

## OPTIMIZED CROSS-DOMAIN MEDICAL IMAGE RETRIEVAL VIA MULTIMODAL FEATURE FUSION AND DEEP LEARNING- DRIVEN COMPUTATIONAL INTELLIGENCE

**Hareem Ayesha**

*Institute of Computer Science and Information  
Technology, The Women University Multan,  
Pakistan.*

**Muhammad Ans khalid**

*Department of Computer Science and  
Information Technology University of Southern  
Punjab Multan.*

**Muhammad Tanveer Meeran\***

*Faculty of Computer Science and Mathematics,  
Universiti Malaysia Terengganu, Malaysia.*

**Sami Ullah**

*Department of Computer Science , University of  
Hull, London.*

*\*Corresponding author: Muhammad Tanveer Meerean ([Tanveer\\_meeran@yahoo.com](mailto:Tanveer_meeran@yahoo.com))*

### Article Info



This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

<https://creativecommons.org/licenses/by/4.0>

### Abstract

Content-Based Image Retrieval (CBIR) has emerged as a pivotal research domain in computer vision, driven by the exponential growth of visual data and the demand for efficient retrieval systems that transcend traditional text-based methods. While existing tools leverage audio-visual content to navigate large-scale media databases, challenges persist in handling ever-expanding datasets, particularly in specialized fields like medical imaging. Medical images, acquired through diverse modalities (e.g., MRI, CT, X-ray), require precise modality identification to refine diagnostic workflows and enhance search accuracy. This study addresses this need by proposing a robust framework for modality classification and retrieval of medical images, leveraging advanced visual feature fusion and machine learning. Our approach integrates seven complementary visual features to capture texture, edge, and color characteristics: Scale-Invariant Feature Transform (SIFT), Local Binary Patterns (LBP), Local Ternary Patterns (LTP), Edge Histogram Descriptor (EHD), Color and Edge Directivity Descriptor (CEDD), wavelet-based color edge detection, and color histograms. These features are fused into a unified feature vector, enabling comprehensive representation of image content. The framework was evaluated on the ImageCLEF2012 modality classification dataset, a benchmark comprising 31 distinct modality classes. For classification, a Support Vector Machine (SVM) with a chi-square kernel was employed, chosen for its effectiveness in handling high-dimensional, non-linear data. The proposed system achieved an overall accuracy of 72.2%, surpassing the top ImageCLEF2012 visual feature-based result by 2.6%. This improvement underscores the efficacy of feature fusion in capturing discriminative patterns across modalities. Key innovations include the integration of wavelet-derived edge features with texture descriptors and the strategic use of a chi-square kernel to optimize similarity measurement in SVM decision boundaries. This work advances CBIR in medical imaging by demonstrating that hybrid feature engineering, combined with tailored classifier design, can significantly enhance modality identification. Future directions include exploring deep learning-based feature extraction and expanding the framework to multi-label classification scenarios.

### Keywords:

*Digital Image, Machine Learning, Edge Histogram Descriptor, Scale-Invariant Feature transform.*

## 1. Introduction

In every era, images are used as a wellspring of information. Images are the artifacts that are used to portray or records the pictorial observation of an era. They are used to freeze a period of life some images discovered from Mohenjo-Daro archeological site, these pictures clearly shows that images were used to represent knowledge even in ancient age. Images are used to demonstrate certain solid item, as well as expected to express certain inclination that can't be passed on by some other implies. Languages are synthetic and may get the opportunity to be incomparable after some time, however, images are common and natural. That is the reason we can't read what was composed in Mohenjo-Daro pictures yet can make out what were the creatures lived in that period of time.

A standout amongst the most brilliant things about the image is that they permit individuals to communicate crosswise over hindrances, for example, language obstructions. Analyst Albert Mehrabian exhibited in his study [2], that 93% of correspondence is nonverbal. The retina of the human eye is truly an augmentation of the human brain. The human brain can comprehend images much quicker than whatever other source. Studies demonstrate that human mind unravels image components concurrently, while languages are interpreted in a successive way utilizing more time. The response of human mind to a visual stimulus is very diverse.

Concepts and ideas are easy to express with the image. It is easy to show a square rather expressing in the words. Using the textual description, there is a need to decode. The users easily understand the ideas and concepts by just looking image, rather than reading and decoding the textual explanation. Medical imaging is the system and strategy of making pictorial illustrations of within a body for clinical examination and therapeutic intervention, furthermore visual representation of the limit of a couple of organs or tissues. These images uncover the inside structures shielded up by skin and bones and also to analyze the remedies. Medical images are acquired using three different techniques.

Content-Base retrieval (CBIR) systems focus on the content of image rather than focusing on text associated with the image, contents include color, texture and other low-level feature of the image [6]. CBIR is appealing on the grounds that most online web crawlers depend absolutely on metadata [7]; this creates a considerable measure of Junk in the outcomes. Likewise having people manually entering keywords is an expensive, ineffective and may not catch every word that portrays an image. The goals of CBIR is to retrieve images that are semantically related and decreasing human interference during retrieval stage [8], the major tasks of any CBIR system are the identification of the correct features also called image signature and resemblance calculation mechanism that sets the visually related image are vital land marks of a CBIR system.

While working with the medical images the modality identification plays a vital role. In medical Sciences any of the dissimilar sorts of equipment or tests used to acquire an image of the body part is call modality of acquired image, study exhibited that the modality base retrieval strategies provide key information for image retrieval [9]. Modality identification filters out the result to restrict the results of retrieval to a specific modality level.

Traditional retrieval systems for images were based on text, images were annotated with some text that describes the objects and modality of the image in the form of a caption, and this caption was used to identify the modality of the image. With the development of gigantic image databases, conventional text-based retrieval systems undergo from several limitations such as high cost of annotation and the subjects

of medical images is often challenging to appropriately define in words. Many test and online collections do not contain textual information about the imaging modality. Although many image retrieval websites (Goldminer, Yottalook) permit users to limit the search results to a particular modality. But, this modality is usually extracted from the caption which is often not correct or present. Studies have shown that the modality can be extracted from the image itself using visual features. Hence, there is a strong need for an automatic method, which can classify the modality of given image on the basis of visual features.

## 2. Related work

ImageCLEF is an assessment campaign and organized as part of the CLEF Initiative labs launched in 2003, it supports the domain of visual media examination, recovery and characterization by working on imperative frameworks for appraisal of visual media recovery systems. The ultimate objective of imageCLEF is to support the field of image retrieval and classification. Kalpathy-Cramer et al. investigate the retrieval techniques used in last ten assessment of imageCLEF from 2004 to 2013[9].

Muller et al. present an overview of imageCLEF2012 image retrieval task [11] for the 9<sup>th</sup> edition of imageCLEF. The task was subdivided into three subcategories, Modality Classification, Image-based and case-based, where modality classification works as a filter to restrict the possible outcomes, for modality classification new hierarchy was introduced which is discussed in detail in Chapter 4. The dataset for modality classification and retrieval was divided into 31 distinct modalities of diagnostic images; which consist of five subcategories, compound and Multi-pane images and generic biomedical illustration. The tasks aim to retrieve images that ensemble the provided case depiction. The retrieval results can be more accurate and filtered by incorporating modality information in the retrieval process. Cramer et al. proposed two modality classification models for grayscale and color images [12]. The input to the model was a feature vector extracted by using low-level feature extraction technique. In feature extraction process image was first resized, afterward, the image was divided into five overlapping regions and GLCM [13] are calculated for every block, a feature vector with 132 dimensions for every image is formed. For classification technique, Multi-layer Perceptron [14] was used based on neural network, in which hidden layer contains 50-150 neural cells or neural nodes. For evaluation of the model, the CISMef and the ImageCLEF 2006 dataset were used. Both classifiers achieved an accuracy of more than 95 %.

Wei et al. present their contribution for BUAA AUDR research group in imageCLEF 2012 medical retrieval and classification chores [15]. The researchers concentrated on mono-modular and visual based classification. For the feature extraction, they utilized edge histogram descriptor [16]. For the classification purpose SVM [17] was utilized by adopting one vs. all approach to enhance the execution of the retrieval framework. The maximum accuracy achieved by the group using visual features was 39% on imageCLEF 2012 modality classification dataset. The major cause of such low accuracy was the use of single feature extraction technique, single feature extraction technique does not cover all the visual aspects of image and does not provide enough visual information which helps in modality classification of the image.

Liangliang et al. present the demonstrating systems that were connected by the IBM T.J. Watson research group to the modality classification and case-based retrieval undertakings of ImageCLEF 2012 [18]. They adopted two different strategies. The first strategy was to supplement the training examples with an extra case of every class; the second strategy was to try out with different feature extraction techniques. They made seven different set of features and experiments were carried out on these feature sets and with

different level of granularities. For multi-class classification, LIB-SVM with chi-square kernel was used and set of one vs.one classifier were learned. The maximum accuracy observed was 69.9% on the augmented dataset and it was the maximum accuracy achieved in imageCLEF 2012 modality classification task at that time. Although authors extracted and experimented with a different set of features, but the proposed technique is a lack of feature reduction process, because a combined feature vector may subject of redundancy, feature reduction enhances generalization by reducing overfitting and eliminating redundancy from combined feature vector, and removal of multicollinearity may improve the performance of the classification model.

