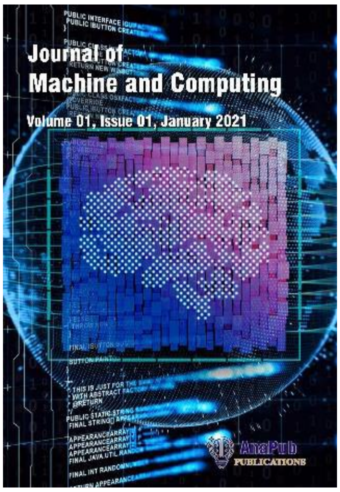
Journal Pre-proof



An Effective Content-Based Image Retrieval Using Multi-Feature Fusion Algorithm with Optimized Retrieval Technique of Soft Computing Approach

Pushpalatha N, Sumendra Yogarayan, Selvi A, Gunapriya D and Siti Fatimah Abdul Razak

DOI: 10.53759/7669/jmc202505153

Reference: JMC202505153

Journal: Journal of Machine and Computing.

Received 28 March 2025

Revised from 03 May 2025

Accepted 20 June 2025

Please cite this article as: Pushpalatha N, Sumendra Yogarayan, Selvi A, Gunapriya D and Siti Fatimah Abdul Razak, "An Effective Content-Based Image Retrieval Using Multi-Feature Fusion Algorithm with Optimized Retrieval Technique of Soft Computing Approach", Journal of Machine and Computing. (2025). Doi: https:// doi.org/10.53759/7669/jmc202505153.

This PDF file contains an article that has undergone certain improvements after acceptance. These enhancements include the addition of a cover page, metadata, and formatting changes aimed at enhancing readability. However, it is important to note that this version is not considered the final authoritative version of the article.

Prior to its official publication, this version will undergo further stages of refinement, such as copyediting, typesetting, and comprehensive review. These processes are implemented to ensure the article's final form is of the highest quality. The purpose of sharing this version is to offer early visibility of the article's content to readers.

Please be aware that throughout the production process, it is possible that errors or discrepancies may be identified, which could impact the content. Additionally, all legal disclaimers applicable to the journal remain in effect.

AnaPub
PUBLICATIONS

An Effective Content-Based Image Retrieval Using Multi-Feature Fusion Algorithm with Optimized Retrieval Technique of Soft Computing Approach

Pushpalatha N[1], Sumendra Yogarayan[2*], Selvi A[3], D. Gunapriya[4], Siti Fatimah Abdul Razak[5]

[1]*Faculty of Information Science and Technology, Multimedia University, Melaka 75450, Malaysia, Department of Electrical and Electronics Engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India.* *pushpalatha.n@sece.ac.in*

[2]*Faculty of Information Science and Technology, Multimedia University, Melaka 75450, Malaysia.* *sumendra@mmu.edu.my*

[3]*Artificial intelligence and Data science, M.Kumarasamy College of Engineering, Karur-639113.* *selvia1@gmail.com* .

[4]*Department of Electrical and Electronics Engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India.* *gunapriya.d@sece.ac.in*

[5]*Faculty of Information Science and Technology, Multimedia University, Melaka 75450, Malaysia,* *fatimah.razak@mmu.edu.my*

***Corresponding author:** Sumendra Yogarayan, Faculty of Information Science and Technology, Multimedia University, Melaka 75450, Malaysia. sumendra@mmu.edu.my

Abstract:

With the increasing digitization of healthcare, hospitals generate and store thousands of medical images daily, creating large-scale datasets that demand efficient retrieval solutions. Content-Based Image Retrieval (CBIR) systems address this by identifying relevant images based on visual features rather than textual metadata. While various CBIR approaches exist, many suffer from low precision, redundant retrievals, and slow query processing times. This paper introduces a novel hybrid CBIR framework that significantly improves retrieval accuracy and efficiency by integrating Principal Component Analysis (PCA) for texture extraction, Wavelet Transform (WT) for shape feature extraction, and Canonical Correlation Analysis (CCA) for advanced feature fusion. Unlike previous methods that rely on single-feature analysis or basic fusion strategies, our approach combines multiple complementary features into a unified representation, enhancing the system's ability to discern subtle patterns in medical images. CCA helps to find features from the medical images that are maximally related, e.g., the part of the breast that usually co-occur when someone is under observation. Additionally, we apply a customized classification strategy using Fuzzy Support Vector Machine optimized with Modified Whale Optimization Algorithm (FSVM-MWOA), which enhances model adaptability and retrieval precision. FSVM is a variant of SVM that incorporates fuzzy logic to handle uncertainty and noisy data, MWOA an enhanced version of the bio-inspired Whale Optimization Algorithm, used here to optimize the parameters of the FSVM. Experimental results show that the proposed system achieves over 90% retrieval accuracy, reduces query response time by up to 40%, and minimizes redundancy, outperforming conventional CBIR techniques. This integrated approach not only addresses the limitations of existing methods but also introduces a scalable and robust solution tailored to the specific challenges of medical image datasets.

Keywords: Wavelet Transform, CBIR Efficiency, Mammography Image, PCA, Feature Fusion, CCA, Fuzzy SVM, Optimization Algorithm, Medical Image Retrieval.

1. Introduction

Diagnosis of diseases is a critical preventive step for minimizing the death rate and giving proper treatment to the patient. Hence, other medical multi-imaging modalities like Ultrasound sonograms (ULS), Bio-psy (histological images), Computerized thermography and magnetic resonance imaging (MRI) assists the physicians in the early diagnosis of breast cancer and the various stage of breast cancer as well as staging of the breast cancer [1].

Among the woman-related diseases, breast cancer is considered the most dangerous and it is also a disease that can endangers someone's life. It is a type of invasive tumor and it originates from the breast ductal and lobular cells. A malignant tumor is also known to be a mass comprised of cancer cells that invade nearby tissues within the breast and other parts. Carcinoma is still one of as common cancers in women. This particular kind of tumor originates in epithelial cells found in the lining of organs and tissues throughout the body. Breast cancers start most often as a kind of carcinoma known as adenocarcinoma. This term describes the neoplasm arising from glandular tissue that has a secretory function [2].

In working with images, deep learning proves to be a very useful technology in practice. A content based image retrieval system is great technique to making breast cancer diagnosis with high accuracy using the medical images conventionally obtained from multi image modalities. Physicians are able to compare the current mammography images and past mammography images along with pathologic conditions which are used for the diagnosis of earlier stages of diseases like breast cancer. It also filters out unrelated medical images by using input query images as a references [3].

