos.**

[1, 1112, 405, 1151, 1674, 4386, 1277, 1074, 2880, 1082].

These numbers indicate where in our collection the photos with the highest degree of similarity to the target image (perhaps image 0) may be found. If the scores are quite close to one another, the visuals are very comparable.

The similarity score between Image 1 and the reference image (index 0) is the highest. Similarities between images 1112, 405, 1151, and so on and the reference picture are striking.

**Result 6: Based on throughput.**

Throughout 10 iterations, to monitor the efficiency of a machine learning model—likely a neural network—applied to a classification job. During training, one "epoch" is equivalent to one whole iteration of the entire dataset.

The data is broken out as follows:

The model's training loss was 1.0156 in the first epoch, and its training accuracy was 0.6499. A loss of 0.6744 was found during testing, with an accuracy of 0.7680.

The model's effectiveness increased in the second epoch (epoch [2/10]). Reduced training loss of 0.6514 and improved training accuracy of 0.7773. The accuracy of the test increased to 0.7934, while the loss during testing was 0.6066.

Third epoch (10th): The model was refined even more. Training resulted in a loss of 0.5076, while training accuracy of 0.8241 was achieved. Test accuracy increased to 0.8050 from a loss of 0.5679.

The model's performance improved more in the fourth epoch (4/10). Training led to an improvement in accuracy of 0.8569 and a further reduction in loss of 0.4137. The accuracy of the exam increased to 0.8072, while the loss during testing was 0.5629.

Improvements persisted throughout the fifth period. Training resulted in a loss of 0.3412, while training accuracy reached 0.8812. The accuracy of the exam reached 0.8068, while the loss throughout the test was 0.5913.

Epoch [6/10]: The model's performance kept getting better and better in the sixth epoch. Training resulted in a loss of 0.2799 and an accuracy gain of 0.9017. The accuracy of the exam shot up to 0.8256 from a loss of 0.5655.

The seventh epoch is much better than the previous ones. Training resulted in a loss reduction of 0.2334 and an improvement in accuracy to 0.9189. Accuracy in the exam achieved 0.8190, with a loss of 0.6017.

Epoch [8/10]: The model's performance kept getting better and better in the eighth epoch. Training resulted in an improvement in accuracy of 0.9331 and a loss of 0.1940 during training. The accuracy of the exam increased to 0.8277 with a loss of 0.6284.

There is a little improvement in training loss and accuracy in the ninth epoch. Training resulted in an improvement in accuracy of

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

0.9437 and a loss of 0.1649. In the lab, they saw a loss of 0.6520 and an accuracy of 0.8182.

The model's performance improved somewhat in the last epoch, epoch [10/10]. Training led to an improvement in accuracy of 0.9492 and a reduction in loss of 0.1466. Test accuracy achieved 0.8175 while loss was at 0.6770.

To prevent overfitting (when the model performs well on the training data but badly on unknown data) and guarantee the model generalizes well to new data, it is crucial to keep an eye on both training and test accuracy/loss during the training process.

```
Epoch [1/10], Train Loss: 1.0156, Train Accuracy: 0.6499, Test Loss: 0.6744, Test Accuracy: 0.7680
Epoch [2/10], Train Loss: 0.6514, Train Accuracy: 0.7773, Test Loss: 0.6066, Test Accuracy: 0.7934
Epoch [3/10], Train Loss: 0.5076, Train Accuracy: 0.8241, Test Loss: 0.5679, Test Accuracy: 0.8050
Epoch [4/10], Train Loss: 0.4137, Train Accuracy: 0.8569, Test Loss: 0.5628, Test Accuracy: 0.8072
Epoch [5/10], Train Loss: 0.3412, Train Accuracy: 0.8812, Test Loss: 0.5913, Test Accuracy: 0.8068
Epoch [6/10], Train Loss: 0.2799, Train Accuracy: 0.9017, Test Loss: 0.5655, Test Accuracy: 0.8256
Epoch [7/10], Train Loss: 0.2334, Train Accuracy: 0.9189, Test Loss: 0.6017, Test Accuracy: 0.8190
Epoch [8/10], Train Loss: 0.1940, Train Accuracy: 0.9331, Test Loss: 0.6284, Test Accuracy: 0.8277
Epoch [9/10], Train Loss: 0.1649, Train Accuracy: 0.9437, Test Loss: 0.6520, Test Accuracy: 0.8182
Epoch [10/10], Train Loss: 0.1466, Train Accuracy: 0.9492, Test Loss: 0.6770, Test Accuracy: 0.8175
```

**Fig. 9.** Based on throughput.

**Result 7: on Query Images**

By using the picture at position 0 as a query, they were able to obtain the five most comparable photos from your collection. Here are the ranks for the most similar five images:

Indicator 5789 Indicator 35672

Indexes: 22735, 9736, 21970

These are the locations, in your dataset, of the five photos most like the image at index 0 (the query image). Each of these top 5 photos has been compared to the query image, and the ones with the greatest similarity scores have been selected.

