IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 1 August 2023, accepted 26 August 2023, date of publication 30 August 2023, date of current version 5 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3310245

RESEARCH ARTICLE

Fusion of Textural and Visual Information for Medical Image Modality Retrieval Using Deep Learning-Based Feature Engineering

SAEED IQBAL[1], ADNAN N. QURESHI[2], MUSAED ALHUSSEIN[3], IMRAN ARSHAD CHOUDHRY[1], KHURSHEED AURANGZEB[3], (Senior Member, IEEE), AND TARIQ M. KHAN[4], (Member, IEEE)

[1]Department of Computer Science, Faculty of Information Technology and Computer Science, University of Central Punjab, Lahore 54000, Pakistan
[2]Faculty of Arts, Society and Professional Studies, Newman University, Birmingham, B32 3NT Birmingham, U.K.
[3]Department of Computer Engineering, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia
[4]School of Computer Science and Engineering, UNSW, Sydney, NSW 1466, Australia

Corresponding author: Saeed Iqbal (saeediqbalkhattak@gmail.com)

This Research is funded by Researchers Supporting Project Number (RSPD2023R947), King Saud University, Riyadh, Saudi Arabia.

ABSTRACT Medical image retrieval is essential to modern medical treatment because it enables doctors to diagnose and treat a variety of illnesses. In this study, we present an innovative technique for selecting the methodology of medical images by combining textural and visual information. Knowing the imaging process behind an idea, such as a chest X-ray, skin dermatology, or breast histopathology image, may be extremely helpful to healthcare professionals since it can aid in image investigation and provide important information about the imaging technique used. We use deep learning-based feature engineering to do this, using both the textural and visual components of healthcare images. We extract detailed visual information from the images using a predefined Convolutional Neural Network (CNN). The Global-Local Pyramid Pattern (GLPP), Zernike moments, and Haralick are also used to physically separate the pertinent parts from the images' other visual and factual aspects. These essential characteristics, such as image modality and imaging technique-specific characteristics, provide additional information about the technology. We employ a feature fusion method that incorporates the depictions obtained from the two modalities in order to combine the textural and visual elements. This fusion process, which improves the discrimination capacity of the feature vectors, makes accurate modality classification possible. We conducted trials on a sizable dataset consisting of various medical images to assess the effectiveness of our proposed method. The results indicate that, in comparison to conventional methods, our technique outperforms modality retrieval, with a precision of 95.89 and a recall of 96.31. The accuracy and robustness of the classification task are greatly creased by the combination of textural and visual data. Through the integration of textural and visual information, our work offers a unique method for recovering the modality of medical images. This method has the potential to greatly improve the speed and accuracy of medical image processing and diagnosis by helping experts rapidly and accurately identify the imaging technology being utilized.

INDEX TERMS Medical image retrieval, textural information, visual information, modality retrieval, deep learning, feature engineering, convolutional neural network.

I. INTRODUCTION

Due to the exponential growth of medical images stored in hospital databases, healthcare professionals are struggling to keep up with managing them effectively. Medical image retrieval techniques must be successful and efficient if they are to improve patient care, research, and development [1], [2]. To comprehend the distinctive qualities of particular patients that set them apart from the general patient population, healthcare practitioners look for relevant insights

The associate editor coordinating the review of this manuscript and approving it for publication was Bing Li.

from accumulated prior instances as well as literary data. The diagnosis, response to therapy, and prognosis of the disease are all tightly correlated with its radiographic phenotype. As a result, Content-based Medical Image Retrieval (CBMIR) has received the majority of attention in the development of medical modality retrieval. Based on the content of medical images, it enables the evaluation of image similarity [3].

The development of techniques to measure image similarity by capturing the unstructured features of clinical data has been the focus of the CBMIR challenge, which has previously been thought of as a computer science issue. In a query-by-example approach [4] CBMIR systems frequently use an example image as the query image in a query-by-example technique and extract a matching query vector that encapsulates the needed information. The reference images with the query vector's most comparable reference vectors are found in a database during the retrieval process. A method for calculating similarity based on visual characteristics, including aspects like color, texture, form, and spatial interactions between regions of interest, has evolved [5], [6]. CBMIR has recently integrated deep learning [7], [8] to efficiently fragment semantic characteristics from sample images and generate query vectors [9]. This method reduces the semantic gap between high-level semantic ideas and low-level visual aspects in medical images, as stated in [10].