The linguistically correct understanding of CBIR is focused on compressing the relevance of information and searching for medical images within huge information. The model of CBIR is applied in many areas e.g.: medical image processing systems, commercial advertisement, military etc. In the last few years, an increased death rate is reached due to increased breast cancer prevalence. The breast cancer is diagnosed with the help of mammograms. For effective diagnosis of breast tumors, health professionals utilize the CBIR methods to evaluate the stages of the disease and plan the treatment of breast cancer correctly [4].

In their study, Zhu, M et al. [5] suggested that breast cancer can be detected through the use of long non-coding RNAs (lncRNAs) that play a vital role in tumorigenesis. According to Bick, U et al. [6], Using MRI images that identify the breast and ovarian cancerous types based on the features like texture, shape, and size of regions, and applying morphological operators.

For the implementation of retrieval of medical images based on CBIR, many research works have been implemented. Considering the existing research works, the procedures are ineffective and time-consuming considering the amount of effort that goes into searching for mammography images from the database or even finding a search image that persists in the retrieval of irrelevant images. To address these challenges, this study incorporates the FSVM-MWOA based image retrieval. The recommended system offers an efficient retrieval mechanism by restricting the number of unnecessary images retrieved while using the smallest possible time to fetch the images from the database. The aim of this proposed work is:

- To get a more accurate retrieval of the image, the removal of noise from the mammography image is implemented by applying a Kalman filter.
- Retrieval of the high similarity of medical images from the large data set both texture and shape features are extracted by applying PCA for extracting the texture feature and Wavelet transform for the shape feature of the image.
- Applying fusion of feature extraction is implemented by a canonical correlation algorithm.
- To classify the similarity of images by using a fuzzy-based SVM algorithm with an optimized algorithm of the modified whale optimization algorithm.
- To compute similarity images between input query images and images in the large data set by using Euclidean distance.

2. Review of Literature

This part focuses on the literature regarding concepts surrounding CBIR in medical mammograms used in breast cancer diagnosis. In the health sector, images are obtained from different modalities including, Digital Mammogram, DSA, Doppler Ultrasound, MRI, CT, Digitized X rays Films, DEM. These other images, on the other hand, pose a problem of collecting, storing images and managing the server of the hospital. A number of Hospitals have been able to effectively apply Picture Archiving and Communication Systems (PACS) with the aim of augmenting efficient storage systems and fast retrieval systems for images. PACS is text-oriented and it is also time-consuming. Hence, CBIR concept is implemented [7].

The volume of medical databases continues to grow, and although it's impossible to label every image in the database, it is neither effective nor practical to do so. Hence, a system to assist based on image retrieval using keywords has been developed whereby the features of the image allow to search for specified images. The CBIR model uses the supplied query image and retrieves related images from the extensive collection [8].

According to Basile and his collaborators, microcalcifications can be diagnosed in digital mammograms through a fully automated computer-aided system. R. Vijayarajeswari and his colleagues describe analysis of the images of the database using SVM classifier and the Hough transform, which contributes to earlier detection of breast cancer using mammograms. H. Cai and colleagues used a deep convolutional neural network for the purpose of handling digital mammogram images in cancer diagnosis. Authors G. V. Ionescu and his colleagues focused their investigation on the convolutional neural networks and used them for the prediction of mammographic density and for retrieval of the mammographic images as well.

According to Amelio [13], similarity encoding can be retrieved from the network of dots employing a new axiomatic methodology. K T. Ahmed and colleagues [14] undertook a combination of feature extraction from image data base and retrieval of similar images from the data base. Shinde et al [15] implemented soft computing approaches for the classification of the dataset of mammographic images and retrieved the mammograms based on segmentation of the pectoral muscle from the mammographic images. This is supported by Ashebir Dohe Dogoma et al [16], who applied content-based image retrieval systems in the task of breast cancer detection and the classification of mammography images in the dataset.

Sonia Jenifer Rayen et.al [17] presented that retrieval of mammography images from the dataset uses optimized classifiers. Initially, it filtered the mammography image by applying the Modified Weiner filter. This filter removes the pectoral of the image. Then extracting the features and classifying the images as benign, malignant, and normal by implementing the optimized classifier. It employs and optimizes the "ABC Algorithm" Neuro Fuzzy System as suggested by M. Suebir.

In this review of literature, we examine the range of research works pertaining to the detection and diagnosis of breast cancer based on assessment of mammography images as per the provided dataset. Using a content-based image retrieval system, a mammographic image dataset is constructed. The problems in the previous works include ineffective retrieval of images, dependence of image retrieval on text, and long duration taken.

3. Retrieval of mammography images

Retrieval of mammography images from the large dataset using feature fusion technique along with retrieval process of FSVM-MWOA. This proposed work FSVM-MWOA contains five modules. Figure 1 shows the framework of FSVM-MWOA. It consists of five modules of data collection, pre-processing, Feature Extraction, Feature Fusion using canonical correlation analysis (CCA), and Retrieval of similarity of images using FuzzySVM with Modified Whale Optimization Algorithm (FSVM-MWOA). The Modified Whale Optimization Algorithm (MWOA) is employed to optimize the parameters of the FSVM, such as kernel functions and fuzzy membership values, thereby enhancing classification accuracy and model robustness.

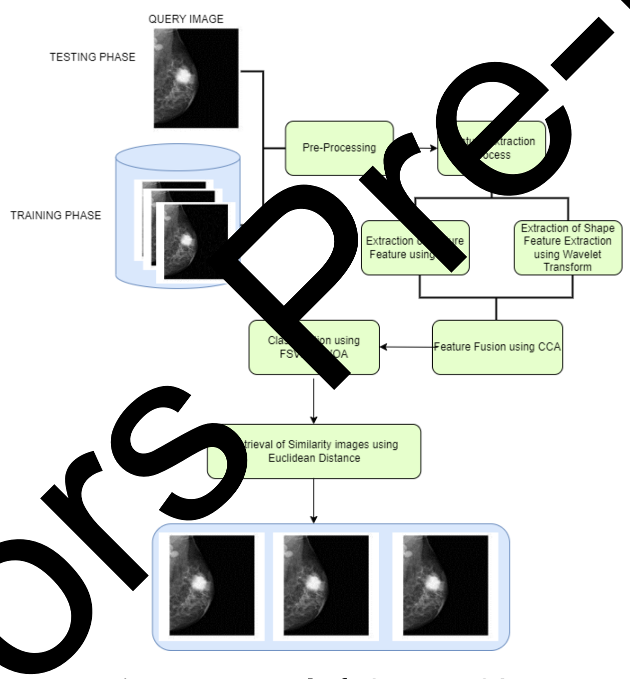


Figure 1 Framework of FSVM-MWOA

3.1 Data Collection

The FSVM-MWOA leverages the breast cancer screening database established by MIAS/Mini-MIAS and DDSM/CBIS-DDSM in order to effectively retrieve mammographic images from a massive dataset. This can be established as follows [19, 20].