Fatemeh proposed a new set of features that are mined from transform domain for radiographic images [19]. This technique comprises of three steps. In the first step, images were preprocessed in order to remove noise keeping in mind the end goal to acquire center-symmetric local binary features; every image is apportioned into 25 intersecting sub-images. In the second step, Gabor transform of every sub-image was calculated and a novel version of local binary pattern LBP called center-symmetric local binary pattern was calculated. This proposed scheme is implemented on a subset of IRMA dataset from 15 different categories. Authors utilized two SVM [17] classifiers using polynomial and Gaussian kernels and results are presented in the form of accuracy rate.

Computer-aided diagnostic (CAD) is quickly entering the radiology standard. It is now turned into a business-as-usual clinical work for the recognition of cancer. These systems play a vital role in human health care systems. Agarwal et al. proposed a CAD system for early detection of lung cancer [20] by using Computed Tomography (CT) images of the chest. The proposed system works in three different phases. In the first phase, extraction of affected region is performed using region base extraction method [21]. In the second phase, the feature from affected region is extracted using gray level co-occurrence matrix (GLCM) [13] and 2D wavelet transform [22] and finally. In the last phase, the classification is performed whether cancer exists or not by using SVM [17] with polynomial kernel and above 80% accuracy was achieved.

The proposed system by Deserno et al. is another CAD tool for retrieval of mammography images by using CBIR [23]. They developed a classification tool that is able to classify the affected tissues pattern. The dataset was partitioned in 12 and 20 classes. Prior to feature extraction, the images were divided into the small patches of 128 x 128. Features were extracted using two-dimensional principal component analyses (PCA) [24]. For classification, they utilized support vector machine (SVM) [17]. For 12 and 20 class problem, an accuracy rate of 61.6 % and 52.1 % were obtained respectively.

Rafi and Fatemizadehi proposed a retrieval technique based on the content of image [25]. They present a contextual analysis, which depicts the approach of CBIR framework that is helpful for the retrieval of MRI images from huge databases, the approach is based on the texture of the image. The proposed model expects to spread information of CBIR way to deal with the utilization of image management and to separate between an ordinary, abnormal and MS image. A primary record is discovering and grouping of abnormal images into two certain categories. Gray level co-occurrence (GLCM) [13] used for the texture feature extraction, and then features are reduced using principal component analysis PCA [24]. For the classification process, support vector machine SVM [17] was used, which classify the image into one of three categories. The proposed model achieved 95% of accuracy.

Ultrasound images are widely used in health care system. These images are produced by a special machine that generates sound waves which are too high to be heard. These images are used to depict the interiors of the human body. Sohail et al. proposed a combined technique for medical image classification and retrieval for ultrasound Images for three types of ovarian cysts [26]. Gray level co-occurrence matrix (GLCM) [13] and histogram moment [28] proposed as features extraction techniques, for both retrieval and classification. In the case of retrieval, Gower's coefficient [29] was used, for similarity comparison between images and for the classification of images the fuzzy K-Nearest Neighbors (FKNN) [30] was used. The first 20 and 40 retrieved image revealed 75% and 20% average precision respectively. In the case of classification for three types of ovarian cysts 88.1 % average accuracy was achieved. Though technique displays very much better results in the terms of accuracy and precision, but this work focuses on a specific modality of medical images and has its particular characteristics for classification and retrieval.

Color and texture are important image features of the digital image and plays a vital role in image retrieval systems. Retrieval strategy proposes by Kavitha et al. based on local color and texture features [30]. The images were divided into sub-blocks prior to feature extraction. Texture features were calculated using GLCM [13] and color information were extracted using hue, saturation, and intensity (HSI) color model. The distance between the feature vectors was measured using minimum distance measure. This technique focuses on the color images, which favor for the particular medical domain. Since in specialized fields, specifically in the medical field, the absolute color and gray level features are very limited. Drawings are the essential segment of patients and numerous exploration tasks in licensed innovation space depend on the correlation of patent drawings. Hanbury et al. present an overview of retrieval of patent image retrieval [31], they in progress with an assessment of calculations produced for the computerized recapture of comparative pictures in the patent area. The research introduces a review of this exploration for specialized drawings, outlines, diagrams, plots also, substance structures. At long last, they examine the valuation including CLEF-IP and TREC-CHEM releases in 2011.

### **3. Methods and materials**

The fundamental applications of CBIR are medical analysis, research, and teaching. A substantial advantage originates from the territory of educating. Here instructors can utilize tremendous storages to locate some fascinating cases to show to students. These selected cases are Chosen cases are shown and construct in light as well as outwardly comparative images are introduced to improve the instructive abilities. And huge archives are also utilized by medical students for educational purpose. Another advantage is originated from research territory; the researcher can incorporate their case in their explorations. At last, the most critical and troublesome application region of CBIR is diagnostic; this leads to the combination of the system to the daily life. Content-based image retrieval (CBIR) system allows perusing the database and finding analysis based on visual resemblances. In this chapter, we are going to discuss proposed methodology and different techniques used in our framework.

#### **3.1. Proposed Framework**

The proposed framework consists of several stages which include feature extraction, feature vector construction, dimensionality reduction and training of classifier and finely the modality classification for the given query image. Figure 1 explains the working of proposed retrieval and classification framework.



3.2. Feature Extraction

When the input data to an algorithm is too large to be processed and it is suspected to be redundant, then it can be transformed into a reduced set of features (features vector). This process is called feature extraction. The selected features are expected to contain the relevant information from the input data so that the desired task can be performed by using this reduced representation instead of the complete initial data. Features for a digital image are distinct quantifiable properties. The selection of separating and autonomous features is a vital step for a successful CBIR system, the type of the features are generally numeric; however basic components, for example, strings and diagrams are utilized as a part of recognition and retrieval frameworks.

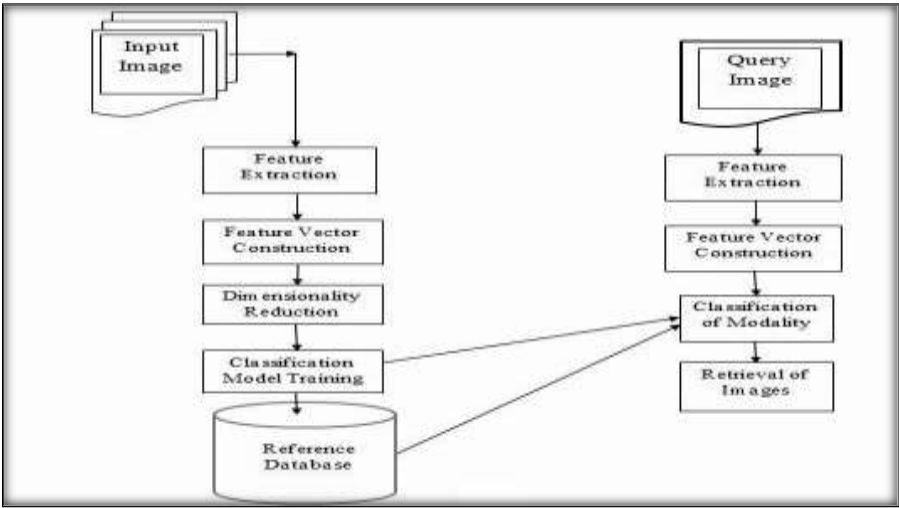


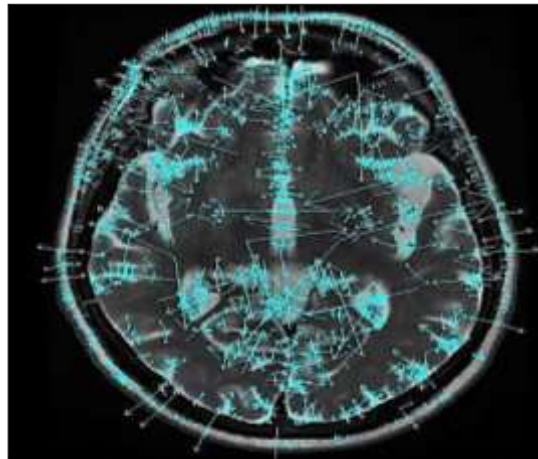
Figure 1. Block Diagram of Proposed Classification Framework

The idea of feature correlated to the controlled variable in the linear regression. The process of mining and selection of a feature is a blend of skill and science, the development of such framework to do so is known as feature engineering [32]; it requisites several experiments with various possibilities and grouping of many automated techniques and domain knowledge. Mechanizing this process is known as feature learning [33], where machine learns the feature by itself. In our framework we have used a set of features, our methodology can be portrayed as scaling up to use as many features and data as possible. The experiments exhibit, that expanding either axis have a tendency to increase performance. Features used in our framework are; scale invariant feature transform (SIFT) [34], local binary pattern (LBP)[35], local ternary pattern (LTP)[36], color edge directivity descriptor (CEDD)[37], color edge detector using wavelet transform [38], edge histogram descriptor (EHD)[39] and RGB histogram. Figure 3-1 Working of Proposed Classification and Retrieval Framework.

3.2.1. Scale Invariant Feature Transform (SIFT)

SIFT is a feature extraction technique which is used to extract local features from the image [34]. It gives an arrangement of features that are not affected by rotation and scaling of the object. An object can be recognized using SIFT features in several images of the similar place that are acquired from different

positions or viewpoints. Additionally, the technique is robust to noise [40]. Figure 2 displays the detected SIFT features from an image.



