E-ISSN 2581 – 7957
P-ISSN 2277 – 3916

Extensive Content Feature based Image Classification and Retrieval using SVM

A. Srinivasa Reddy
Assoc. Professor, CVR College of Engineering/CSIT Department, Hyderabad, India
Email: srinivas.asr@gmail.com

Abstract: **The classification and retrieval of picture advances in the field of image retrieval, particularly content-based image retrieval, are presented in this work. Scaling Invariant Feature Transform (SIFT) and developed K-Means clustering approach can be used to first arrange the features extracted based on the bag of visual words (BOW). The two stages of our retrieval method are retrieval and classification. The k-Nearest Neighbor (kNN) and Support Vector Machine (SVM) techniques were used to classify the photos based on their attributes and results were compared. This will categorize the images into different groups to improve the precision and recall rate. Following image classification, similar images matching the query image are pulled from the appropriate class.**

Index Terms: **Bag of visual Words, Support Vector Machines, k-Nearest Neighbor, *Scaling Invariant* Feature Transform, classification, retrieval, classification.**

I. Introduction

Searching for digital images in huge databases is known as the image retrieval problem. Content-based image retrieval, also known as query by image content (QBIC) applies computer vision techniques to this problem for a scientific overview of the CBIR field [1,2]. "Content-based" refers to a search that examines the contents of the image rather than its associated metadata, such as its keywords, tags, or descriptions. Colors, shapes, textures, and any other information that can be inferred from the image itself are all examples of "content" in this context. CBIR is preferable since searches that only employ metadata rely on the accuracy and comprehensiveness of the annotations.

Toshikazu Kato, an engineer at the Japanese Electrotechnical Laboratory, is credited with coining the phrase "content-based image retrieval" in 1992 to describe research involving the automatic retrieval of photos from a database based on the colours and forms present. The methods, devices, and algorithms are derived from statistics, pattern recognition, signal processing, and computer vision, among other disciplines.

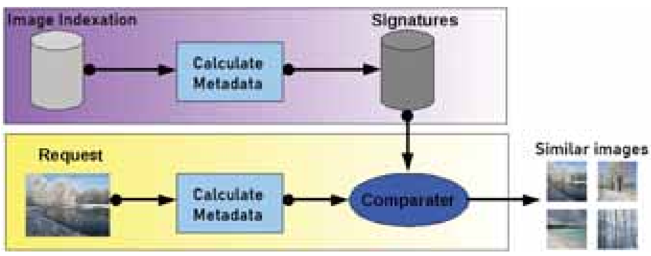

Figure 1. General CBIR scheme

The image metadata or keywords that are related to the visual content or properties of the picture file are what the annotation-based retrieval is dependent on. An automatic image annotation process assigns significant words to an image while taking into account its content. This method is predicated on the idea that zooming is significant enough to allow for the indexing, retrieval, and comprehension of enormous image data collections. Breast density texture characterisation effectively makes use of the two-dimensional principal component analysis in order to characterise the texture effectively while allowing for dimensionality reduction. The retrieval function is performed using a support vector machine [3]. A novel approach to interactive image segmentation-based content-based picture retrieval with relevance feedback has been developed [4]. It is based on the random walker algorithm.

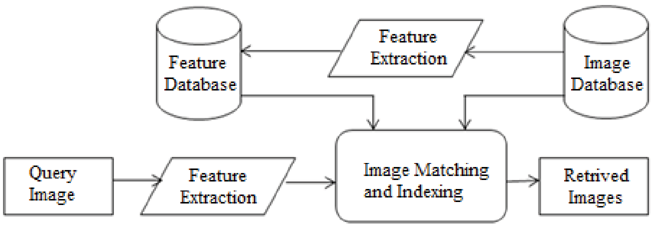

Figure 2. CBIR Model

The technique used to extract components from given images is a critical step in CBIR, and how well it works depends on how highlights are isolated from the images. Two images' similarity is calculated as a percentage of the discrepancies between their element vectors. The goal of CBIR [5,6] is to recover images from a query image that have photos with comparable visual elements, which has proven to be a challenging task in the realm of computer vision and artificial intelligence. The regular CBIR can be expanded by the creative method, much like how the content-based recovery is. Here, only visual highlights are used; access to substantive information is not available.