3.2 Pre-Processing

The first step in collecting the relevant mammographic images is pre-processing. In this research activity pre-processing work refers to the elimination of noise, the eradication of artifacts, and the normalization process. Certain visuals depicting the primary steps in the pre-processing of images are shown in Figure 2. The model begins with pre-processing steps such as noise reduction using median filtering and contrast enhancement via histogram equalization to improve lesion visibility. This is followed by multi-feature extraction using Principal Component Analysis (PCA) for dimensionality reduction and Wavelet Transform (WT) to capture both texture and edge-related features from the mammogram images.

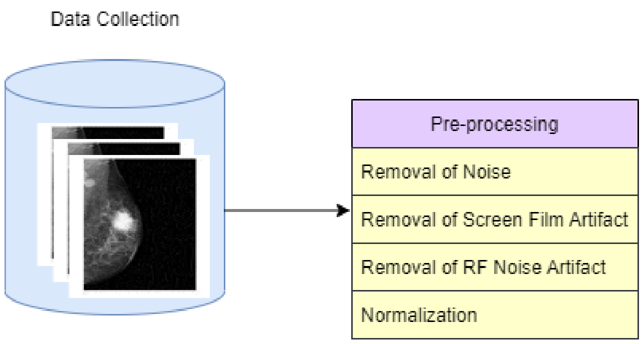


Figure 2 Pre-Processing Phase

3.2.1 Noise Removal

Mathematically, it is based on the principle of the linear system with Gaussian perturbation. This filter utilizes the best estimate of the neighbouring pixel of the mammographic image at a given pixel. For the pixel value of a mammographic image m_image (x, y), it can be stated that a certain spatial region is affected by the image m_image (x-p, y-q) at the coordinates defined by its neighbouring pixel values at a defined area M, for (p, q) epsilon M. It can be expressed mathematically as follows:

$m\_img(x,y) = \sum_{(p,q)\in N} \sum m_{(p,q)}\, k(x-p) + (y-q) + u(x,y)$ (1)

Where, M denotes the neighboring pixel value of mammographic image $m\_img\ mimg\ x, y$ that is involved in the assessment of linear sum as M means the neighbouring image pixel. (x, y)Denotes the coordinate value of mammography image is $k\ a$, $b.\ u\ x, y$ denoted the noise signal in a mammography image. Images in the image region were denoised using additive and blurred noise in the picture. Then the original image can be written as:

$m\_img(l) = Bm\_img(k-1) + u(k)$ (2)

Here, $m\_img(l) = [m\_img_0(l), m\_img_1(l), \ldots, m\_img_n(l)]^T$

3.2.2 Removal of Screen Film Artifact

In the context of a mammographic picture, screen film artifact is a structural trait on the image that signifies something that can be identified and contains such information, in this case the title and age of the patients. To overcome the problem of screen film artifact this paper is using the Text stroke edge height estimation method. The steps that are contained in this algorithm are mentioned below.

Step 1: Let the input mammography image be denoted as m_img(X,Y). First the original image of breast mammography is converted to the binary image by first applying the Otsu's threshold function which will let us define a binary image as the Bim_img(X,Y) in the following manner.

Step 2: In the second step, the scheme employs a scanning method whereby a binary image is scanned from top to bottom and left to right on the x axis. The aim of the scan is to identify the set of neighbouring pixel values that are the same in intensity as the pixel being scrutinised in the centre. When one of a certain intensity pixel values has been identified, draw connections to the same set of pixel values that are next to it. Neighbours' pixels being used in distance-based component labelling.

Step 3: Thirdly, commence by eliminating the recorded structure whose axial height measures a minimum internal volume. In doing this, artifacts produced on the film screen are eliminated like m_img(X,Y).

3.2.3 Removal of RF Noise Artifact

Due to the environment in the procedure room of taking mammography image some unavoidable noise signals are included in the image. This type of noise signal is called as radio frequency noise (RF). Because of excessive electromagnetic emissions is available in the scanning devices, mammography image contains RF noise artifact. By using Wiener Filter RF noise is removed from the mammography image. The wiener filter $wie(a,b)$ is applied mammography image $m\_img'(a,b)$ by taking product and applying Fourier transform and it produce$m_i\ (a,b)$. Wiener filter is evaluated as:

$wie(a,b) = \frac{H(\omega_1,\omega_2)S_{u,u}(\omega_1,\omega_2)}{|H(\omega_1,\omega_2)|^2 S_{u,u}(\omega_1,\omega_2) + S_{n,n}(\omega_1,\omega_2)}$ (3)

Where, S_ {(u,u)} and S_{(n,n)} are the Fourier Transform of the autocorrelation function of the mammogram image, and this is used for RF noise removal from the meager mammogram.

3.2.4 Normalization

For processing, the input mammography image of size 512 × 512 pixels can be enhanced using the max-min normalization technique. The range of intensity is between [0, 1] which can be calculated through the following expression

$p_i = \frac{(b_i - \min(b))}{(\max(b) - \min(b))}$ (4)

Accordingly, in the image b_i, where i=1,2,3, ...m, p_i denotes the intensity values that have been normalized such that their maximum and minimum values are given the notation of b_max and b_min, respectively. Normalization results in reducing the dimensions of the image to 256 * 256 pixels.

3.3 Feature Extraction

In this paper texture and shape features are extracted from the mammographic images to enhance the retrieval accuracy of mammographic images in the large dataset. Therefore, Principal component analysis (PCA) is utilized for texture feature extraction while the wavelet transform (WT) technique is used for shape feature extraction in the medical image.

3.3.1 Texture Feature Extraction using PCA

Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms high-dimensional image data into a lower-dimensional space, where the resulting features are mutually uncorrelated (orthogonal). The goal is to retain the most significant variations in the data while discarding redundancy. In the context of mammography image analysis, PCA can be effectively applied to extract texture features that are both computationally efficient and capable of reducing classification error rates.