**. Figure 2. Detected SIFT Features from an Image [11].**

Because of some low contrast images (e.g. radiographs), it would be hard to utilize for the point of interest (detector). We extracted SIFT and Dense SIFT features from images. The obvious difference is that with dense SIFT you get a SIFT descriptor at every sampled location, while with normal SIFT you get a SIFT descriptions at the locations determined by Lowe's algorithm. DSIFT uses the descriptor in densely sampled locations. We use a fixed pixel interval of 6 pixels and the value of  $\sigma = 1.2$  for the Gaussian filter [41, 42, 43].

### **3.2.2. Bag of Visual Words on SIFT and DSIFT**

BoVW (Bag of Visual Words) generates a histogram of the visual word by counting the occurrences visual vocabulary in the image [44]. The SIFT feature extraction and applying BoVW on extracted descriptors is utilized in several state-of-the-art frameworks [18, 41]. The idea behind the approach is to extract the descriptors and then matching them with specified visual vocabulary and develops histogram. This is a three step process, feature detection, visual vocabulary formation and generation of a bag of visual words for images. In the first step of feature detection, we extracted SIFT and DSIFT features from the image, in the second step we performed clustering on extracted descriptors, the K-mean algorithm is applied to the descriptors, and each cluster center is considered as a visual vocabulary. In the third step, we used an approximation of nearest neighbor algorithm to form a feature histogram for every image in the collection. The algorithm then increases the histogram bin in the light of the closeness of the descriptor to a specific group center. The parameter to consider here is the size of codebook or visual vocabulary, the dimensions of the output feature histogram relate to this size. In the case of small size of vocabulary, dissimilar Keypoints are grouped in the same cluster, which makes the vocabulary not so discriminative. While a substantial size of vocabulary has a tendency to enhance performance, but it doesn't imply that a bigger vocabulary certainly prompts to a higher performance, studies shows that if we keep increasing the size of vocabulary the performance increases dramatically then peaks and after those levels of or drop mildly [45], which employees that an overlarge vocabulary not simply fabricates the computational load in clustering process, but also impacts contrarily on matching performance. A comparison of codebook

representation is presented by Hou et al. [46]. We experimented with different codebook size 200, 400, 600, 800, and 1000 and by combined histograms obtained by different codebook sizes.

### 3.2.3. Local Binary Pattern (LBP)

The fundamental thought behind the Local binary pattern (LBP) [35] methodology is to utilize the data about the surface from a nearby neighborhood. To begin with, we specify a radius  $R$ . The algorithm then forms a binary pattern or code that depicts the neighboring texture in the area set of  $P$  pixels, the binary pattern is generating by considering the center value as a threshold value. The obtain value is then converted to decimal by using Equation 1 and 2.

$$LBP(x_c, y_c) = \sum_{n=0}^7 m(g_n - g_c) 2^n \tag{1}$$

$$m(k) = \begin{cases} 1, & k \geq 0 \\ 0, & k < 0 \end{cases} \tag{2}$$

Where  $g_c$  represents the value of the center pixel and  $g_n$  represents the value of eight surrounding pixels and function  $m(k)$  returns a binary value. The length of LBP histogram depends on a number of neighbors. We utilized uniform pattern for experiments because it reduces the size of the histogram by combining the entire non-uniform pattern into a single bin. We experimented with the radius size of 1, 2, and 3 and setting the neighborhood size to 8, 16 and 24 respectively. The results of the experiments are outlined in Section 4.3.3.

### 3.2.4. Local Ternary Pattern (LTP)

Local binary pattern (LBP) has turned out to be exceedingly discriminative components for texture classification and they are impervious to lighting impacts as in they are invariant to gray-level transformation [36]. However, in light of the fact that they restrict at the center pixel value, they have a tendency to be sensitive to noise, particularly in close uniform and smooth regions.

$$m(k) = \begin{cases} 1, & \text{if } p > c + t \\ 0, & \text{if } p > c - t \text{ and } p < c + t \\ -1, & \text{if } p < c - t \end{cases} \tag{3}$$

LTP are the augmentation of LBP, in the case of LTP a user defines a threshold value  $T$ . and A ternary code is generated by using equation (3) for each pixel. When the ternary code is calculated the upper and lower patterns are computed using that ternary code. Fundamentally, any value that given -1 get allotted 0 for upper pattern and value that given 1 allotted out 1 for the lower pattern. Also, for the lower pattern, any values that are 1 from the original window get mapped to 0. We experimented with the radius size of 1, 2, and 3 and setting the neighborhood size to 8, 16 and 24 respectively. The results of the experiments are outlined in Section 4.3.4.



3.2.5. Color Edge Detection using Discrete Wavelet Transform

The color is a standout among the most critical low-level components of the image and utilized as a part of CBIR frameworks. Nevertheless, CBIR systems utilizing the only color may give exceptionally inadmissible results on the ground that image with comparative color does not have comparative contents. The feature extraction method described in [38] forms a feature vector by utilizing color edge detection and wavelet transform. The color information is extracted from the image by converting the image from RGB color space to YCbCr color space. In YCbCr color space Y component indicates luminance, Cb component denotes blue difference and Cr component signifies the red difference. The selection of the YCbCr color space is made on the ground that luminance part Y is free of color so can be embraced to tackle color variety. In the first step, the image is converted to YCbCr space after that canny edge detection [47] is applied to Y component and edge map is combined with Cb and Cr are combined to make a single RGB image.

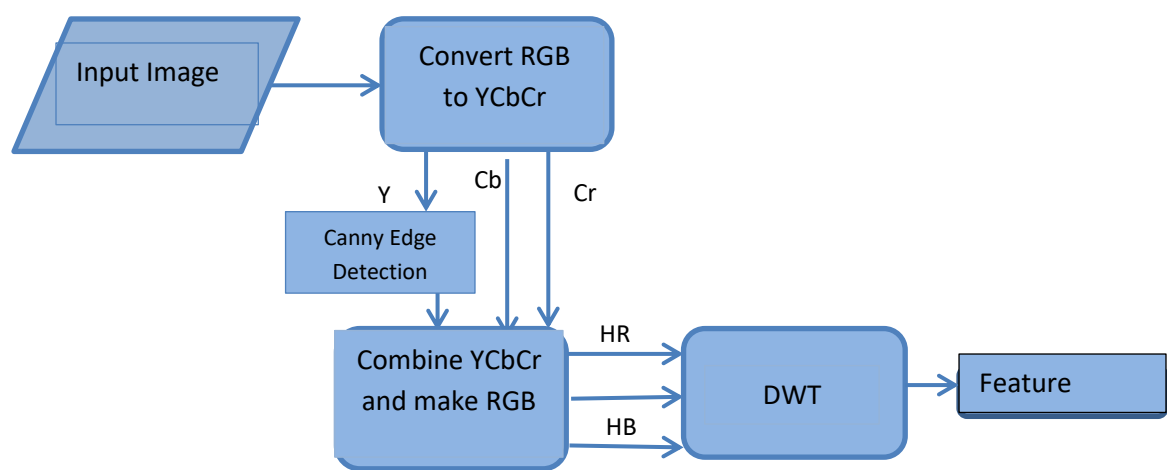


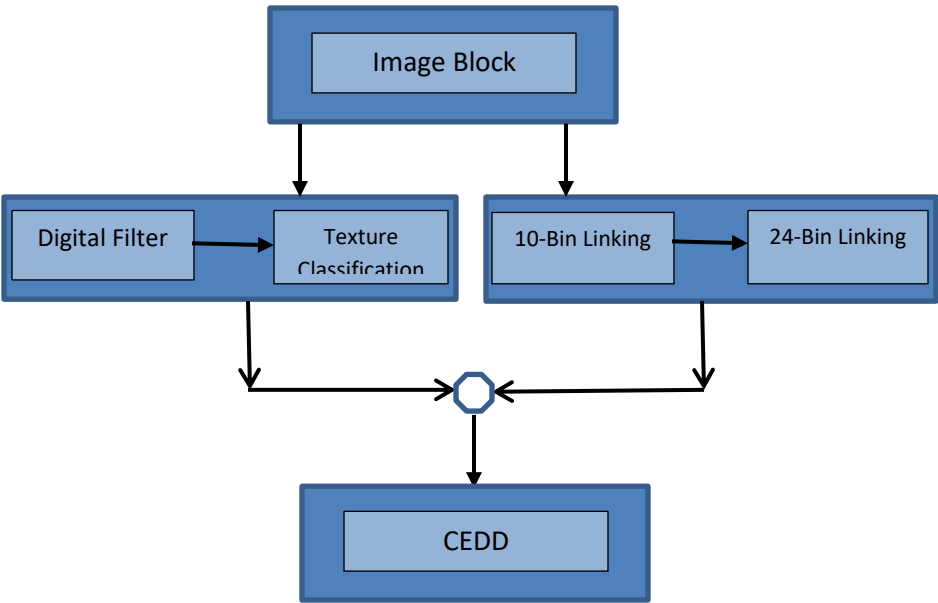
Figure 3. Description of Color-Edge Extraction with Wavelet Transform

In the second we have computed RGB values of the image which is obtained from step one and find the histogram of each matrix separately and call them HR, HG, and HB. In the third step, discrete wavelet transform is then applied of level two on HR and level three on HG and HB to reduce the size of the feature vector and preserving the content detail. Figure 3 depicts the working of the algorithm.