The goals of healthcare professionals when seeking information are complex and go beyond the limitations of simply using image similarity calculations. To support their clinical judgments, healthcare practitioners frequently require access to a variety of case types. This includes instances that have had comparable clinical results in the same area in the past, cases that have had similar findings in various locations, and even cases that have had different findings in the same location. Furthermore, it might be difficult to start the search without a ready sample image when looking for specific clinical findings from online medical image archives, which have been more and more common in the latest years [11]. In the realm of clinical medicine, uncommon instances, including rare illnesses and clinical findings, possess substantial reference value [12], [13], yet healthcare professionals may face difficulties retrieving relevant images from a database due to the scarcity of similar example images. These challenges highlight the need for more advanced information-seeking approaches in healthcare.

Even though image retrieval methodologies have been thoroughly investigated for normal image retrieval jobs, applying the retrieval methodology in medical images, particularly radiology images, still poses a significant challenge [14], [15]. One reason for this difficulty is that medical images are complex to examine compared to normal images due to various factors, including intricate imaging parameters, connections between various illnesses, as well as minute differences between images with various diagnoses [6], [16]. Despite these challenges, researchers have made efforts to evolve medical image retrieval methodlogies in present years. Much of the literature focuses on retrieval based on features extracted through handcrafted or shallow learning approaches from various medical image modalities [17], [18], [19]. For retrieval systems based on shallow learning-based characteristics, large-scale datasets present difficulties [6]. Systems based on deep learning may be functional to successfully handle the needs of large-scale medical image retrieval. Deep learning has, however, only been partially used for content-based image retrieval, mostly because there aren't many large radiology datasets available [20]. Existing deep learning-based methods for retrieving images from radiology databases frequently use models that have already been trained on data from other image databases or use models with fewer layers that have been trained on smaller datasets [21], [22], [23], [24], [25]

Medical images are distinct from normal images, so pre-trained models may not serve as an ideal feature extractor in the medical field. The prevalence of deep learning is largely due to the accessibility of large training datasets. By training domain-specific approaches on a well-tagged dataset of huge images from a single modality, it is possible to fully use the promise of deep learning approaches for medical image retrieval applications. By using this strategy, the model is able to grasp the nuances and distinctive traits of the medical domain, producing retrieval results that are more precise and effective.

The research focuses on usability and searchability issues while discussing the drawbacks of query-by-example-based traditional CBMIR systems. Users who need to locate information without examples or who want to hone their search terms depending on search results have a usability constraint. The searchability restriction has to do with how challenging it is to find unusual images that are scattered across the database and distant from other images. The research suggests the creation of CBMIR approaches that can automatically classify and retrieve images in order to overcome these constraints and improve clinical practice and medical Decision Support Systems (DSS). Traditional CBMIR methods rely on labor-intensive, ineffective human annotation and expert tagging. The research instead highlights the requirement for automatic image classification and retrieval methods that make use of computer vision models. With the use of distance measurements like Euclidean distance, these models calculate query images by analyzing image qualities like color, texture, and structure in big datasets. They do this by recognizing image features in a high-dimensional feature space. However, the article emphasizes that despite their importance, present techniques frequently ignore deep learning models and linked visual characteristics between classes [26].

The research proposed using a query-by-patch approach, which has shown promising results in computer vision, to address the usability and searchability difficulties. Instead of depending exclusively on example images, this method enables users to convey shape, location, and item attributes

through a patch of the image. Medical image retrieval has proven to be difficult when it comes to patching exact anatomical information, which has prevented the technique from being used widely. Therefore, the research promotes the creation of a successful CBMIR system that does not rely on intricate anatomical patches or reference images in order to get over this constraint. In order to increase the strength and usefulness of medical image retrieval in diverse clinical contexts, the aim is to develop a system that can strongly recover pertinent medical images without the need for complex anatomical annotations. The primary contributions of this study are summarized as:

- Development of a novel technique that combines textural and visual information to determine the modality of medical images.
- Propose a deep learning architecture to extract in-depth visual information from medical images, including pretrained CNN models.
- Employing manual feature extraction methods like Zernike moments, Haralick, and Global-Local Pyramid Pattern (GLPP) to the images to extract key visual and statistical characteristics.
- Propose a fusion approach to blending the textural and visual components and improve the feature vectors' ability to discriminate.
- Our approach has the potential to be applied in both research and clinical settings, enabling the processing of data and offering second opinions that closely align with the intuitive expertise of medical professionals.