Algorithm 1: Texture Feature Extraction using PCA

Step 1: In order to make the original pixel values of mammography image consistent, mean value of the feature is subtracted from all sample pixel values of the mamography image as shown by

$$\bar{B_j} = \frac{1}{m}\sum_{i=1}^{m} B_{ij} \qquad (5)$$

Step 2: Evaluate the covariance matrix $Cov\ (cov = (B_{jl})_{m\times m}$ where m is the number of features, B_{jk} is correlation between j^{th} and l^{th} feature value; where $j = 1,2, \ldots, m; l = 1,2, \ldots, m.$

$$Cov = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1m} \\ B_{21} & B_{22} & \cdots & B_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m1} & B_{m2} & \cdots & B_{mm} \end{bmatrix} \qquad (6)$$

Step 3:Evaluate the eigen value of λ_i and eigen vector value is ei

$$\lambda_i ei_i = Ceig_i \qquad (7)$$

Step 4: Preserve the values of eigen vector in a descending order where the highest value comes first so that one gets large as one goes down the list.

The principal components will also be ordered by their contribution ratio, which is m> ... ≥ λ_1 ≥ λ_2 and so on. In addition, determine the contribution rate of each of the principal components. The contribution rate is provided below:

$$\frac{\lambda_l}{\sum_{l=1}^{m}\lambda_l} \qquad (8)$$

Step 5 The original matrix B(Ob) is converted to matrix D. The final formula is:

D = Djl, where j=1,2…m and [illegible]…m.

Transform the original matrix B into new matrix D $(D = (D_{jl})_{n\times n1}$, where $j = 1,2 \ldots m$ and $l = 1,2, \ldots m.$

$$D = B \times \lfloor fe_1, fe_2, \ldots, fe_m \rfloor \qquad (9)$$

Where, ⟦fe⟧_1,⟦fe⟧_2,…⟦fe⟧_[illegible] indicates a new feature space which consists of m1 vector feature values and m2 feature values where m is the PCA extracted features. In the same manner as step 1 in the Algorithm 1, the mammographic image matrix undergoes transformation once again with h number of vector values to derive covariance matrix. As the number of dimension of vector values increases, the dimension of this covariance matrix also increases. In this manner, utilize the eigenvectors and form the covariance matrix in ascending order. The feature vector is used to extract and represent image spectral features.

3.3.2 Shape Feature Extraction using Wavelet Transform

Da[illegible]unau (2007) says that the wavelet image recompilation procedure is to the left and can be expressed as the Fourier spectrum. The wavelet transform is basic in edge detection methods and provides the necessary efficiency. The wavelet transform makes it possible to resolve the problem of localizing the image frame in distortions. So, the wavelet method decomposes the shape of the mammography image with three dimensional influences, horizontal, vertical and diagonal to create a complete picture. The same researchers confirm that image edges are enhanced with the Haar filter bank which works as both low and high pass filters. Thus, they conclude that any specific lateral low frequencies on the mammography image highlight features of edges and curves whereas high impulses provide texture characteristics. Therefore, to explain the wavelet transformation, it is expressed as follows:

$$l_0[m] = \begin{cases} \frac{1}{\sqrt{2}}, m = 0, -1 \\ 0, otherwise \end{cases} \qquad (10)$$

The high pass filter is represented by:

$$h_1[m] = \begin{cases} \frac{1}{\sqrt{2}}, m = 0 \\ -\frac{1}{\sqrt{2}}, m = -1 \\ 0, otherwise \end{cases} \quad (11)$$

In the given input mammography image of $M \times M$ and applying the Haar transform F will be evaluated as: $T = HFHT$, here H is the Haar basis functions. H can represented in matrix format as:

$$H_m = \begin{cases} l_0(0) l_0\left(\frac{1}{m}\right) \cdots l_0\left(\frac{M-1}{M}\right) \\ l_1(0) l_1\left(\frac{1}{m}\right) \cdots l_1\left(\frac{M-1}{M}\right) \\ \cdots \quad \cdots \quad \cdots \quad \cdots \\ l_{M-1}(0) l_{M-1}\left(\frac{1}{m}\right) \cdots l_{M-1}\left(\frac{M-1}{M}\right) \end{cases} \quad (12)$$

2 X 2 Haar transformation matrix is defined by:

$$H_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad (13)$$

Figure 3 shows that Hierarchical Subband Technique of mammography image

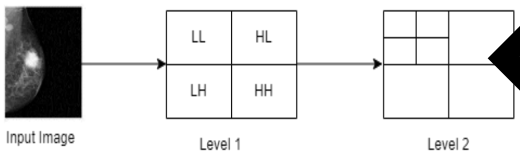


Figure 3 Hierarchical Sub band Technique

In Figure 3, the depiction of DWT approach based system with subbands is demonstrated. A feeling of separations of the original mammography image into core parts known as subbands is created in the text. Also, let us present to the reader the original image further filtered horizontally and vertically to obtain greater detail. This leads to the formation of four subbands, as categorised below: • Low Frequency Horizontal and Vertical Layer 1 subband LL 1 – Low Frequency Horizontal and vertical Frequencies in the image component as well as other elements. • High Frequency Horizontal and Vertical Components 1 subband HH1- High Frequency Horizontal and vertical frequencies about respective images. • Low Frequency Component; High Frequency subband LH1 - These include horizontal low frequency and Vertical high frequency components. • Horizontal High Frequency and Vertical Low Frequency subband HL1 - Components containing high horizontal frequencies and low frequencies in the vertical direction. The filters are again applied in the last subband. Again, LL2, HH 2, LH 2, HL 2, in that order is what it has decomposed into. These are depicted in Figure 4

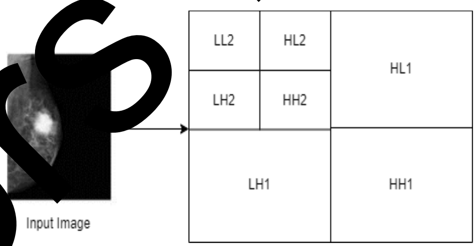


Figure 4 Decomposition Map

By decomposing the original mammography image using a wavelet transform into three components of horizontal, vertical and diagonal. The edges and curves of the image are detected by this component in a prominent way. To extract the shape, features, these components are used.