With the use of machine learning (ML), which is a form of artificial intelligence (AI), software programmes can predict outcomes more accurately without having to be explicitly instructed to do so. To forecast new output values, machine learning algorithms use historical data as input. Support vector machines (SVMs) [3,7] are a group of supervised learning techniques for classifying data, performing regression analysis, and identifying outliers. Support vector machines' benefits include efficiency in high-

Received on 20.08.2022, Revised on 19-10-2022, Accepted on 10-12-2022.

CVR Journal of Science and Technology, Volume 24, June 2023
DOI: 10.32377/cvrjst2409

dimensional environments. Still useful in situations where the number of dimensions exceeds the number of samples.

II. LITERATURE REVIEW

Stefanos Vrochidis et al. [1] mentioned that "to address the demand for content-based patent picture search and retrieval, we first explore the possible advantages, the needs, and the obstacles associated with patent image retrieval. Then, we present a framework that includes cutting-edge image analysis and indexing algorithms. To successfully facilitate content-based image retrieval in the patent domain, the proposed system applies document image pre-processing, image feature, and textual information extraction. We put in place a patent picture search engine to gauge the potential of our idea. Results from a variety of interaction modes, comparisons with other systems, and a quantitative assessment of our engine show that image processing and indexing technologies are at a stage of development where they are ready to be included into practical patent retrieval applications".

Srinivasa Rao et al. [2] stated that "There are Moment Invariants (MI) and Zernike Moments (ZM)-based Content Depending Picture Retrieval (CBIR) systems that use invariant image moments based on shape. The shape features of an image can be effectively represented by MI and ZM. However, their use in CBIR is limited by the MI's non-orthogonality and the ZM's poor reconstruction. Therefore, a CBIR system based on orthogonal moments must be effective. Legendre Moments (LM) can compactly express the features of an image's shape and are orthogonal and computationally faster. This paper proposes a CBIR system for grayscale images utilizing Exact Legendre Moments (ELM). The suggested CBIR system outperforms the MI and ZM moment-based approaches in terms of retrieval effectiveness and retrieval time. Using the Support Vector Machine (SVM) classifier also increases the efficiency of classification. Over the traditional CBIR technique, better retrieval outcomes are obtained using the Stacked Euler Vector (SERVE) in conjunction with Modified Moment Invariants (MMI)".

Srinivasa Reddy A. et al. [3] discussed that "One of the methods for detecting tumors in any part of the body is magnetic resonance imaging (MRI). The brain tumor is becoming one of the leading causes of mortality for many people. A brain tumor is among the deadliest malignancies, so it is important to find it quickly and get the right therapy to save a life. Due to the development of tumor cells, the detection of these cells is a challenging issue. Comparison of the MRI treatment for a brain tumor is crucial. Using basic imaging techniques, it is quite challenging to see the aberrant brain structures. To solve a problem, automated methods for classifying and detecting brain tumors are suggested in this work".

Nidhi Singh [4] et al. proposed that "It addresses the issue of content-based picture retrieval in dynamic environments. Systems that evaluate photos in real-time cannot function in a situation where new or additional images are constantly being added or stored. The authors of this research suggest a system that may choose the best features to evaluate recently acquired photos, increasing retrieval efficiency and accuracy. Here, a better algorithm is suggested. After segmentation, the process entails creating feature vectors that will be used to compare query photos to database images for similarity. The framework has been trained on several database pictures. When compared to the performance of traditional methods of content-based picture retrieval, the suggested algorithm's performance on a variety of real photos is found to be rather satisfactory".