II. BACKGROUND RELATED WORK

Apart from direct diagnosis and lesion detection, the CBMIR system is another frequently utilized approach for examining medical images. CBMIR operates by utilizing the image itself to conduct a search on large datasets, rather than using keywords or database structure for queries. CBMIR has been extensively studied for its probable in clinical systems, including content-based approach to pathology images, where medical practitioners or radiologists can locate reference specimen slides from an actual database to make a diagnosis, and digital mammography reading for radiologists, where the mammogram retrieval system can provide visual aids to aid in a simpler diagnosis [27], [28].

Our hypothesis is that a CBMIR system has the potential to be highly beneficial in managing the pandemic situation in healthcare by enabling almost real-time retrieval of medical images for both doctor/radiologist examination from vast and multi-site datasets. The CBMIR system works by supplying images with labels that match the query image from a database that are both aesthetically and semantically meaningful. The matched image's label or diagnosis can, therefore, serve as a sign or hint for the query sample image. Image embedding, which entails converting images from their original domain to a more realistic, lower-dimensional manifold, is the essential component of a CBMIR system. Effective image representation through embedding can improve retrieval accuracy and speed. There have been several image embedding methods designed specifically for biomedical images, such as kernel approaches like hashing [29], [30], custom image filters such filter banks [31] and SIFT [32].

The progress in deep learning has also sparked the evolution of CBMIR architecture based on deep neural networks [33], including convolutional neural networks (CNN) for classification [34] and deep autoencoder [35], which have exhibited better achievement than other methods. Present deep learning-based methods, which explicitly train image representations (i.e., embedding) using the correlation underlying image characteristics and tags, might not be the best strategy for image retrieval jobs, though.

According to [36], pair-wise contrastive error might be more effective than cross-entropy cost, which is commonly used in present deep learning architectures, in leveraging label information. Consequently, CBMIR systems based on metric learning for examining histological images have been developed in recent years [37], [38]. Metric learning approaches, both traditional (non-deep learning) and deep learning-based, have also been discovered for analyzing CT [39] and magnetic resonance imaging (MRI) images [40]. However, to the best of our knowledge, there are no studies on metric learning for chest X-ray images in a clinical setting.

Retrieval with Clustering-guided Contrastive Learning (RetCCL), a system that enhances the performance of content-based Whole-Slide Image (WSI) retrieval for medical diagnosis, investigation, and teaching, is introduced by Wang et al. [41]. The framework can precisely evaluate image similarity and identify comparable patches or sub-regions in each image since it includes self-supervised feature learning, global ranking and aggregation algorithms. Using more than 22,000 slides, the scientists compared the framework's performance against other cutting-edge techniques for retrieving anatomical location and cancer subtype information. With the help of their feature representation, they also demonstrated enhanced patch retrieval.

Reddy et al. [42] describe Content-Based Image Retrieval (CBIR), which uses image data attributes rather than labels to scan a library of images for images that are visually similar to a user query. For the purpose of extracting high-level and deep properties from images, the usage of deep learning methodologies in particular Densenet-121-is described. The authors provide a technique that compares training images to the sample using a Bidirectional LSTM (BiLSTM) classifier in order to get relevant images. Using f-measure, recall, and accuracy measures, the method is compared to other image retrieval methods and performs better on the publicly accessible Corel dataset.

Singh et al. [43] explores Federated Learning (FL), which enables organizations to create a machine learning algorithm without sharing data, in the context of healthcare. FL entails disclosing model inputs while maintaining data privacy. The limitations and promise of FL are examined by the

writers. They offer an overview of existing FL techniques and recommend that conventional approaches to large-scale machine learning need to be reconsidered. Future trends for research communities are outlined in the article.