3.4 Feature Fusion using canonical correlation analysis (CCA)

The main aim of using CCA is to fused the feature vector of texture feature and shape feature of the mammography image. These feature vectors are represented by two variables. The correlation between the projections of twofeature vector variables are mutually maximized [26-28]. Let us considered the $t_1 \in R^m, s_2 \in R^n$ be the two-feature vector set of variables. CCA detects the pair of directions ω_{t_1} and ω_{s_2} which maximize the correlation between the projections of two canonical vectors are $T_1 = \omega_{t1}^T t_1$ and $S_2 = \omega_{s2}^S s_2$ and it is mathematically represented as:

$$arg \max_{\omega_{t_1}, \omega_{s_2}} \omega_{t1}^T R_{t_1 s_2} \omega_{s_2} \quad (14)$$

Here $R_{t_1 s_2} = t_1 {s_2}^T$ is the cross-correlation matrix of feature vector of t_1 and s_2. At the same time t_1 and s_2 must satisfy the constraints as below:

$$\omega_{t1}^T R_{t_1 t_1} \omega_{t_1} = \omega_{s_2}^T R_{s_2 s_2} \omega_{s_2} = 1 \quad (15)$$

Here $R_{t_1 t_1} = t_1 {t_1}^T$ and $R_{s_2 s_2} = s_2 {s_2}^T$

By applying Eqn (14) using Lagrange multipliers [21] we get the Eqn (16) as follows:

$$\begin{bmatrix} 0 & R_{t_1 s_2} \\ R_{s_2 t_1} & 0 \end{bmatrix} \omega = \mu \begin{bmatrix} R_{t_1 t_1} & 0 \\ 0 & R_{s_2 s_2} \end{bmatrix} \omega \qquad (16)$$

Here μ is the canonical correlation value and $\omega = \left[\omega_{t_1}^T, \omega_{s_2}^T\right]^T$ is the projected vector.
By this method it correlates the two-feature vector set values as a single fused feature vector value.

3.5 Classification using Fuzzy SVM with Modified Whale Optimization Algorithm (FSVM-MWOA) (Proposed)

As the last stage of this work, the authors have noticed how texture and shape features are acquired from the pre-processed pictures in the large dataset and the query input image and are combined through CCA And for the classification of similarity of images in the dataset this paper proposes to employ a Modified Whale Optimization Algorithm Fuzzy SVM (hereon referred to as FSVM-MWOA). This proposed work consists of two modules of applying fuzzy SVM and Modified Whale Optimization Algorithm.

3.5.1 Fuzzy SVM

Support vector machine (SVM) classification has the constraint of processing an overabundance of data because the SVM algorithm's complexity is proportional to the amount of data which it is operating on. In response to this challenge, this paper proposed fuzzy based SVM Classification algorithm. In training the dataset, fuzzymembership acquisition is a challenging task. The distance from a given input sample data to the class center in the space of a high dimensional function determines the value of fuzzy membership.

Let us consider the T is the set of labels$n = l$ and the binary classification of delinquent is defined by (x_n, z_n, s_n) via $n = l$. Based on the binary class label of $z_n \in \{-1,1\}$ and its contribution of data is obtained using $x_n \in R^m$. The fuzzy membership degree of SVM is defines by $s_n \in [0,1]$ and x_n belongs to z_n. The binary classification of delinquent of FSVM is required for the discrimination restriction form. Now quadratic programming problem of FSVM is obtained by:

$$Z_n[Ve^T \delta(x_n) + t] \geq 1 - \omega_j \qquad (17)$$

$$\omega_j \geq 0, n = 1,2, \dots, l \qquad (18)$$

By applying the Lagrangian function which solves the quadratic optimization problem and it transforming into dual problems by using the following Equations:

$$max_\beta \sum_{n=1}^{l} \beta_n \beta_o z_n z_o \, K(x_n, x_o) s \qquad (19)$$

Subject to:

$$\sum_{n=1}^{l} \beta_n z_n = 0 \qquad (20)$$

Here β_n is considered as Langrange multiplier in which value is not equal to 0 and n is represented as support vector and $K(x_n, x_o)$ is a kernel function.
Gaussian Kernel function is implemented by using:

$$K(x_n, x_o) = e^{\left(-\frac{1}{2\sigma^2}\|x_n, x_o\|^2\right)} \qquad (21)$$

The outcome of the FSVM is implemented by using:

$$p(x) = \left[\sum_{m=1}^{l} a\beta_n z_n \, K(x_n, x_o) + t\right] \qquad (22)$$

Here class label of x can be executed and FSVM technique classify the similarity of images in the large dataset. For enhancing and most accurate retrieval of similarity of images by applying the optimization algorithm. Therefore, this work implements the Modified Whale Optimization Algorithm.

3.5.2Modified Whale Optimization Algorithm

The large image dataset has been classified for similarity through the application of a modified whale optimization algorithm, MWOA, a method that draws its fundamentals from the characteristics of a whale and utilizes a spiral-shaped bubble surface to capture its prey [22]. The MWOA algorithm is based on a randomly selected whale and generates the best solution based on its best whale value. It can be represented as:

$$\vec{H}(it+1) = \overrightarrow{H_{rand}}(it) - \vec{P} . \vec{D} = \left|\vec{Q} . \overrightarrow{H_{rand}} - \vec{H}\right| \quad (23)$$

On this graph, it(it) is the iteration index and H(it+1) is the vector of location of the prey, its update with the multiplication operator. As for this time choice, three whales are chosen randomly and do not take into considerations the position of the leader. It follows that placement (23) may be re-written in the following way:

$$\vec{H}(it+1) = \overrightarrow{wt_1} * \vec{H}_{rand1} + \vec{y} * \overrightarrow{wt_2} * \left(\vec{H}_{rand2} - \vec{H}_{rand3}\right) + (1 - \vec{y}) * \overrightarrow{wt_3} * (\vec{H} - \vec{H}_{rand1}) \qquad (24)$$

Where, $\vec{H}_{rnd1}$, $\vec{H}_{rnd2}$ and $\vec{H}_{rnd3}$ are randomly select the solutions (prey). $\overrightarrow{wt_1}$ is randomly choose the value between [0,0.5] . $\overrightarrow{wt_2}$ and $\overrightarrow{wt_3}$ values between [0,1] randomly. $\vec{y}$ decreases the value and make it smoothly between exploration and exploitation by using

$$\vec{y} = 1 - \left(\frac{it}{Max_{it}}\right)^2 \qquad (25)$$

Here, n indicates the current iteration number while Max_it signifies the maximum allowed iteration number. The algorithm is given below:

Algorithm 2: Modified Whale Optimization Algorithm (MWOA)

Step 1: Initialize Population $\overrightarrow{H_i}(i = 1,2, \dots m)$, maximum iteration max_it, function of fitness Fi_n.