3.2.6. Color and Edge Directivity Descriptor (CEDD)

A color edge directivity descriptor (CEDD) [37] incorporates color and texture information in a single histogram. The algorithm consists of two different units, texture unit is responsible for the texture extraction and color unit is related to the extraction of color information. The CEDD then divide the image into a number of blocks and the extracted histogram is a compromise of 6 bins, as dictated by the texture unit, every bin is further organized in 24 singular regions and the length of the final histogram contains  $6 \times 24 = 144$  bins. In the event that we characterize the receptacle that results from the texture unit as T and as C the container that outcome from the Color Unit, the output histogram position:  $T \times 24 + C$ . In the Texture Unit, the Image Block is isolated into 4 locales, the sub-blocks. The estimation of every sub-block is the mean estimation of the radiance of the pixels that take an interest in it. The radiance

qualities are derived by converting the image into YIQ color space, Y represents luminance information; and IQ represents chrominance information, and then texture information is extracted using MPEG-7 edge histogram.



**Figure 4. Description of CEDD Feature Detection**

The color unit transform every image block into HSI color space and the mean value of H, S and V component is computed and organize input of the fuzzy system which forms fuzzy histogram of 10 bins, every bin signify a color: white, black, magenta, cyan, green, blue, yellow, orange, gray and red. To expand the histogram coordinate logic filter is applied; it expands the histogram to 24 bins. After that operation are applied to every HSI components, the H component (Hue Component) is divided into 8 bins every bin represents different color orange, green, yellow, cyan, magenta, blue red to orange and blue to red and orange, the Saturation components separated into 2 fuzzy regions which define the shade of the color, the value component is divided into 3 areas. Every image block interconnects well with the fuzzy system. Defining the number of bins produced by the texture information fuzzy system as  $j$  and the bins produced by the fuzzy linking system as  $k$ , every block is placed in a bin position  $(j \times 24 + k)$  Figure 4 shows the working of CEDD.

**3.2.7. Color Histogram**

Histogram of image characterizes the dissemination of intensity values in the image. It visually represents the number of intensities in an image color. It represents the number of pixels that have colors in each of a fixed list of color ranges that span the image's color space; it can be built for any kind of color space the set of all possible colors. Studies [48, 49] show the importance of color histogram in content-based image retrieval system. In our experiments, we computed RGB histograms and then all three histograms were concatenated to form a single histogram. The dataset consists of almost 70% of RGB images, for gray-scale images the extracted histogram is then concatenated two more times by itself, so the size of final histogram equal the size of histogram computed for RGB images.

3.2.8. Edge Histogram Descriptor

Edge histogram descriptor (EHD) signifies local edge dispersion in the image [39]. When the primary texture is not consistent, then edge dispersion is a valuable mark for image comparison. EHD defines the edge by first isolating image in 4x4 grids and then describes edge in every sub-image. The edges extracted from sub-image are grouped into five categories; vertical, horizontal, 45-degree, 135-degree, and non-directional existence of every category develops a histogram bin, which makes a histogram of 80 bins. Figure 5 show the distribution of image into image blocks and sub image. A remarkable variant of EHD is to calculate an extended histogram by utilizing 80-bins previously extracted histogram. The extension of 80-bit histogram is achieved by combining the image blocks. The prolonged bins are denoted as a global histogram (formed by grouping entirely 16 image blocks) and semi-global histogram (formed by pooling by image blocks four rows and four columns and five groups). This outcome in five bins for the global histogram and for the semi-global histograms from the 80 local histogram bins. The overall number of bins is hence 150.

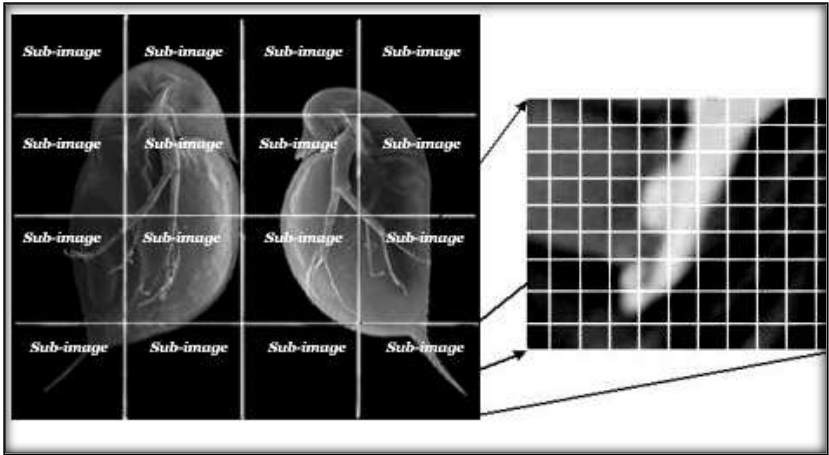


Figure 5. Distribution of Image into Sub-Image and Blocks

3.3. Feature Vector Construction

In machine learning and computer vision, the feature vector is used to characterize some object, feature vector consists of n-dimensions each dimension of the vector contains some numerical information regarding an object, which is used to encourage statistical investigation. Hybrid feature vector was constructed by applying low-level fusion scheme [50]. Figure 6 shows the process of feature vector construction.

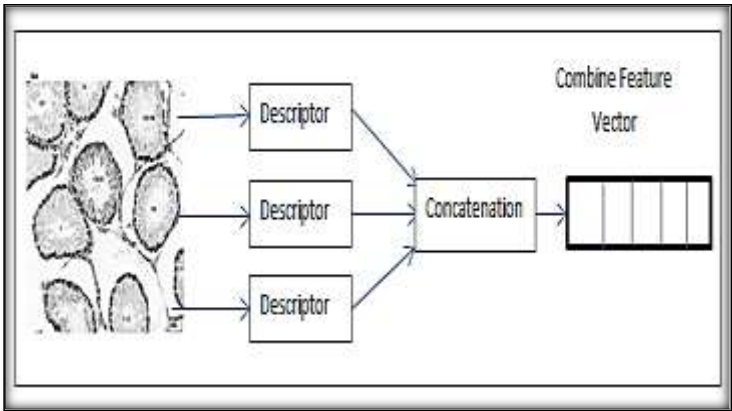


Figure 6. Process of Combined Feature Vector Construction

3.4. Feature Vector Reduction

In classification process, the dimensionality of data is generally too high. Feature reduction is a process to map the high dimensional data to a smaller number of dimensions. Although we have to ensure that the step preserves the accuracy and must not lead us to loss of information. Dimensionality reduction is a preprocessing step carried out prior to classification. The major objectives of this step are, to enhance generalization by removing the redundant dimensions and reducing over-fitting, which support to design classifier in easy to calculate way and improves the performance of classification model. For dimensionality reduction we utilized WEKA [51]; the tool was created to support machine learning and several other tasks such as preprocessing, classification, clustering, and visualization. For evaluator we selected CfsSubsetEval and for Search, we selected Genetic Search [52], Table 1 describes selected parameters for genetic search and their values.

Table 1. Parameter Selection for Genetic Search

Parameters	Values
Population Size	100
Number of Generations	100
Crossover Probability	0.9
Mutation Probability	0.03
Number of Total Features	7826
Number of Reduced Features	2796
Percentage of Achieved Reduction	64%

4. Implementation and results

The process of identifying to which set of sub-population a new observation belongs to is called classification. In this section, we are going to discuss the classification setup, which we used to predict the modality of given medical image.

4.1. Classification Setup

We have utilized support vector machine (SVM), which is widely adopted supervised learning model [17]. The studies show that SVM with an early fusion of several features produces the best performance [18]. For multiclass SVM we have considered chi-square kernel in the form of Equation 4, here  $X, Y$  are  $N$  dimensional inputs and  $K$  is the kernel function that maps the data from  $N$  dimensional space to  $M$  dimension space, usually  $M$  is much larger than  $N$ . Studies show that if features are in the form of a histogram, the Chi-square kernel yields better performance as compared to the others kernels [18, 41 53].

$$K(X,Y)=\sum_i x_i \frac{y_i}{x_i+y_i}$$

(4)

In this research work, we have utilized LIB-SVM tool box for classification [54]. To solve multi-class classification problem, we adopted one-vs-all approach [55], in this approach, we train a binary classifier for each class. Observations related to that class are marked as positive instances and outstanding cases are labeled as negative instances. The data set was highly imbalance few classes had very low number of training instances and while working with, even if the class distribution is balanced in the training set, the binary classification learners see unbalanced distributions because typically the set of negatives they see is much larger than the set of positives, one-vs-all approach most of the time perceived as imbalanced to the model, to solve such issue we assigned weight to every positive and negative class [41]. The weights for positive class were set as  $((\#positive + \#negative) / \#positive)$  and the weights for the negative class was set as  $((\#positive + \#negative) / \#negative)$  where  $\#positive$  is number of positive instances and  $\#negative$  is number of negative instances, so the positive instances got higher weight as compared to negative instances in classification process. Parameter tuning and training have been performed on training dataset by using 10-fold cross validation.

#### 4.2. Evaluation Measures

The performance of proposed framework is evaluated in the terms of Accuracy, Precision, Recall, F-score and ROC analysis. These are exceptionally basic and generally utilized evaluation measures for all CBIR systems. In following equations  $TP, FP, TN$  and  $FN$  represents true positive, false positive, true negative, false negative respectively.