III. CONTENT BASED IMAGE RETRIEVAL

In 1992, T. Kato coined the phrase "content-based picture recovery" to describe his work with database-based automatic image retrieval. In the CBIR, images are retrieved based on colour and shape [1,2,4]. Since then, the method of utilizing linguistic image qualities to extract desired photographs from a large collection has been referred to as CBIR. Among the fields where methods, tools, and calculations are applied are insights, design recognition, signal preparation, and computer vision. As it is based on the visual analysis of contents that are present in the query image, content-based image retrieval (CBIR), a framework, can get around the issues outlined above. Numerous uses of the CBIR technology have been developed, including fingerprint identification, biodiversity information systems, digital libraries, crime prevention, medical research, and history.

While CBIR does have relatively high accuracy on common benchmark datasets, some jobs, such datasets for trademarks, still have low accuracy. Unsolved issues include partial similarity, semantic similarity, domain variance, and resistance against adversarial attacks.

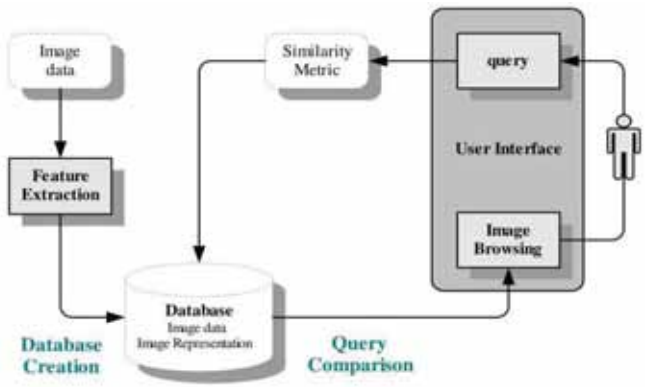

Figure 3. Architecture of CBIR

Using an image distance measure is the most popular technique for comparing two images in content-based image retrieval. An image distance measurement examines how similar two photos are in terms of their shape, colour, texture, and other attributes. For instance, 0 denotes a perfect match with the query in terms of the dimensions considered. A score greater than 0, as one might guess, denotes varying degrees of similarity between the photos. Then, you can arrange search results according to how close they are to the image you are looking for.

The semantic gap is the discrepancy between the low-level features provided by CBIR systems and the user's required semantic categories. The semantic gap between the low-level visual features (color, shape, texture, etc.).

A. Color

It is possible to calculate distance measurements based on colour similarity by creating a colour histogram for each image that shows the percentage of pixels that have a particular value. One of the most popular methods is examining photos based on the colours they contain because it can be done regardless of the size or orientation of the image. Research has, however, also tried to break down colour proportion by region and by the spatial relationships between various colour zones.

B. Texture

Measures of texture look for visual patterns in images and how those patterns are defined spatially. Depending on how many textures are found in the image, Texel's - which represent textures - are divided into a variety of sets. These sets not only specify the texture but also the location of the texture within the image. Additional techniques for categorizing textures include, Matrix of co-occurrence, Wavelet transform, and laws texture energy.

C. Shape

Shape does not refer to the shape of an image but to the shape of a particular region that is being sought out. Shapes will often be determined first applying segmentation or edge detection to an image. Other methods use shape filters to identify given shapes of an image. Shape descriptors may also need to be invariant to translation, rotation, and scale. Some shape descriptors include Fourier transform and Moment invariant.

D. Image Retrieval

In the context of image retrieval, the terms Accuracy, Recall and Precision are described in terms of a set of recovered images, a collection of relevant images, or a list of all online pictures that are pertinent to a particular image, and a set of pictures acquired by a web search engine for a query.