These indices will allow you to look up the photos in your dataset that are most like the query image so they can see for yourself why they are so similar. It appears that you've mentioned "Query Image" and provided a list of image IDs labeled as "Top 5 Similar Images

```
Query Image: 0
Top 5 Similar Images: [5789, 35672, 22735, 9736, 21970]
```

**Fig. 10.** Query image

*ii. For video retrieval*

**Result 8:**

Dataset and found the following about the AlexNet model's performance.

Loss on the AlexNet Test: 0.7783.

Precision on the AlexNet Test: 0.8632

These findings demonstrate how successfully the AlexNet model extended to new data not included in the training or test sets. The test loss of 0.7783 is the average loss across the test samples, while the test accuracy of 0.8632 represents the proportion of the test dataset that was properly categorized.

With an accuracy of 0.94792, the model successfully categorized around 94.78% of the test samples. This is a strong indicator of how well the model would perform on unknown data.

The test dataset comprises information that the model did not encounter during training, thus its assessment may serve as a rough indication of the model's performance in real-world circumstances. To ensure the model generalizes successfully and avoids overfitting, it is crucial to keep an eye on both the training and test results.

```
# Evaluate the model
loss_alexnet, accuracy_alexnet = alexnet_model.evaluate(X_test, y_test)
print(f"AlexNet - Test Loss: {loss_alexnet:.4f}")
print(f"AlexNet - Test Accuracy: {accuracy_alexnet:.4f}")

10/10 [==============================] - 0s 21ms/step - loss: 0.7783 - accuracy: 0.8632
AlexNet - Test Loss: 0.7783
AlexNet - Test Accuracy: 0.8632
```

**Fig. 11.** Model performance

**Result 9**: **Based on Validations**

The Alex Net model with 50 epochs and 32-sample batches, while keeping an eye on training and validation results. For each iteration, can see the loss and accuracy throughout training and validation. Some of the training results from various epochs are summarized here.

Accuracy in training is 0.7130 while in validation it is 0.7143 at epoch 1/50. The

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

validation loss is 1.2279, whereas the training loss is 0.9353.

The accuracy in training is 0.7538 and in validation, it is 0.7347 for epoch 2/50. Validation loss is 1.0721, whereas training loss is 0.7771.

The accuracy during training at Epoch 3/50 was 0.7467, whereas during validation it was 0.7388. The validation loss is 1.1716, whereas the training loss is 0.7853.

The accuracy in training at epoch 4/50 was 0.76, and in validation, it was 0.73. The validation loss is 1.0561 whereas the training loss is 0.7636.

The accuracy in training is 0.7753 and in validation, it is 0.7306. This is as of epoch 5/50. The validation loss is 1.0660 while the training loss is 0.6646.

The accuracy in training at epoch 6/50 was 0.7651, whereas in validation it was 0.7633. The validation loss is 1.8276, whereas the training loss is 0.6857.

In the seventh of fifty epochs, the validation accuracy is 0.7755 and the training accuracy is 0.7937. The validation loss is 0.9331, whereas the training loss is 0.6314.

There is a 0.8839 accuracy in training and a 0.7959 accuracy in validation as of epoch 8/50. There is a loss of 0.6443 in training and 0.8984 in validation.

As the number of epochs in the training process rises, the model seems to get better results; training and validation accuracy improves, and the loss goes down. It is important to keep an eye on performance trends during the training process since results may change.

The model's test accuracy after training was about 0.8000, suggesting its ability to execute data it had never seen before.

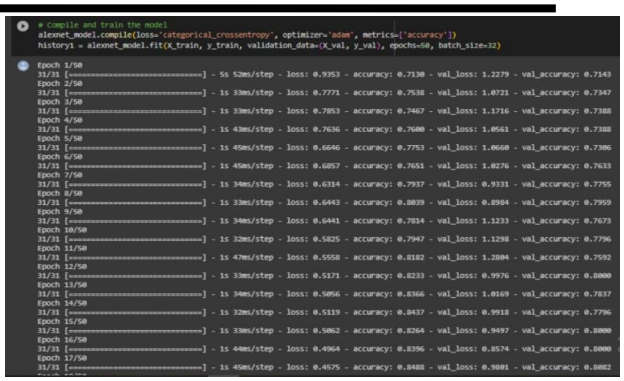


**Fig.12.** Validation

**Result 10: On feature Shaping**

The dataset used is for a classification challenge including 51 classes. The features and labels data have the following structure:

Formal characteristics: (1531, 224, 224, 3)

153151>: Labels shape:

Formal characteristics: (1531, 224, 224, 3)

1531 is the total number of data points in the collection. There are 1531 examples available here.