**Accuracy:** Represents the fraction of test set whose class or modality was predicted correctly.

$$Accuracy = \frac{TP + TN}{T + N} \quad (5)$$

**Precision:** Percentage of observations that the classifier predicted as positive are actually positive.

$$Precision = \frac{TP}{(TP + FP)} \quad (6)$$

**Recall:** Completeness – what % of positive tuples did the classifier label as positive.

$$Recall = \frac{TP}{(TP + FN)} \quad (7)$$

**F-score:** Harmonic mean of precision and recall

$$F = \frac{2(Precision.Recall)}{(Precision + Recall)} \quad (8)$$



**ROC Analysis:** ROC (Receiver Operating Characteristics) curve graphically illustrates the trade-off between true positive instances and false-positive instances for a certain classification model. For a given binary classification problem, it permits us to envision the balance among the rate at which the classification model precisely perceives positive instances against the rate at which it erroneously recognizes negative instances as a positive instance. The vertical axis of ROC plot signifies the true-positive rate and horizontal axis denotes the false-positive rate. The measure of correctness for the model is the area under the ROC curve.

4.3.Dataset

We used imageCLEF 2012 medical image classification and retrieval dataset [11]. ImageCLEF is an assessment campaign and organized as part of the CLEF Initiative labs. The standard target of imageCLEF is to support the domain of visual media examination, recovery and characterization by working on imperative frameworks for appraisal of visual information recovery systems functioning in monolingual and cross-language settings. Both academic and industry research groups from worldwide participate in imageCLEF competitions. Since 2004, ImageCLEF has run the medical task. In our work, we concentrated on the dataset utilized in 2012’s competition. We settled on this decision principally in view of the availability of data and the diversity of the data set. The dataset consists of 31 different classes. Figure 7 shows the hierarchy for the imageCLEF 2012modality classification task, the dataset was divided into two parts, train and test set. Train set consists of 1001 images and train set consists of 1000 images from 31 different classes.

4.4.Results for Individual Feature Extraction Methods

This section, we will discuss the selected parameters for feature extraction algorithms and the results of individual feature extraction techniques adopted in our framework.

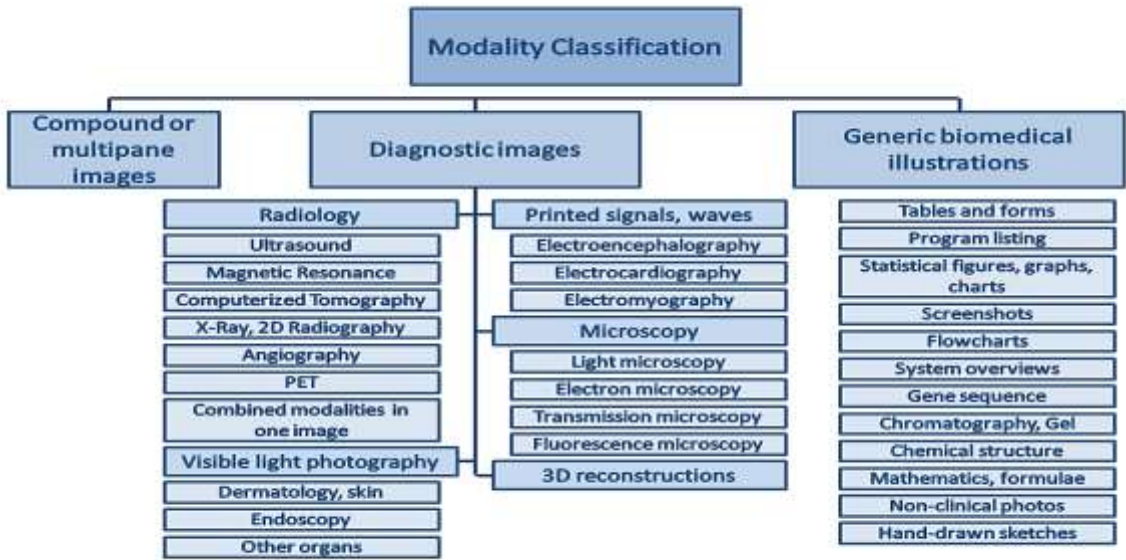


Figure 7. Hierarchy for Modality Classification Task ImageCLEF 2012 [11]

4.5. Classification Based on Scale Invariant Feature Transform

We made extensive experiments with different codebook size such as 200, 400, 600, 800, and 1000 and by combined histograms obtained by different codebook sizes. The model was trained and tested by using chi-square kernel. Table 2 shows the results obtained by using different codebook sizes. The results show that dense-SIFT (DSIFT) descriptors provide better performance as compared to SIFT descriptors. Furthermore, the best performing histogram was obtained by combining (Concatenating) the histogram of vocabulary size 200, 400, 800 and 1000 respectively, the combined histogram obtained the accuracy of 63.1 and 66.7 for SIFT and DSIFT respectively, thus the total length of combined histogram was 2400 bins.

Table 2. Classification Performance for SIFT and DSIFT

Vocabulary Size	Number of Features	Accuracy of DSIFT	Accuracy of SIFT
200	200	52.5%	50.3%
400	400	59.3%	56.1%
600	600	61.5%	58.0%
800	800	63.0%	59.2%
1000	1000	62.5%	59.3%
Combined Vocabulary (200+400+800+1000)	2400	66.7%	63.1%

4.6. Classification Based on Color Edge Directivity Descriptor (CEDD)

CEDD yields a histogram of 144 bins for each image. Classification accuracy of 37% has been achieved by using CEDD features alone.

Table 3. Classification Performance of CEDD Descriptor

Number of Features	Accuracy Rate
144	37.0%

4.7. Classification Based on Local Binary Pattern Descriptor (LBP)

We extracted LBP feature from the image on the basis of different size radius and a different number of neighbors. Table 4 shows the results of experiments. Best performing histogram was found by combining the all the histogram that are obtained by setting radius value to 1,2,2,3 and number of neighbors to 8, 16, 24, 24 respectively. The combined histogram achieved 54% accuracy. But the technique threshold the central pixel, which tends to be sensitive to noise, especially in near-uniform regions and to smooth weak illumination gradients. Many image regions are relatively uniform and it is valid to investigate whether the robustness of the features can be improved in these regions. That is why we did not consider LBP alone for classification.

Table 4. Classification Performance of LBP

Radius	Neighbors	Number of Features	Accuracy Rate
1	8	59	31.5%
2	16	243	43.4%
2	24	555	48.0%
3	24	555	48.6%
Combined	Combined	1412	54.0%

4.8. Classification Based on Local Ternary Pattern

Experiments for LTP carried using the same number of neighbors and same size of radius as in the case of the local binary pattern (LBP). Table 4.4 describes the classification results for LTP.

Table 5. Classification Performance of LTP

Radius	Neighbors	Number of Features	Accuracy Rate
1	8	118	38.0%
2	16	486	47.3%
2	24	1110	51.3%
3	24	1110	53.4%
Combined	Combined	2824	55.0%

Best performing histogram was found by combining the all the histogram that are obtained by setting radius value to 1,2,2,3 and number of neighbors to 8, 16, 24, 24 respectively. The combined histogram attained 54% accuracy so for next of experiments we selected the combined histogram.

4.9. Classification Based on Color Edge Detection using Discrete Wavelet Transforms

Color edge features yields a histogram of 128 bins and shows the accuracy of 29%. Table 4.5 displays the results of color edge features.

Table 6. Classification Performance of Color Edge Features

Number of Features	Accuracy Rate
128	29.0%

4.10. Classification Based on Color Histogram Features

Combined color histogram for all the three-color channels obtained 29.2%. Table 4.6 describes the obtained results for the color histogram.

Table 7. Classification Performance of Color Histogram

Number of Features	Accuracy Rate
768	29.2

4.11. Classification Based on Edge Histogram Descriptor

EHD characterizes local edge spreading in the image; the features are computed for the whole image by setting the threshold value to 0.10. Table 8 shows the obtained results for edge histogram descriptor.

Table 8. Classification Performance of Edge Histogram Descriptor

Number of Features	Accuracy Rate
150	34.5

4.12. Results for Hybrid Features

A number of studies have shown that use of diverse features captures different information of the image and thus, their combination offers a complete representation of the visual contents of an image and clearly provides better performance as compared to single feature approach [11, 41]. Classification results are presented in the form of accuracy, precision, recall and F-score. Table 4.8 displays the detail performance in the term of accuracy for every individual class for the imageCLEF dataset. The results show that some classes have a very low accuracy. For the compound multi-pane images (COMP) class, the obtained accuracy was only 33.3%, this is because images from this class combine multiple modalities into a single image, which cause the classifier to commit mistakes. For the class of PET (DRPE) image, classifier obtained 38.8% accuracy this is due to the fact that, images from this class were blurred and do not contain adequate color and edge information. For other class such as System overview (GSYS), Electromyography (DSEM) and Mathematical formula (GMAT) classifier obtained 31.7%, 42.9%, and 50% accuracy respectively, the reasons behind the low accuracy of these classes are; images from these classes are consist of simple black and white drawings, numerical formulas, and simple flow charts, which do not contains sufficient visual information in the terms of shape, texture, and color, so lack of lack visual information for these classes affected the performance of classifier. Table 9 shows detail performance of the classifier in the terms of Precision, Recall, and F-Score.