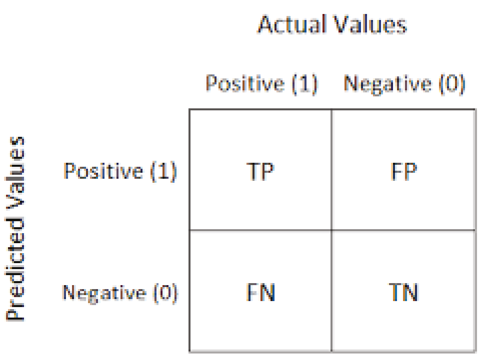

Figure 4. Confusion Matrix

True Positive:	Actual-Yes	Predicted-Yes
True Negative:	Actual-No	Predicted-No
False Positive:	Actual-No	Predicted-Yes
False Negative:	Actual-Yes	Predicted-No

Accuracy

Accuracy reveals how frequently the ML model was overall correct.

$$\text{Accuracy} = (TP+TN) / (TP+TN+FP+FN) \quad (1)$$

Precision

The model's precision measures how well it can forecast a particular category.

$$\text{Precision} = TP / (TP+FP) \quad (2)$$

Recall

How frequently the model was able to identify a particular category is indicated by recall.

$$\text{Recall} = TP / (TP+FN) \quad (3)$$

IV. Support Vector Machine

Support vector machines (SVMs) are a collection of supervised learning techniques used to classify images and identify outliers. The SVM algorithm's objective is to establish the best line or decision boundary that can divide n-dimensional space into classes, allowing us to quickly classify fresh data points in the future.

A hyperplane is the name given to this optimal decision boundary. SVM selects the extreme vectors and points that aid in the creation of the hyperplane. Support vectors, which are used to represent these extreme instances, form the basis for the SVM method. Consider the diagram below, where a decision boundary or hyperplane is used to categorise two distinct categories.

SVM comes in two varieties:

Linear SVM: Linear SVM is used for data that can be divided into two classes using a single straight line. This type of data is called linearly separable data, and the classifier employed is known as a Linear SVM classifier.

Non-linear SVM: Non-Linear SVM is used for non-linearly separated data. If a dataset cannot be classified using a straight line, it is considered non-linear data, and the classifier employed is referred to as a Non-linear SVM classifier.

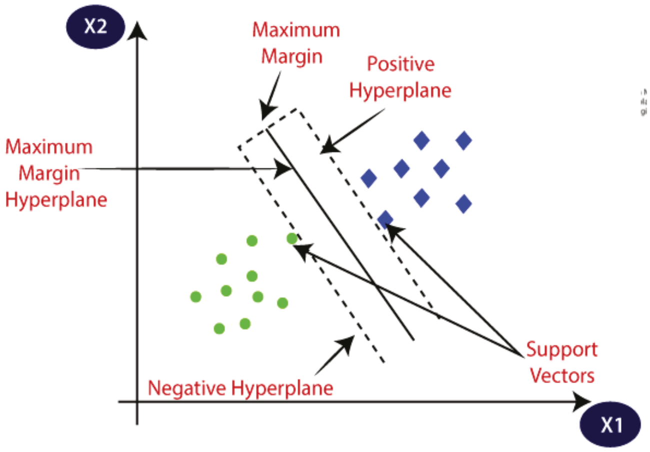

Figure 5. Architecture of SVM

E-ISSN 2581 – 7957 CVR Journal of Science and Technology, Volume 24, June 2023
P-ISSN 2277 – 3916 DOI: 10.32377/cvrjst2409

In n-dimensional space, there may be several lines or decision boundaries used to divide classes; however, the optimal decision boundary for classifying the data points must be identified. The hyperplane of SVM is a name for this optimal boundary.

V. Proposed Methodology

A computer system for browsing, searching, and retrieving images from a database of digital images is called an image retrieval system. The two methods used for searching and retrieving images from an image database are context-based image retrieval and content-based image retrieval. Content-based image retrieval is a system for retrieving different images from an image database. Most web image retrieval solutions now use text-based image retrieval. A keyword-based search is used in the text-based approach. We use a high-level feature as a label for getting the image based on the names in context-based image retrieval. Our recommended approach consists of two essential components, namely: Process of feature extraction and retrieval.