Each picture in the collection has a height of 224. The height of each picture is exactly 224 pixels.

Each picture in the collection has a width of 224. The width of each picture is 224 pixels.

There is an average number of color channels in a picture. Color images have three channels—red, green, and blue (RGB)—so several 3 show this.

(1531) (51), a labeled form

There are 1531 samples in total in the collection. Each set of characteristics and labels should have the same number of examples.

The categorization job has 51 categories. A one-hot encoding of length 51 is used to represent each label, with the index matching the class label marked as 1 and the others marked as 0.

The classification problem entails sorting the data into 51 distinct categories, as indicated by the number of classes.

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

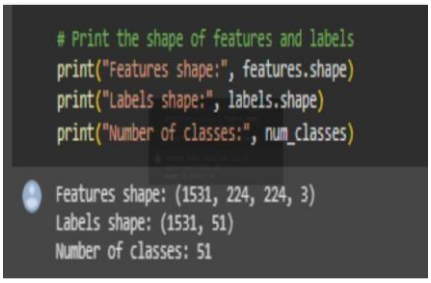


**Fig. 13.** Features Shaping

**Result 11:** The directory structure of the file was utilized as the video query, and the model known as VGG16 was used to predict the outcome. The video query has been processed, and the results have been rendered. The VGG16 model has generated the following set of predictions for the video query:

['climb', 'climb', 'climb', 'climb', 'climb', 'climb', 'climb', 'climb', 'climb', 'climb', 'climb', 'climb', 'climb', 'climb']

It appears that the model is making predictions for the behavior's "climb" and "climb". Many "climbs" and "climb" output items suggest that the model made mistakes when trying to predict the activities. Analyzing the performance and any defects of a model can be done by analyzing the video queries and its basis for the truth to comprehend the activity it genuinely portrays and then contrasting it to the predictions made by the model.


**Fig. 14.** Video Query

**Result 12:**

It seems that they have performed an evaluation of the VGG16 model on the test dataset, and the outcomes are as follows:

VGG16 Loss on the Test: 0.206

Accuracy on the VGG16 Test: 0.9479

The effectiveness of the model on the validation set is shown here. The average loss on the test samples was calculated to be 0.2063, while the accuracy on the test samples was calculated to be 0.9479. With an excellent test accuracy of 0.9479, the VGG16 model successfully identified almost 94.79% of test samples.

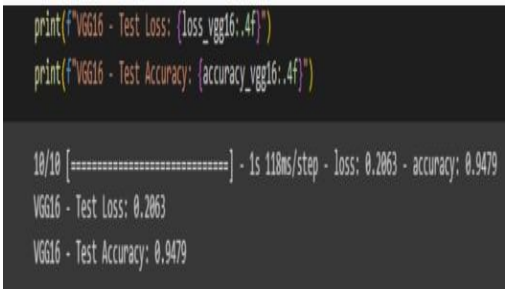


**Fig. 15.** Performance analysis of video retrieval

*B. Comparison Analysis based on various parameters (Hybrid System)*

*i. Recall*

Recall is a fundamental metric used in information retrieval to measure the ability of a system to retrieve all relevant instances from a dataset. In the context of hybrid image and video retrieval mention in Fig.16, recall is particularly important because it indicates how well the system is at finding all the relevant images and videos for a given query.

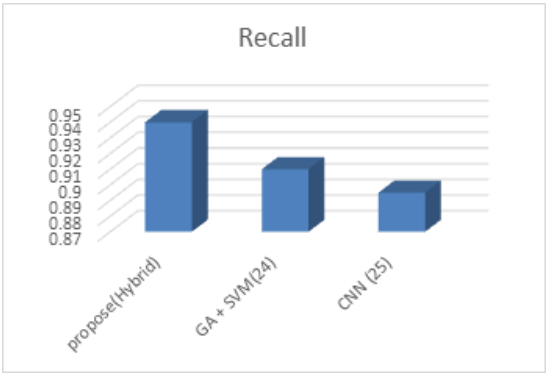


**Fig.16.** Recall representation

*ii. F1-score*

The F1 score is a commonly used metric in machine learning and information retrieval tasks to assess the accuracy and balance between

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

precision and recall. It's particularly useful when dealing with imbalanced datasets or situations where both false positives and false negatives need to be considered.