Gao et al. [44] proposed a methodology for the extraction of Chinese architectural history using images. At Jiangmen City, Guangdong Province, China, the project emphasizes using deep learning to extract images of the architectural history of the Chinese diaspora. To categorize and recover images based on resemblance, a Convolutional Neural Network Attention Retrieval Framework (CNNAR Framework) is devised and used in two steps. Using the heritage image datasets, the suggested technique obtains a classification accuracy of 98.3% and a mean Average Precision (mAP) of 76.6%. The algorithm's findings demonstrate a strong resemblance to the search images. Using the open datasets Paris500K and Corel5K, the suggested system obtains accuracy levels of 71.8% and 72.5%, respectively. The study finds that the CNNAR Framework may be used to analyze various topics' datasets in an efficient manner and gives information for further research on the cultural implications of diaspora Chinese homeland life.

The difficulty of using cutting-edge Convolutional Neural Networks (CNNs) on cellular gadgets that have limited resources is covered by Zhang et al. [45]. Offline Mobile Content-Based Image Retrieval (OMCBIR), a unique offline retrieval framework built on a minimal neural network model, is suggested by the authors as a solution to this problem. At the bottleneck phase, they provide pointwise group convolution and channel shuffle, and they suggest an incredibly light network using attention-based tiny networks (ALNet). For every dataset in OMCBIR, ALNet outperforms MobileNetV2 while decreasing model parameters by more than 62% and model size by more than 63%. To show the effectiveness of the suggested OMCBIR architecture, the authors undertake comprehensive experiments on five open-source datasets.

Hashimoto et al. [46] suggests a case-based Similar Image Retrieval (SIR) approach for histopathology images of malignant lymphoma marked with hematoxylin and eosin (H&E). In order to concentrate on tumor-specific areas and use immunohistochemistry (IHC) labeling features to establish acceptable similarity across diverse malignant lymphoma instances, the technique uses attention-based multiple-instance learning and contrastive distance metric learning. In a test with 249 individuals who had malignant lymphoma, the suggested technique demonstrated higher assessment methods than the standard case-based SIR methods. The similarity measure was also subjectively assessed by pathologists using IHC labeling trends, and they determined that it was suitable for illustrating the resemblance of H&E labeled tumor images for malignant lymphoma.

Mahmoud et al. [47] used different pre-trained CNN models for the retrieval of COVID-19 images using different image modalities. The high frequency of chest conditions including pneumonia and COVID-19 is discussed in the study, as well as the possibility of chest X-ray and CT-scan images for diagnosing these conditions. In order to create a functional feature descriptor, the study suggests a content-based image retrieval technique based on pre-trained CNN models like ResNet-50, AlexNet, and GoogleNet. While comparing images, the system compares them using similarity indices like City Block and Cosine. With accuracy rates of up to 99% for COVID-19 identification, ResNet-50 and GoogleNet are shown to outperform the best when the suggested method is tested on chest X-ray and CT scan datasets. Using a k-nearest neighbor classifier during voting increases accuracy by 1% to 4%.

A CBIR system for medical images was developed by Agrawal et al. [48], with a focus on the early identification and categorization of lung illnesses using lung X-ray images. Deep neural models and transfer learning are used to build the system, and they are trained using datasets of typical COVID-19 Chest X-ray images. When compared to previous approaches, experimental analysis on the standard dataset revealed a considerable enhancement in accuracy and Area Under the Precision-Recall Curve (AUPRC) values.

In order to extract and categorize biomedical images from sizable databases, the study proposes a CBMIR application that uses deep learning. Three procedures are used in the proposed Multimodal Biomedical Image Retrieval and Classification (M-BMIRC) technique: feature extraction, resemblance analysis, and classification. For extracting features, the model combines a variety of deep and manually produced features. Hausdorff Distance is used to determine how similar two features are, and Probabilistic Neural Network (PNN) is used to classify data. The suggested model performs better, according to experimental experiments on two benchmark medical datasets, in terms of a variety of metrics, including Average Precision Rate (APR), Average Recall Rate (ARR), F1-score, accuracy, and Computation Time (CT) [49].