There are two types of feature extraction: textual feature extraction and visual feature extraction. An example of a visual feature is Bag of Visual Words. text-based features such as filenames, notes, and keywords. By using our recommended technique, the features are extracted from the Bag of Visual Words (BoW). The BoW includes feature description, code book production, and point of interest detection. This graphic piece could be a representation of an image in a histogram. Scaling Invariant Feature Transform (SIFT) and the newly created K-means clustering algorithm can be used to frame the BoW.

A better multi-texton approach can be used to extract the texture. Following the feature extraction, the retrieval technique will be used. The two stages of our retrieval procedure are categorization and retrieval. Applying the kNN algorithm will be the primary method of classifying the photos based on their attributes. To improve the precision and recall rate, this will divide the photos into several classes. The performance of the recommended picture retrieval is assessed based on the precision, recall, and accuracy values after the classifications of the similar images, which are obtained from the pertinent class in accordance with the provided query image.

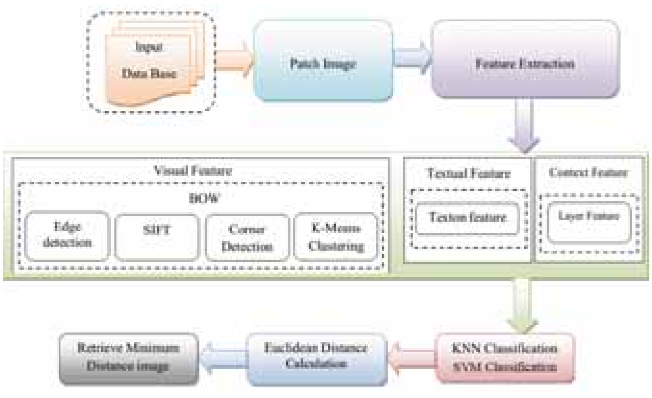

Figure 6. Architecture of Proposed method

A. Feature Extraction

By removing picture features, content-based images retrieval (CBIR) was launched. In general, the features are divided into visual elements like colour and texture, to mention a few, as well as text-based features like annotations and keywords. A lot of information about a picture's content can be found elsewhere outside the image itself. The context of an image can be thought of as all information about the image that isn't related to its visual qualities. The context for retrieval fundamentally alters in this case of a considerably larger collection of photographs because the feature space is suddenly more heavily loaded with images. A high level, layer-based feature is the context feature.

B. Texture Images

Unevenness is visible as tone or intensity fluctuations over a neighborhood in an intensity image. A recurrent array of a specific essential model comes from the relative change in most instances. Large-scale fundamental models are mostly responsible for the coarser texture. The basic models of minute size always represent the finer texture in a similar way. The spotlight has been shone on a feast of creative techniques meant to coerce the texture features in this way. There are many different types of texture feature extraction techniques, such as spectral texture feature extraction techniques and spatial texture feature extraction techniques, depending on the domain from which the feature is extracted.

C. Multi-Texton Features

Textons are generally portrayed as a collection of blobs or embryonic patterns that exhibit a consistent character across the image, while a specific definition of textons is not yet possible. The pre-attentive distinction is somewhat diminished if the texture pieces are stretched out very far in one orientation. The texton-gradients at the texture boundaries are improved if the stretched-out elements are not jittered in orientation. Because the texture gradients only exist at the texture boundaries, the texture discrimination can be increased with a small element dimension like this. This, together with the block's simplicity of expression, is why textons detection is used in this page.

D. Feature Extraction from Visuals

The visual search is a type of conceptual function that invites attention and traditionally permits an active scan of the visual scene for the goal amid other objects or features (the distractors). We first extort the Bag of visual words as the feature extraction from the photos (BOW). Additionally, we implement this functionality using the SIFT and K-Means techniques.

A group of mathematical approaches known as edge detection are focused on locating the locations in digital images where there are sharp brightness variations or, more precisely, breaks. Edges are likely integrated by the detection of brief linear edge segments and the accumulation of edges into prolonged edges. There are numerous techniques to position the edges, including the Laplacian Roberts, Sobel, and gradient.