Step 2: Initialize parameters of WOA $\vec{D}, \vec{d}, \vec{E}, \overrightarrow{s_1}, \overrightarrow{s_2}, \overrightarrow{s_3}, k$ and modified parameters $\overrightarrow{wt_1}, \overrightarrow{wt_2}, \overrightarrow{wt_3}$.

Step 3: Initialize $it = 1$.

Step 4: Convert the solution into binary values either 0 or 1

Step 5:compute the fitness value Fit_n for each $\overrightarrow{H_i}$.

Step 6:Identify best individual value by $\overrightarrow{H^*}$.

Step 7:While$it \leq \max_iter$ do

Step 8:for$i = (1; i < m + 1)$do

Step 9: If $(\overrightarrow{s_3} < 0.5)$ then

Step 10: If $\left(|\vec{D}| < 1\right)$ then

Step 11:Update the $\vec{y}$ by the exponential form by using Eqn (25)

Step 12: Update current position of agent for search by using Eqn(24)

Step 13: Else

Step 14: Choose three search agents randomly $\vec{H}_{rnd1}, H_{rnd2}, \vec{H}_{rnd3}$.

Step 15: End if

Step 16: Else

Step 17: Update current position of agent for search by using

$$\overrightarrow{H}(t+1) = \overrightarrow{D'} \text{[illegible]} \cos(2\text{[illegible]}) + \overrightarrow{H^*}(t)$$

Step 18: Endif

Step 19: End For

Step 20: for $(i = 1; i < m + 1)$do

Step 21: Evaluate $\overrightarrow{H_i^*} = Gaussian\left(\mu\overrightarrow{H^*}, \sigma\right) \text{[illegible]} \times \overrightarrow{H^*} - \eta' \times \overrightarrow{P_i})$

Step 22:End For

Step 23: Update $\vec{D}, \vec{d}, \vec{E}, \overrightarrow{s_3}, k$

Step 24: Binary optimizer th[illegible]ed so[illegible]tion/prey by using

$$\overrightarrow{H_n^{(t+1)}} = \begin{cases} 1 & if\ sigmoid(H_{best}) \geq 0.5 \\ 0 & Otherwise \end{cases}$$

$$sigmoid(H_{best}) = \frac{1}{1 + exp^{-10(H_{best-0.5})}}$$

Step 25: E[illegible]uate fit[illegible]s value Hi_n for each $\overrightarrow{H_i}$.

Step 26: [illegible]ind be[illegible]dividual value by $\overrightarrow{H^*}$.

[illegible]p 27:it [illegible] $+ 1$

Step [illegible]nd While.

Step 29: Return$\overrightarrow{H^*}$.

In the improved whale optimization approach, the based on the hyper parameter is used as an input value and is subjected to evaluating the fitness function. Re-evaluate the individual performance corresponding to the fitness value. Keep on executing the process in a repeat cycle, until optimum solution is achieved.

3.6 Retrieval of Image

In this work retrieval of similarity images from the large dataset this paper uses Euclidean distance. The strong image uses the query image to fetch similar mammography images from the collection of images. The Euclidean distance formula is stated below:

$$Ed = \sqrt{\sum(x_i - y_i)^2} \qquad (26)$$

The minimum distance value signifies an exact match with the query.

4. Result & Discussion

The proposed feature extraction techniques using a Fuzzy SVM based modified whale optimization algorithm for the retrieval of mammmogram images, effectively locate the required similar images from the dataset.

4.1 Data Set Description

The images in the MIAS database, which were scanned with a pixel edge of 60 microns but lo[illegible] to a pixel edge of 150 microns during the process, have been trimmed and downscaled to dimensions of 1024×1024 pixels. For the dataset CBIS-DDSM, images provided are in the DICOM 16 Bit format and have a resolution of 3232 x 5295 pixels (width x height).

4.2 Performance of parametric measures

One of the features of this proposed work of efficiency retrieval of similar image from large data set is measured, calculated and evaluated in order to compare existing algorithms with that of SVM[23] FSVM [24] what have been explained earlier.

Precision is the positive predictive value (PPV) as previously described. It evaluates the true positives of the population with all positive values with the help of

$$Precision = \frac{TP}{TP+FP} \qquad (27)$$

Recall is a technique, it computes the true negatives for all negative numbers.

$$Recall = \frac{TP}{TP+FN} \qquad (28)$$

$$F - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (29)$$

MCC (Matthews Correlation Coefficient)

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TP+FN)}} \qquad (30)$$

False Rejection Rate (FRR)

$$FRR = \frac{FN}{FN+TP} \qquad (31)$$

It can be seen that recall and precision can be combined to one value known as the F-Score which comprises both factors. In graph form, the F-Score would range from a maximum of 1 to a minimum of 0, no further or less. With regards to the MCC, its regression one would be the aforementioned which is constrained within the interval [-1, +1]. For example, Table 1 depicts the description of dataset.

Table 1 : Description of Dataset

Categories of data	MIAS	DDSM	MIAS-DDSM
Benign	25	680	735
Malignant	25	570	590

Table 2 depicts precision and recall as parametric metric measures.

Table 2: : Performance of Metric Measures

Algorithm	Precision	Recall	F-Score
SVM	75.45%	83.24%	76.31%
FSVM	83.52%	83.15%	84.76%
FSVM-MWOA (Proposed)	89.42%	92.13%	94.34%

The proposed approach FSVM-MWOA attains an admirable 89.42% precision rate. The recall rate was recorded at 92.13%, while the F-score achieved 94.34%. Graphical correlations of FRR and MECC of several algorithms are presented in Figure 5.