Table 9. Detailed Performances per Class in terms of Accuracy

Class Description	Class Code	Train Images	Test Images	Accuracy
Compound Multi-pane Images	COMP	49	57	33.3
3D Reconstructions	D3DR	25	30	56.6
Electron Microscopy	DMEL	22	29	65.5
Fluorescence Microscopy	DMFL	21	13	84.6

Light Microscopy	DMLI	46	46	89.1
Transmission Microscopy	DMTR	29	18	55.5
Angiography	DRAN	38	17	82.3
Combine Modality	DRCO	12	13	84.5
Computerized Tomography	DRCT	49	64	82.5
Magnetic Resonance	DRMR	43	55	69.1
PET	DRPE	09	18	38.8
Ultrasound	DRUS	48	13	100.0
X-ray	DRXR	48	23	86.1
Electrocardiography	DSEC	05	24	62.5
Electroencephalography	DSEE	06	15	80.0
Electromyography	DSEM	05	14	42.9
Dermatology skin	DVDM	47	33	81.7
Endoscopy	DVEN	32	32	78.1
Other organs	DVOR	48	21	76.2
Chemical structure	GCHE	21	50	94.0
Statistical Figures	GFIG	48	61	77.0
Flow charts	GFLO	48	50	94.0
Chromatography, GEL	GGEL	49	20	85.0
Gene sequence	GGEN	47	42	73.7
Hand written sketch	GHDR	17	29	58.6
Mathematical formula	GMAT	06	14	50.0
Non-Clinical photos	GNCP	47	49	79.3
Program listing	GPLI	10	18	61.1
Screen shots	GSCR	40	54	96.3
System overview	GSYS	48	47	31.7
Tables and Forms	GTAB	38	31	58.0

Table 10. Detailed Performances per Class in terms of Precision, Recall and F-Score

Class Description	Class Code	Precision	Recall	F-Score
-------------------	------------	-----------	--------	---------



<b>Compound Multi-pane Images</b>	COMP	0.33	0.57	0.42
<b>3D Reconstructions</b>	D3DR	0.56	0.70	0.62
<b>Electron Microscopy</b>	DMEL	0.65	0.79	0.71
<b>Fluorescence Microscopy</b>	DMFL	0.84	0.57	0.68
<b>Light Microscopy</b>	DMLI	0.89	0.83	0.86
<b>Transmission Microscopy</b>	DMTR	0.55	0.43	0.48
<b>Angiography</b>	DRAN	0.82	0.63	0.71
<b>Combine Modality</b>	DRCO	0.84	0.84	0.84
<b>Computerized Tomography</b>	DRCT	0.82	0.82	0.82
<b>Magnetic Resonance</b>	DRMR	0.69	0.71	0.70
<b>PET</b>	DRPE	0.38	1.00	0.56
<b>Ultrasound</b>	DRUS	1.00	0.65	0.78
<b>X-ray</b>	DRXR	0.86	0.57	0.68
<b>Electrocardiography</b>	DSEC	0.62	0.93	0.75
<b>Electroencephalography</b>	DSEE	0.80	0.80	0.80
<b>Electromyography</b>	DSEM	0.42	0.75	0.54
<b>Dermatology skin</b>	DVDM	0.81	0.75	0.78
<b>Endoscopy</b>	DVEN	0.78	0.80	0.79
<b>Other organs</b>	DVOR	0.76	0.61	0.68
<b>Chemical structure</b>	GCHE	0.94	0.90	0.92
<b>Statistical Figures</b>	GFIG	0.77	0.72	0.74
<b>Flow charts</b>	GFLO	0.94	0.59	0.72
<b>Chromatography, GEL</b>	GGEL	0.85	0.58	0.69
<b>Gene sequence</b>	GGEN	0.73	0.83	0.78
<b>Hand written sketch</b>	GHDR	0.58	0.80	0.68
<b>Mathematical formula</b>	GMAT	0.50	0.58	0.53
<b>Non-Clinical photos</b>	GNCP	0.79	0.75	0.77
<b>Program listing</b>	GPLI	0.61	0.57	0.59
<b>Screen shots</b>	GSCR	0.96	0.78	0.86
<b>System overview</b>	GSYS	0.31	0.75	0.44
<b>Tables and Forms</b>	GTAB	0.58	0.60	0.59

4.13. ROC Analysis

ROC analysis is also carried out on results of hybrid features. The area under the ROC curve is calculated for each class and then average area under the curve is also computed. Figure 8 shows the ROC plot for all 31 classes.

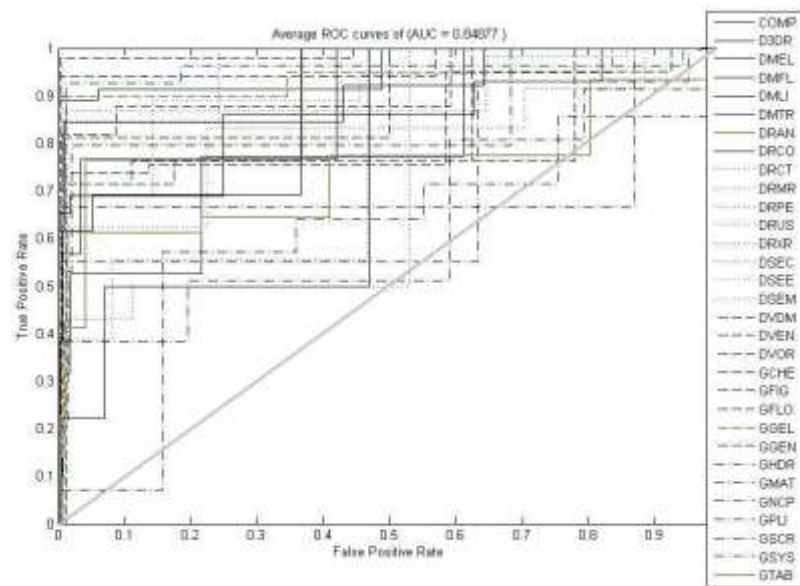


Figure 8. ROC Curve for ImageCLEF Dataset using SVM classification

Table 11 is showing area under the ROC curve for each class, according to ROC plot for 3 classes less than 0.7 area under curve is perceived as compared to other 28 classes, the classes GMAT (Mathematical Formula), GSYS (System overview) and GHDR (Hand Writing Sketch) shows the 0.61,0.63 and 0.69 area under the curve respectively. This is due to fact that images from these classes contain mathematical formulas, hand written made drawing and system flow charts, these images provide less visual information in the terms of colors, shapes, and texture because of their plainness and simple black and white color nature. For other 28 classes, more than 0.7 areas under the curve are observed. The average area under the ROC was 0.85. Table 4.11 shows the average performance of the proposed model in terms accuracy, precision, recall, F-score and area under the ROC curve.

Table 11. Area under the ROC Curve for each Class

Class Description	Class Code	Area Under Curve
Compound Multi-pane Images	COMP	0.80
3D Reconstructions	D3DR	0.89
Electron Microscopy	DMEL	0.86
Fluorescence Microscopy	DMFL	0.91
Light Microscopy	DMLI	0.96
Transmission Microscopy	DMTR	0.75

Angiography	DRAN	0.90
Combine Modality	DRCO	0.88
Computerized Tomography	DRCT	0.96
Magnetic Resonance	DRMR	0.90
PET	DRPE	0.78
Ultrasound	DRUS	0.99
X-ray	DRXR	0.90
Electrocardiography	DSEC	0.82
Electroencephalography	DSEE	0.91
Electromyography	DSEM	0.73
Dermatology skin	DVDM	0.92
Endoscopy	DVEN	0.90
Other organs	DVOR	0.84
Chemical structure	GCHE	0.96
Statistical Figures	GFIG	0.83
Flow charts	GFLO	0.99
Chromatography, GEL	GGEL	0.93
Gene sequence	GGEN	0.80
Hand written sketch	GHDR	0.69
Mathematical formula	GMAT	0.61
Non-Clinical photos	GNCP	0.85
Program listing	GPLI	0.71
Screen shots	GSCR	0.95
System overview	GSYS	0.63
Tables and Forms	GTAB	0.74

Table 12. Average Performances in terms of Accuracy, Precision, Recall and F-Score and Area under ROC Curve

Overall Accuracy	Average Precision	Average Recall	Average F-Score	Average ROC
72.20	0.71	0.72	69.90	0.85

4.14. Performance Comparison

In this section, the performance of proposed framework is compared with other frameworks mentioned in the literature for imageCLEF 2012 modality classification task using visual features. Several groups participated in imageCLEF 2012 competition for medical retrieval task, there were 8 groups, who experimented with visual features. The best average accuracy reported by using visual features was 69.6% by IBM Multimedia Analytics group [18]. Other participants which include UESTC Group, ITI Group, ETFBL Group, medGIFT Group, GEIAL Group, IPL Group and UNED-UV Group reported 57.8%, 51.6%, 48.5%, 42.2%, 39.5%, 26.6% and 15.7% average accuracy respectively [11]. Afterward, Dimitrovski et al [41] experimented with the dataset in 2015 and they reported 71.2% accuracy by using visual features. Obtained accuracy by proposed methodology is 72.2%, which is 2.6 % and 1 % higher than IBM Multimedia Analytics group [18] and Dimitrovski et al [41] respectively. Figure 4-3 shows performance comparison between our obtained results and other reported results in the literature, the x-axis shows group name, and y-axis show their obtained accuracy. According to best of our knowledge, the principal conclusion from experimentations is that the results acquired with our methodology are better from previously reported results. It is due to fact that the dataset is so much diverse that single feature extraction algorithm with particular settings is not sufficient to extract information that provides better classification and retrieval results. We utilized a different type of features extraction algorithms that capture color, texture and other visual information; furthermore, we tried to employ different setting for every particular algorithm to extract more and more information that helped us in the classification task.