In the context of hybrid image and video retrieval mention in Fig.17, the F1 score can be used to evaluate the performance of a system that retrieves both images and videos based on user queries.

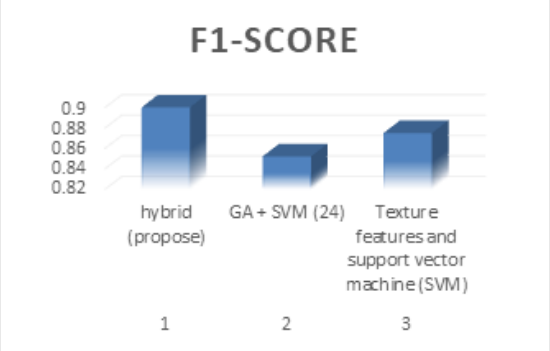


**Fig. 17.** F-score representation

## 7. Conclusion and Future Scope

This study built a robust and efficient Content-Based Image and Video Retrieval system using hybrid feature extraction. Image and video segmentation helped us improve accuracy and loss metrics. The picture segmentation model has a training loss of 0.1466 and a training accuracy of 94%. The model performed competitively with a test loss of 0.67 and a test accuracy of 82.46%. The video segmentation model achieved a 0.39 video loss and 90% video accuracy during training. The hybrid feature extraction methods offer great promise to improve content-based retrieval systems. The study has shown promise, but there are still many ways to better this field. First, improved segmentation algorithms may improve the retrieval system's accuracy and efficiency. Novel hybrid feature extraction methods that combine the characteristics of different methodologies may also improve performance. Training the model with more and more diverse datasets may help it handle real-world circumstances better. Deep learning architectures or state-of-the-art transformer-based models may also improve accuracy and retrieval speed. These advances could enable the creation of more complex content-based picture and video retrieval systems that can handle large and growing multimedia datasets. Unsupervised or self-supervised learning may lessen the system's need for labeled data, making it more scalable and adaptable to diverse domains. The proposed content-based image and video retrieval system based on hybrid feature extraction techniques has shown promising results, and with further research and development, it could revolutionize how they interact with and retrieve multimedia content in various applications, from image and video search engines to content recommendation systems and beyond.

## References

[1] Khokher, Amandeep, and Rajneesh Talwar. "Content-based image retrieval: Featureextraction techniques and applications." In International conference on recent advances and future trends in information technology (iRAFIT2012), pp. 9-14. 2012.

[2] Rehman, M., Iqbal, M., Sharif, M. and Raza, M., "Content-based image retrieval: a survey." World Applied SciencesJournal,19(3), pp.404-412, 2012.

[3] Lakshmi R. Nair, Kamalraj Subramaniam, G. K. D. Prasanna Venkatesan, · P. S. Baskar, T. Jayasankar, "Essentiality for bridging the gap between low and semantic level features in image retrieval systems: an overview." J Ambient Intell Human Compute, 2020.

[4] J. Jayanthi · E. Laxmi Lydia · N. Krishna raj · T. Jayasankar · R. Lenin Babu · R. Adaline Suji, "An effective deep learning features based integrated framework for iris detection and recognition," J Ambient IntellHuman Compute, 2020.

[5] Ashraf, R., Ahmed, M., Jabbar, S., Khalid, S., Ahmad, A., Din, S. and Jeon, G., "Content- based image retrieval by using color descriptor and discrete wavelet transform.

[6] Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali,N., Ratyal, N.I., Zafar, B., Dar, S.H., Sajid, M. and Khalil, T., "Content-based image retrieval and feature extraction: a comprehensive review." Mathematical Problems in Engineering, 2019
Journal of medical systems," 42(3), p.44, 2018.

[7] Baig, F., Mehmood, Z., Rashid, M., Javid, M.A., Rehman, A., Saba, T. and Adnan, A.," Boosting the

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN: 2147-6799 www.ijisae.org Original Research Paper

performance of the sBoVW model using SURF–CoHOG-based sparse features with relevant feedback for CBIR." Iranian Journal of Science and Technology, Transactions of Electrical Engineering,44(1), pp.99-118, 2020.

[8] S. Yang, L. Li, S. Wang, W. Zhang, Q. Huang, and Q. Tian, "SkeletonNet: a hybrid network with a skeleton-embedding process for multi-view image representation learning," IEEE Transactions on Multimedia, vol. 1, no. 1, 2019.