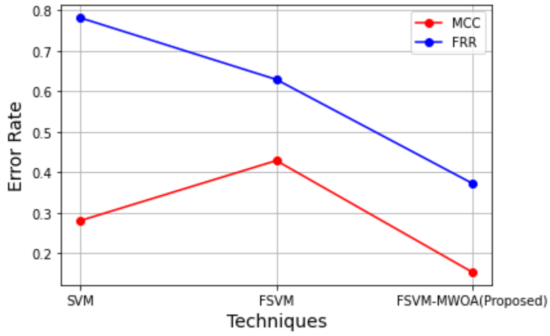
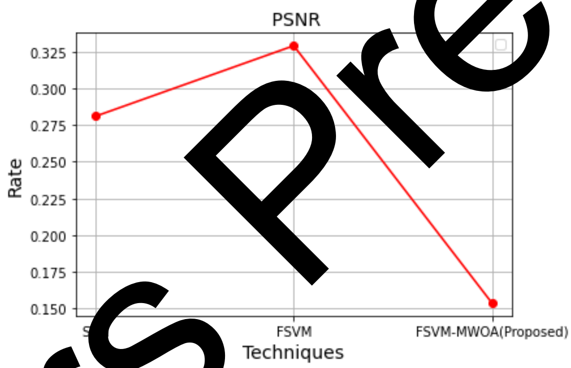


Figure3 Graphical representation of the MCC and FRR

As salient in Figure 3, the FRR and MCC are executed in diverse ways. In particular, the re[illegible]men[illegible] work of [illegible]VM-MWOA acquired MCC = 0.854, FRR = 0.373. Concerning the SVM MCC stated is 0.281, F[illegible] 0.2[illegible] for [illegible]VM 0.429 FRR acquired 0.092 are observed. The Image Quality to Noise Ratio value PSNR (Peak Sig[illegible] to Nois[illegible] Ratio) [illegible]ssessed as a measure of the images quality during suppression of noise in the mammographic image [illegible]g a me[illegible] filter as follows:

$$PSNR = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [mh(i,j) - nk(i,j)]^2 \quad (25)$$

Let M and N be the row and column counts respectively. nk(i,j) represents the noisy image in the context, while mh(i,j) is for the monochromatic image. Figure 4 PSNR v[illegible]ues o[illegible] different algorithms may be presented for comparison.



[illegible]gure4 PSNR Rate

According to the analysis of Figure 4, [illegible]r FSVM-MWOA work demonstrates minimum error rate. Figure 5 indicates the accuracy rate of different te[illegible]nique[illegible]mpl[illegible]ed.

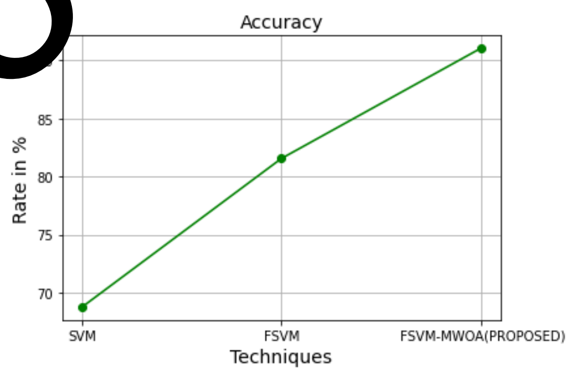


Figure 5 Accuracy

According to Table 5, the accuracy of the Proposed work of FSVM-MWOA was achieved with 95.04%, SVM was 68.76% and FSVM was 81.54%. Figure 6 indicates the duration of SVDD based approach for retrieval of mammography image from large dataset,

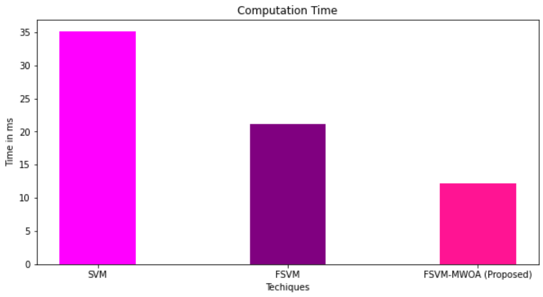


Figure 6 Computation Time

Figure 6 revealed that the average computation time of the work in this proposal is the least in comparison to the other existing techniques. Figure 7 demonstrates retrieval of sample images.

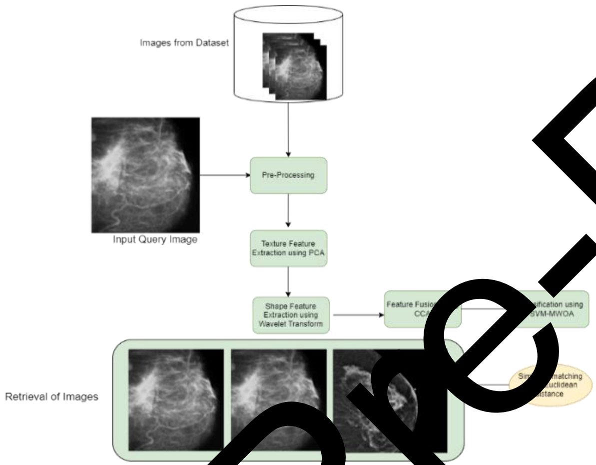


Figure 7 Sample Retrieval of Mammography Image

The viewing of Figure 7 depicts an example of retrieval of mammogram images from the database. Here three images are retrieved from the database based on the closeness of the query images employing Euclidean distance. First two images are more relevant (similarity) images whereas the third one is a dissimilar one.

4. Conclusion

This paper demonstrated the retrieval of mammography images from the large dataset by applying fusion of feature extraction using CCA. In order to get more accurate similarities of retrieval images this paper developed extracting the texture feature by using PCA and for the shape feature using Wavelet transform. Classification of similarities of images by using proposed work of fuzzy based SVM with optimized algorithm of modified whale optimization algorithm. The accuracy rate of Proposed work of FSVM-MWOA got 9[illegible]4 %, SVM got 68.76 % and FSVM got 81.54%. In future, this work may extend up to implement for the fine tuning of retrieval of images by the removal of other artifacts available in the realistic images.

Acknowledgement

The authors are grateful to M.Kumarasamy College of Engineering, Sri Eshwar College of Engineering, Kebri Dehar University and Balls University for providing research environment to carry out the study.

Funding Support

This work is not sponsored by any agency.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

The data that support the findings will be provided on requirement.

REFERENCES

[1]. Tariq Mahmood, JianqiangLi , Yan Pei, Faheem Akhtar, Azhar Imran, And Khalil Ur Rehman," A Brief Survey on Breast Cancer DiagnosticWith Deep Learning Schemes Using Multi-Image Modalities", Digital Object Identifier 10.1109/ACCESS.2020.3021343,IEEE Access, 2020.

[2]. Edson D. Carvalhoa, Antônio O.C. Filhoa, Romuere R.V. Silvaa, Flávio H.D. Araújoa, João O.B. Dinizb,c, Aristófanes C. Silvac, Anselmo C. Paivac, Marcelo Gattassd,"Breast cancer diagnosis from histopathological images using texturalfeatures and CBIR", https://doi.org/10.1016/j.artmed.2020.101845, 2020.