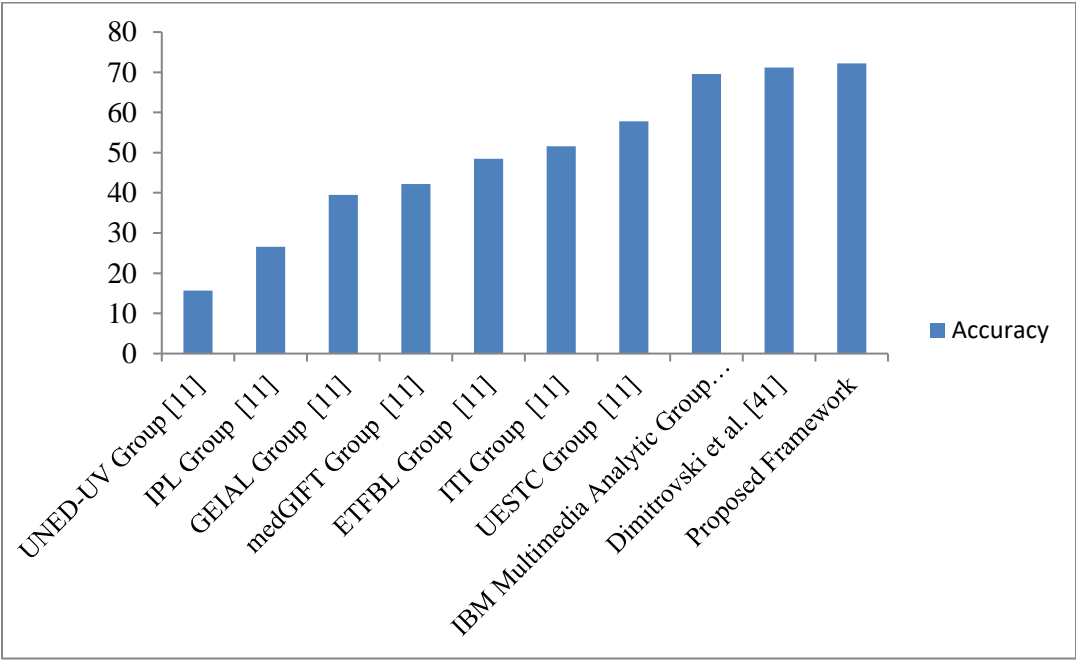


Figure 9. Performance Comparison for Modality Classification Task ImageCLEF2012 using Visual Features.

## 4. Conclusion

In this research, we present the methodology in detail that we utilized for the medical image modality classification for medical images. All the more particularly, we utilized the dataset of the ImageCLEF 2012. The dataset for the modality classification task was consist of 1001 train and 1000 test images, which was further divided into 31 different classes or modalities. We experimented with several types of visual features, the features that we employed in our research consist of, scale invariant feature transform (SIFT) and its variant dense SIFT (DSIFT), local binary pattern (LBP), local ternary pattern (LTP), edge histogram descriptor (EHD), color edges features using wavelet transform and color histogram. First, we experimented with each individual feature extraction technique. We also experimented with combined visual features obtained by different settings for SIFT, DSIFT, LBP, and LTP. Finally, the experiments are carried out for the hybrid feature. The experimentation (for individual feature extraction method) reveal that the scale-invariant feature transform (SIFT) and dense scale invariant transform (DSIFT) are best performing individual features in our framework and exhibited the accuracy of 63.1 and 66.7 respectively. Local binary pattern (LBP) and local ternary pattern (LTP) demonstrated second best performance as individual feature extraction techniques; both techniques exhibited the accuracy of 54% and 55% respectively. Other features like color edge directivity descriptor (CEDD), edge histogram descript (EHD), color histogram, color edge using wavelet transform exhibited the accuracy of 37%, 34.5%, 29.2%, 29 % respectively. After selection of best performing set of features, early fusion (Concatenation) is applied on features, to form a combined hybrid feature vector and experiments carried out on hybrid feature vector. The hybrid features exhibited the accuracy of 72.2%. The comparison of the performance for hybrid features is compared with other performances mentioned in the literature, on this dataset the best-reported accuracy in imageCLEF 2012 competition was 69.6%, which was further enhanced by 71.2% in 2015. The results of our proposed methodology approach 2.6 % and 1 % higher than best reported results in 2012 imageCLEF competition and in 2015 where the accuracy was further enhanced to 71.2% respectively, this is due to fact that, the use diverse features capture different information of image and thus, their combination offers a complete representation of the visual content of an image.



## References:

- [1] Khan, S.U.R., Asif, S., Bilal, O. et al. Lead-cnn: lightweight enhanced dimension reduction convolutional neural network for brain tumor classification. *Int. J. Mach. Learn. & Cyber.* (2025). <https://doi.org/10.1007/s13042-025-02637-6>.
- [2] Khan, S.U.R., et al. Robust&Precise Knowledge Distillation-based Novel Context-Aware Predictor for Disease Detection in Brain and Gastrointestinal. 2025.
- [3] Meeran, M. T., Raza, A., & Din, M. (2018). Advancement in GSM Network to Access Cloud Services. *Pakistan Journal of Engineering, Technology & Science* [ISSN: 2224-2333], 7(1).
- [4] Hekmat, A., et al., Brain tumor diagnosis redefined: Leveraging image fusion for MRI enhancement classification. *Biomedical Signal Processing and Control*, 2025. 109: p. 108040.
- [5] Khan, Z., Hossain, M. Z., Mayumu, N., Yasmin, F., & Aziz, Y. (2024, November). Boosting the Prediction of Brain Tumor Using Two Stage BiGait Architecture. In *2024 International Conference on Digital Image Computing: Techniques and Applications (DICTA)* (pp. 411-418). IEEE.
- [6] Khan, S. U. R., Raza, A., Shahzad, I., & Ali, G. (2024). Enhancing concrete and pavement crack prediction through hierarchical feature integration with VGG16 and triple classifier ensemble. In *2024 Horizons of Information Technology and Engineering (HITE)*(pp. 1-6). IEEE <https://doi.org/10.1109/HITE63532>.
- [7] Khan, S.U.R., Zhao, M. & Li, Y. Detection of MRI brain tumor using residual skip block based modified MobileNet model. *Cluster Comput* 28, 248 (2025). <https://doi.org/10.1007/s10586-024-04940-3>
- [8] Khan, U. S., & Khan, S. U. R. (2024). Boost diagnostic performance in retinal disease classification utilizing deep ensemble classifiers based on OCT. *Multimedia Tools and Applications*, 1-21.
- [9] Asif, S., Awais, M., & Khan, S. U. R. (2023). IR-CNN: Inception residual network for detecting kidney abnormalities from CT images. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 12(1), 35.
- [10] Aslam, M., Salahuddin, Ali, G., & Batool, S. (2024). Assessing the Effects Of Big Data Analytics And Ai On Talent Acquisition And Retention. *Kashf Journal of Multidisciplinary Research*, 1(11), 73-84.
- [11] Khan, M. A., Khan, S. U. R., Haider, S. Z. Q., Khan, S. A., & Bilal, O. (2024). Evolving knowledge representation learning with the dynamic asymmetric embedding model. *Evolving Systems*, 1-16.
- [12] Raza, A., & Meeran, M. T. (2019). Routine of encryption in cognitive radio network. *Mehran University Research Journal of Engineering & Technology*, 38(3), 609-618.
- [13] Al-Khasawneh, M. A., Raza, A., Khan, S. U. R., & Khan, Z. (2024). Stock Market Trend Prediction Using Deep Learning Approach. *Computational Economics*, 1-32.
- [14] Salahuddin, Hussain, M., & hamza Shafique, P. (2024). PERFORMANCE ANALYSIS OF MATCHED FILTER-BASED SECONDARY USER DETECTION IN COGNITIVE RADIO NETWORKS. *Kashf Journal of Multidisciplinary Research*, 1(10), 15-26.
- [15] Khan, U. S., Ishfaq, M., Khan, S. U. R., Xu, F., Chen, L., & Lei, Y. (2024). Comparative analysis of twelve transfer learning models for the prediction and crack detection in concrete dams, based on borehole images. *Frontiers of Structural and Civil Engineering*, 1-17.
- [16] Khan, S. U. R., & Asif, S. (2024). Oral cancer detection using feature-level fusion and novel self-attention mechanisms. *Biomedical Signal Processing and Control*, 95, 106437.