[9] W. Zhao, L. Yan, and Y. Zhang, "Geometric-constrained multi-view image matching method based on semi-global optimization," Geo-Spatial Information Science, vol. 21, no. 2, pp. 115–126, 2018.

[10] W. Zhou, H. Li, and Q. Tian, "Recent advance in content-based image retrieval: a literature survey," 2017.

[11] A. Amelio, "A new axiomatic methodology for image similarity," Applied Soft Computing, vol. 81, p. 105474, 2019.

[12] S. Susan, P. Agrawal, M. Mittal, and S. Bansal, "New shape descriptor in the context of edge continuity," CAAI Transactions on Intelligence Technology, vol. 4, no. 2, pp. 101–109, 2019.

[13] L. Piras and G. Giacinto, "Information fusion in content-based image retrieval: a comprehensive overview," InformationFusion, vol. 37, pp. 50–60, 2017.

[14] L. Amelio and A. Amelio, "Classification methods in image analysis with a special focus on medical analytics," in Machine Learning Paradigms, pp. 31–69, Springer,Basel, Switzerland, 2019.

[15] Madhu, and Raman Kumar. "A hybrid feature extraction technique for content- based medical image retrieval using segmentation and clustering techniques." Multimedia Tools and Applications 81, no. 6 pp: 8871-8904, 2022.

[16] Bhaumik, Hrishikesh, Siddhartha Bhattacharyya, Mausumi Das Nath, and Susanta Chakraborty. "Hybrid soft computing approaches to content-based video retrieval: A brief review." Applied Soft Computing 46, 2016.

[17] Palai, Charulata, Pradeep Kumar Jena, Satya Ranjan Pattanaik, Trilochan Panigrahi, and Tapas Kumar Mishra. "Content-Based Image Retrieval using Encoder based RGB and Texture Feature Fusion." International Journal of Advanced Computer Science and Applications 14, no. 3, 2023.

[18] Reddy, Tatireddy Subba, K. Sanjeevaiah, Sajja Karthik, Mahesh Kumar, and D. Vivek. "Content-Based Image Retrieval Using Hybrid Densenet121-Bilstm and Harris Hawks Optimizati Algorithm." International Journal of Software Innovation (IJSI) 11, no. 1, pp: 1-15, 2023.

[19] Kusi-Duah, Samuel, Obed Appiah, and Peter Appiahene. "An Improved and Efficient Content-Based Medical Image Retrieval Technique vs State-of-Art Texture Feature Extraction Techniques." Journal of BioMed Research and Reports 2, no. 4, 2023.

[20] Devulapalli, Sudheer, Anupama Potti, Rajakumar Krishnan, and Md Sameeruddin Khan. "Experimental evaluation of unsupervised image retrieval application using hybrid feature extraction by integrating deep learning and handcrafted techniques." Materials Today: Proceedings ,2021.

[21] Alsmadi, M.K., "Content-Based Image Retrieval Using Color, Shape, and Texture Descriptors and Features." Arabian Journal for Science and Engineering, pp.1-14, 2020.

[22] Chhabra, P., Garg, N.K. and Kumar, M., "Content-based image retrieval system using ORB and SIFT features." Neural Computing and Applications, 32(7),pp.2725-2733, 2020.

[23] Wang, Fengyuan, Jianhua Lv, Guode Ying, Shenghui Chen, and Chi Zhang. "Facialexpression recognition from image based on hybrid features understanding." Journal of Visual Communication and Image Representation 59, pp: 84-88, 2019.

[24] Singh, Aman, Amit Dixit, and Brajesh Kumar Singh. "Hypertuned Convolutional Neural Network Residual Model Based Content-Based Image Retrieval Syste International Conference on Fourth Industrial Revolution Based Technology and Practices (ICFIRTP), pp. 139-144. IEEE, 2023.

[25] Khan, Umer Ali, Ali Javed, and Rehan Ashraf. "An effective hybrid framework for content-based image retrieval (CBIR)." Multimedia Tools and Applications 80, 2021.

[26] Ansari, Aasif, and Muzammil H. Mohammed. "Content-based video retrieval systems- methods, techniques, trends, and challenges." International Journal of Computer Applications 112, no. 7, 2015.

[27] Hamed Nassar , Ahmed Taha , T M Nazmy, Khaled Ahmed Nagaty. "Retrieving of video scenes using arabic closed-caption." International Journal of Intelligent computing and Information Systems (IJICIS), pp: 191-203, 2008.