The k-Nearest Neighbors algorithm is widely regarded as a non-parametric method for classification and regression. The most common distance function is the Euclidean distance, which is how people typically think about distance in the actual world:

$$d_{Euclidean}(x, y) = \sqrt{\sum_i (x_i - y_i)^2} \quad (4)$$

When measuring the distance, some characteristics with high values, like income, can outweigh the influence of other features that are assessed on a smaller scale. When dealing with continuous data, min-max normalization or Z-score standardization can be used successfully.

Min–max normalization:

$$X^* = \frac{X - min(X)}{range(X)} = \frac{X - min(X)}{max(X) - min(X)} \quad (5)$$

Z-score standardization:

$$X^* = \frac{X - mean(X)}{SD(X)} \quad (6)$$

$$F = 2 * \left(\frac{precision * recall}{precision + recall}\right) \quad (7)$$

VI. Results and Discussion

Based on the assessment criteria, such as Precision, Recall, and Accuracy, the performance of our proposed image retrieval and classification job is assessed. The outcomes of linked recovered photos are created for the test images made available.

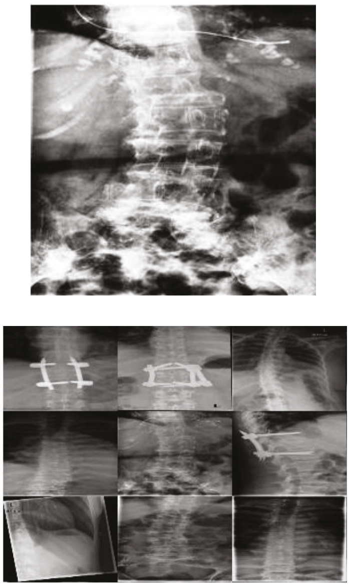

Figure 7. Backbone Input and Retrieved Images

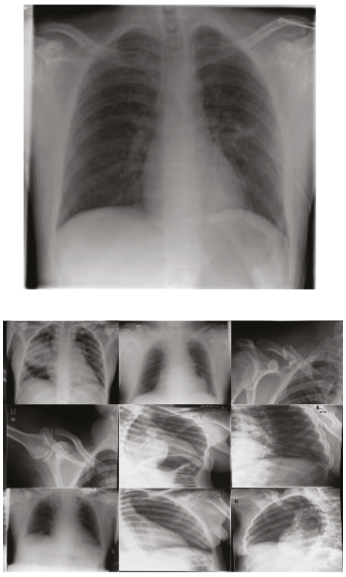

Figure 8. Chest Input and Retrieved Images

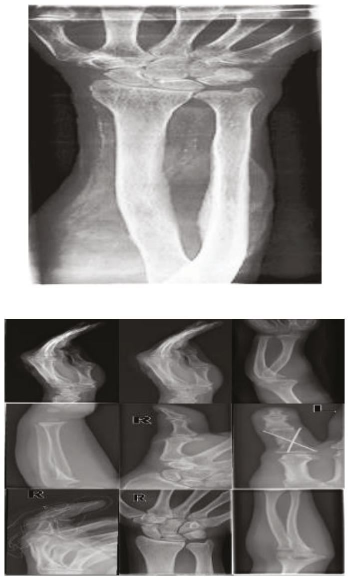

Figure 9. Hand Input and Retrieved Images

TABLE I.
PRECISION

	Hand Bone	Back Bone	Chest Bone
SVM	94.2	93.7	93.3
Naïve Bayes	82.4	87.5	86.6
KNN	92.8	90.5	91.8

TABLE II.
RECALL

	Hand Bone	Back Bone	Chest Bone
SVM	92.1	91.7	90.3
Naïve Bayes	87.2	85.5	87.3
KNN	90.5	89.2	88.7