- [17] Farooq, M. U., Khan, S. U. R., & Beg, M. O. (2019, November). Melta: A method level energy estimation technique for android development. In 2019 International Conference on Innovative Computing (ICIC) (pp. 1-10). IEEE.
- [18] Waqas, M., Tahir, M. A., Al-Maadeed, S., Bouridane, A., & Wu, J. (2024). Simultaneous instance pooling and bag representation selection approach for multiple-instance learning (MIL) using vision transformer. *Neural Computing and Applications*, 36(12), 6659-6680.
- [19] Salahuddin, Syed Shahid Abbas, Prince Hamza Shafique, Abdul Manan Razzaq, & Mohsin Ikhlaiq. (2024). Enhancing Reliability and Sustainability of Green Communication in Next-Generation Wireless Systems through Energy Harvesting. *Journal of Computing & Biomedical Informatics*.
- [20] Raza, A.; Meeran, M.T.; Bilhaj, U. Enhancing Breast Cancer Detection through Thermal Imaging and Customized 2D CNN Classifiers. *VFAST Trans. Softw. Eng.* 2023, 11, 80–92.
- [21] Salahuddin, Abdul Manan Razzaq, Syed Shahid Abbas, Mohsin Ikhlaiq, Prince Hamza Shafique, & Inziam Shahzad. (2024). Development of OWL Structure for Recommending Database Management Systems (DBMS). *Journal of Computing & Biomedical Informatics*, 7(02).
- [22] Dai, Q., Ishfaq, M., Khan, S. U. R., Luo, Y. L., Lei, Y., Zhang, B., & Zhou, W. (2024). Image classification for sub-surface crack identification in concrete dam based on borehole CCTV images using deep dense hybrid model. *Stochastic Environmental Research and Risk Assessment*, 1-18.
- [23] Waqas, M., Ahmed, S. U., Tahir, M. A., Wu, J., & Qureshi, R. (2024). Exploring Multiple Instance Learning (MIL): A brief survey. *Expert Systems with Applications*, 123893.
- [24] Khan, S.U.R.; Asif, S.; Bilal, O.; Ali, S. Deep hybrid model for Mpox disease diagnosis from skin lesion images. *Int. J. Imaging Syst. Technol.* 2024, 34, e23044.
- [25] Khan, S.U.R.; Zhao, M.; Asif, S.; Chen, X.; Zhu, Y. GLNET: Global–local CNN’s-based informed model for detection of breast cancer categories from histopathological slides. *J. Supercomput.* 2023, 80, 7316–7348.
- [26] Hekmat, Arash, Zuping Zhang, Saif Ur Rehman Khan, Ifza Shad, and Omair Bilal. "An attention-fused architecture for brain tumor diagnosis." *Biomedical Signal Processing and Control* 101 (2025): 107221.
- [27] Mahmood, F., Abbas, K., Raza, A., Khan, M.A., & Khan, P.W. (2019 ). Three Dimensional Agricultural Land Modeling using Unmanned Aerial System (UAS). *International Journal of Advanced Computer Science and Applications (IJACSA)* [p-ISSN : 2158-107X, e-ISSN : 2156-5570], 10(1).
- [28] M. Waqas, Z. Khan, S. U. Ahmed and A. Raza, "MIL-Mixer: A Robust Bag Encoding Strategy for Multiple Instance Learning (MIL) using MLP-Mixer," 2023 18th International Conference on Emerging Technologies (ICET), Peshawar, Pakistan, 2023, pp. 22-26.
- [29] Khan, S.U.R.; Zhao, M.; Asif, S.; Chen, X. Hybrid-NET: A fusion of DenseNet169 and advanced machine learning classifiers for enhanced brain tumor diagnosis. *Int. J. Imaging Syst. Technol.* 2024, 34, e22975.
- [30] Khan, S.U.R.; Raza, A.; Waqas, M.; Zia, M.A.R. Efficient and Accurate Image Classification Via Spatial Pyramid Matching and SURF Sparse Coding. *Lahore Garrison Univ. Res. J. Comput. Sci. Inf. Technol.* 2023, 7, 10–23.
- [31] HUSSAIN, S., Raza, A., MEERAN, M. T., IJAZ, H. M., & JAMALI, S. (2020). Domain Ontology Based Similarity and Analysis in Higher Education. *IEEEP New Horizons Journal*, 102(1), 11-16.

- [32] Waqas, M., Tahir, M. A., Al-Maadeed, S., Bouridane, A., & Wu, J. (2024). Simultaneous instance pooling and bag representation selection approach for multiple-instance learning (MIL) using vision transformer. *Neural Computing and Applications*, 36(12), 6659-6680.
- [33] Farooq, M.U.; Beg, M.O. Bigdata analysis of stack overflow for energy consumption of android framework. In *Proceedings of the 2019 International Conference on Innovative Computing (ICIC)*, Lahore, Pakistan, 1–2 November 2019; pp. 1–9.
- [34] M. Wajid, M. K. Abid, A. Asif Raza, M. Haroon, and A. Q. Mudasar, "Flood Prediction System Using IOT & Artificial Neural Network", *VFAST trans. softw. eng.*, vol. 12, no. 1, pp. 210–224, Mar. 2024.
- [35] Shahzad, I., Khan, S. U. R., Waseem, A., Abideen, Z. U., & Liu, J. (2024). Enhancing ASD classification through hybrid attention-based learning of facial features. *Signal, Image and Video Processing*, 1-14.
- [36] Khan, S. R., Raza, A., Shahzad, I., & Ijaz, H. M. (2024). Deep transfer CNNs models performance evaluation using unbalanced histopathological breast cancer dataset. *Lahore Garrison University Research Journal of Computer Science and Information Technology*, 8(1).
- [37] Bilal, Omair, Asif Raza, and Ghazanfar Ali. "A Contemporary Secure Microservices Discovery Architecture with Service Tags for Smart City Infrastructures." *VFAST Transactions on Software Engineering* 12, no. 1 (2024): 79-92.
- [38] Bilal, O., Asif, S., Zhao, M., Khan, S. U. R., & Li, Y. (2025). An amalgamation of deep neural networks optimized with Salp swarm algorithm for cervical cancer detection. *Computers and Electrical Engineering*, 123, 110106.
- [39] Khan, S. U. R., Asif, S., Zhao, M., Zou, W., Li, Y., & Li, X. (2025). Optimized deep learning model for comprehensive medical image analysis across multiple modalities. *Neurocomputing*, 619, 129182.
- [40] Raza, A., Soomro, M. H., Shahzad, I., & Batool, S. (2024). Abstractive Text Summarization for Urdu Language. *Journal of Computing & Biomedical Informatics*, 7(02).
- [41] Khan, S. U. R., Asif, S., Zhao, M., Zou, W., & Li, Y. (2025). Optimize brain tumor multiclass classification with manta ray foraging and improved residual block techniques. *Multimedia Systems*, 31(1), 1-27.
- [42] Khan, S. U. R., Asim, M. N., Vollmer, S., & Dengel, A. (2025). AI-Driven Diabetic Retinopathy Diagnosis Enhancement through Image Processing and Salp Swarm Algorithm-Optimized Ensemble Network. *arXiv preprint arXiv:2503.14209*.
- [43] Asif Raza, Inzamam Shahzad, Ghazanfar Ali, and Muhammad Hanif Soomro. "Use Transfer Learning VGG16, Inception, and Reset50 to Classify IoT Challenge in Security Domain via Dataset Bench Mark." *Journal of Innovative Computing and Emerging Technologies* 5, no. 1 (2025).
- [44] Shahzad, Inzamam, Asif Raza, and Muhammad Waqas. "Medical Image Retrieval using Hybrid Features and Advanced Computational Intelligence Techniques." *Spectrum of engineering sciences* 3, no. 1 (2025): 22-65.
- [45] Khan, Z., Khan, S. U. R., Bilal, O., Raza, A., & Ali, G. (2025, February). Optimizing Cervical Lesion Detection Using Deep Learning with Particle Swarm Optimization. In *2025 6th International Conference on Advancements in Computational Sciences (ICACS)* (pp. 1-7). IEEE.

- [46] Raza, A., Salahuddin, & Inzamam Shahzad. (2024). Residual Learning Model-Based Classification of COVID-19 Using Chest Radiographs. *Spectrum of Engineering Sciences*, 2(3), 367–396.
- [47] Khan, S.U.R., Raza, A., Shahzad, I., Khan, S. (2025). Subcellular Structures Classification in Fluorescence Microscopic Images. In: Arif, M., Jaffar, A., Geman, O. (eds) *Computing and Emerging Technologies. ICCET 2023. Communications in Computer and Information Science*, vol 2056. Springer, Cham. [https://doi.org/10.1007/978-3-031-77620-5\\_20](https://doi.org/10.1007/978-3-031-77620-5_20)
- [48] Hekmat, A., Zuping, Z., Bilal, O., & Khan, S. U. R. (2025). Differential evolution-driven optimized ensemble network for brain tumor detection. *International Journal of Machine Learning and Cybernetics*, 1-26.
- [49] Khan, S. U. R. (2025). Multi-level feature fusion network for kidney disease detection. *Computers in Biology and Medicine*, 191, 110214.
- [50] Khan, S. U. R., Asif, S., & Bilal, O. (2025). Ensemble Architecture of Vision Transformer and CNNs for Breast Cancer Tumor Detection From Mammograms. *International Journal of Imaging Systems and Technology*, 35(3), e70090.
- [51] Khan, S. U. R., & Khan, Z. (2025). Detection of Abnormal Cardiac Rhythms Using Feature Fusion Technique with Heart Sound Spectrograms. *Journal of Bionic Engineering*, 1-20.
- [52] Khan, S. U. R. (2025). Multi-level feature fusion network for kidney disease detection. *Computers in Biology and Medicine*, 191, 110214.