[3]. Sonia Jenifer Rayen1 · R. Subhashini," An Efficient Mammogram Image RetrievalSystem Usingan Optimized Classifier", Neural Processing Lettershttps://doi.org/10.1007/s11063-020-10254-3, Springer 2020.

[4]. Rafael S. Bressan, Pedro H. Bugatti ,PriscilaT.M. Saito," Breast cancer diagnosis through active learning in content-based image retrieval", Neurocomputing 357, 2019.

[5]. Zhu, M.; Lv, Q.; Huang, H.; Sun, C.; Pang, D.; Wu, J. Identification of a four-long non-coding RNA signaturein predicting breast cancer survival. Oncol. Lett., 19, 221–228, 2020.

[6]. Bick, U.; Engel, C.; Krug, B.; Heindel, W.; Fallenberg, E.M.; Rhiem, K.; Maintz, D.; Golatta, M.; Speiser,D.; Rjosk-Dendorfer, D.; et al. "High-risk breast cancer surveillance with MRI: 10-year experience from theGerman consortium for hereditary breast and ovarian cancer", Breast Cancer Res. Treat, 175, 217–228, 2019.

[7]. Sadiq Jaafar Ibrahim, Halidu Ibrahim Umar, Abba Mukhtar, Abdulazeez Muhammad Ahmad," Content Based Image Retrieval in Mammograms: A Survey", DOI 10.4010/2016.1155, ISSN 2321 3361, IJESC, 2016.

[8]. R. Ashraf, M. Ahmed, U. Ahmad, M. A. Habib, S. Jabbar, andK. Naseer, "MDCBIR-MF: multimedia data for content-basedimage retrieval by using multiple features," Multimedia Tools and Applications, pp. 1–27, 2018.

[9] T. M. A. Basile, A. Fanizzi, L. Losurdo, R. Bellotti, U. Bottigli,R. Dentamaro, V. Didonna, A. Fausto, R. Massafra, M. Moschetta,P. Tamborra, S. Tangaro, and D. La Forgia, ''Microcalcification detectionin full-field digital mammograms: A fully automated computer-aidedsystem,'' Phys. Medica, vol. 64, pp. 1–9, Aug. 2019.

[10].R. Vijayarajeswari, P. Parthasarathy, S. Vivekanandan, and A. A. Basha,''Classification of mammogram for early detection of breast cancerusing SVM classifier and Hough transform,'' Measurement, vol. 146,pp. 800–805, Nov. 2019.

[11]. H. Cai, Q. Huang, W. Rong, Y. Song, J. Li, J. Wang, J. Chen, and L. Li,''Breast microcalcification diagnosis using deep convolutional neuralnetwork from digital mammograms,'' Comput. Math. Methods Med.,vol. 2019, pp. 1–10, Mar. 2019.

[12].G. V. Ionescu, M. Fergie, M. Berks, E. F. Harkness, J. Hulleman,A. R. Brentnall, J. Cuzick, D. G. Evans, and S. M. Astley, ''Predictionof reader estimates of mammographic density using convolutional neuralnetworks,'' J. Med. Imag., vol. 6, no. 3, pp. 1–15, 2019.

[13]. A. Amelio, "A new axiomatic methodology for the imagesimilarity," Applied Soft Computing, vol. 81, p. 105474, 2019.

[14]. K. T. Ahmed, M. A. Iqbal, and A. Iqbal, "Content basedimage retrieval using image features information fusion,"Information Fusion, vol. 51, pp. 76–99, 2018.

[15]. Shinde V, Rao BT "Novel approach to segment the pectoral muscle in the mammograms. In: Cognitiveinformatics and soft computing", Springer, pp 227–237, 2019.

[16]. Ashebir Dohe Dogoma, Dipti Kapoor Sarmah, " Survey on Content Based Image Retrieval forBreast Cancer Detection and Classificationfrom Mammography Masses", International Journal of Pure and Applied Mathematics", Volume 118 No. 18, 2697-2702, 2018.

[17]. Sonia Jenifer Rayen, · R. Subhashini,"An Efficient Mammogram Image Retrieval System Usingan Optimized Classifier", Neural Processing Lettershttps://doi.org/10.1007/s11063-020-10254-3, 2020.

[18]. ShubhiMittal, Atul Kumar Verma,"Content Based Image Retrieval Using K-Means Algorithm", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 7 , pp. 5562-5564, http://dx.doi.org/10.37622/IJAER/13.7.2018.5562-5564,2018.

[19] Lei Gao, Rui Zhang, Lin Qi, Enqing Chen, and Ling Guan, "The Labeled Multiple Canonical CorrelationAnalysis for Information Fusion", 2103.00359v1 [cs.CV] 28 Feb 2021.

[20] S. Mirjalili, S. M. Mirjalili, S. Saremi, and S. Mirjalili, "Whale Optim.Algorithm: Theory, Literature Rev., Appl. Designing Photonic CrystalFilters", Cham: Springer International Publishing, pp. 219-238,doi: 10.1007/978-3-030-12127-3_13, 2020.

[21]. Mansoor. Ts, Shoaib & Sharma, A K., "Content Based Image Retrieval System using SVM Technique", 2017.

[22]. Chiranji Lal Chowdhary , Mohit Mittal, Kumaresan P, P. A. Pattanaik andZbigniew Marszalek," An Efficient Segmentation and Classification Systemin Medical Images Using Intuitionist PossibilisticFuzzy C-Mean Clustering and Fuzzy SVM Algorithm", Sensors, 20, 3903; doi:10.3390/s20143903, 2020.

[23] M. M, K. D. V, D. Gunapriya, P. N, S. S. Karthik and S. A, "Machine Learning-Based Pre-Stroke Detection System," 2024 International Conference on Science Technology Engineering and Management (ICSTEM), Coimbatore, India, 2024, pp. 1-5, doi: 10.1109/ICSTEM61137.2024.10560875.

[24] A. Pradeepa, O. V. Shanmuga Sundaram, and N. Pushpalatha, "Graphical image of Trisomy Ultrascan related Total edge magic labelling", EAI Endorsed Trans Perv Health Tech, vol. 10, Mar. 2024.

[25] Aruchamy, P, Gnanaselvi, S, Sowndarya, D, Naveenkumar, P. An artificial intelligence approach for energy-aware intrusion detection and secure routing in internet of things-enabled wireless sensor networks. Concurrency Computat Pract Exper. 2023; 35(23):e7818. doi: 10.1002/cpe.7818