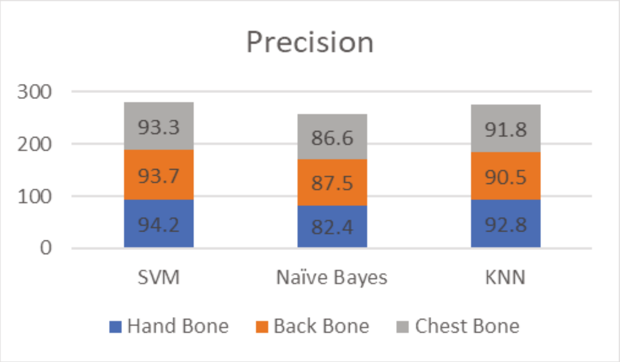

Figure 9. Precioson for SVM, NB, and KNN

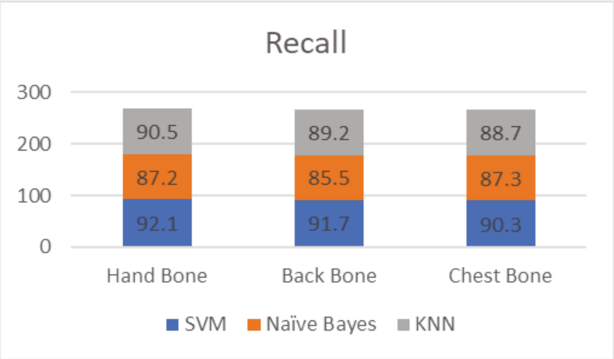

Figure 10. Recall for SVM, NB, and KNN

VII. Conclusions

The two steps implemented are feature extraction and picture retrieval. The evaluation results of our recommended SVM method have shown that it is the most effective. We have applied the SVM algorithm on average to get 93% accuracy. Additionally, a comparison between the Naïve Bayes and KNN was done. Our suggested SVM image retrieval and classification algorithm outperforms the competition, according to analysis from the comparison. The MATLAB platform was used to implement our method. The proposed method achieves greater precision values when compared to existing methods. The same can be implemented on Google Colab environment for more accurate results while considering more data sets.

References

[1] Stefanos Vrochidis, Symeon Papadopoulos, Anastasia Moumtzidou, Panagiotis Sidiropoulos, Emanuelle Pianta and Ioannis Kompatsiaris, "Towards content-based patent image retrieval: A framework perspective", Elsevier Journal of World Patent Information, pp.94-106, 2010.

[2] Ch.Srinivasa Rao, S.Srinivas Kumar and B.Chandra Mohan, "Content-based Image Retrieval Using Exact Legendre Moments And Support Vector Machine", The International Journal of Multimedia and Its Applications, Vol.2, No.2, May 2010.

[3] A. Srinivasa Reddy, P. Chenna Reddy, MRI brain tumor segmentation and predict ion using modified region growing and adaptive SVM, Soft Comput. 25(2021) 4135–4148.

[4] Nidhi Singh, Kanchan Singh and Ashok K. Sinha, "A Novel Approach for Content Based Image Retrieval", Elsevier Journal of Procedia Technology, pp.245-250, 2012.

[5] A. Srinivasa Reddy, Effective CNN-MSO method for brain tumor detection and segmentation, Materials Today: Proceedings, Volume 57, Part 5, 2022, Pages 1969-1974.

[6] Sahith, R. and Reddy, P. V. P. (2019). Decision tree based machine learning algorithms to classify rice plant diseases. International Journal of Innovative Technology and Exploring Engineering, 9(1): 5365–5368.

[7] Reddy, A.S. and P.C. Reddy. Novel Algorithm based on Region Growing Method for Better Image Segmentation. in 2018 3rd International Conference on Communication and Electronics Systems (ICCES). 2018. IEEE.

[8] G Malleswari, "Significance of Genetic Algorithms in Image Segmentation", International Journal of Signal Processing, Image Processing and Pattern Recognition,9(4), 177-184.

[9] G. Malleswari and A. S. Reddy, "Diverse Convolutional Neural Network Models for Feature Extraction from Brain Tumor Images," 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2023, pp. 405-410.

[10] A. S. Reddy and G. Malleswari, "Adaptive Energy Routing Protocol using Spider Optimization in Wireless Sensor Networks," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-6.