The usage of Content-Based Image Retrieval (CBIR) methods, which search through a big dataset for related images using visual metadata, is covered by Pradhan et al. [50]. Education, military, agriculture, remote sensing, biological research, clinical care, and medical imaging are just a few of the industries where CBIR methods are used. This study discusses standard retrieval methods and their shortcomings before focusing on the utilization of machine learning and deep learning methodologies in medical image retrieval. The contemporary retrieval methods make recommendations for future research trajectories.

The relevance of CBIR is discussed by Raja and Karthikeyan [51], with an emphasis on how it might be used in agriculture. It introduces a brand-new model called RSA-DLCBIR that measures similarity using Minkowski distance and deep learning techniques for feature extraction. The model performs better than other methods when evaluated against a benchmark dataset.

In order to handle vast amounts of data and improve clinical decision-making, Haq et al. [52] suggest a machine-assisted method for the automated retrieval of medical images with comparable content. To extract related images, the system combines a network community identification method with a deep learning-based feature-generating methodology. It beat cutting-edge medical image retrieval algorithms with an accuracy of 85%25 in retrieving identical images with illness labels when tested on two sizable chest X-ray datasets. It is asserted that the suggested approach is the first deep learning-based image retrieval system on a sizable chest X-ray database.

For extensive medical image extraction, Xu et al. [53] suggest a unique hashing technique termed Multi-manifold Deep Discriminative Cross-modal Hashing (MDDCH). The low multi-manifold structure retention throughout various modalities and the low discriminability of the hash code are the two key problems with the cross-modal hashing techniques now in use. By combining several sub-manifolds created from heterogeneous data to retain association between instances and by suggesting discriminative items to prepare each hash code encoded by hash functions unique, MDDCH solves these difficulties. For three benchmark datasets, the suggested method surpasses current state-of-the-art hashing techniques and is well-received by medical experts.

In this research, Liu et al. [54] present a Deep Self-taught Hashing (DSTH) algorithm that trains hash functions employing discriminative deep models and creates pseudo-labels for datasets lacking tags. Both supervised and unsupervised situations are supported by DSTH, which also lowers the time complexity without sacrificing accuracy. The scientists conducted comprehensive tests to evaluate DSTH with state-of-the-art approaches in six publically accessible datasets using two distinct deep learning architectures to learn the hash functions. They found that DSTH outperformed the others methodologies in all databases.

A novel unsupervised deep learning-based hashing technique for massive image retrieving is proposed by Deng et al. [55]. An encoder network, a generator network, and a discriminative network make up the generative adversarial framework used in the proposed technique. The model builds a semantic similarity matrix, which directs the hash code training steps, using both feature and neighbor comparisons. The suggested technique may train the encoder network to learn effective hash codes by adversarially training these networks. The empirical findings on three benchmarks demonstrate that the suggested approach beats a number of cutting-edge unsupervised hashing techniques and operates on par with well-known supervised hashing techniques.

Feng et al. [56] proposed a concept for the extraction of encoded images from distant servers is referred to in the article as encrypted image retrieval (EIR). Currently used methods include end-to-end deep learning models or manual feature extraction. These methods do, however, have shortcomings. The authors suggest a not-end-to-end EIR technique dubbed DHAN that encrypts DCT coefficients and uses deep attention networks to recover images in order to overcome these problems. Compared to previous techniques, investigations on two datasets confirm that DHAN offers good image security and enhances retrieval performance.

The Li et al. [57] covers the information loss and lack of sorting knowledge caused by binary hash coding's limitations in image retrieval. The authors provide an ensemble deep neural network for retrieving images that combines rating data through weighted voting and learns small hash codes. On three benchmark datasets, the approach is evaluated, and it produces comparable results.

Fang et al. [58] suggest a study of a deep metric learning approach with mirror attention to improving the distinguishing characteristics of tiny and scattered lesions in fundus images. By offering analogous situations, this framework can aid ophthalmologists in making evidence-based medical decisions. The suggested strategy uses a fine triplet loss to enhance the ranking efficiency of successful items while encoding the lesions into image characteristics. Using the biggest fundus dataset for the diagnosis of Diabetic Retinopathy (DR), the suggested technique outperforms competitors in terms of precision, particularly when it comes to enhancing the indexing quality of DR grades that contain microaneurysms and hemorrhages. To provide highly discriminative image descriptors, the suggested mirror attention may be used for backbones purchased off-the-shelf and learned successfully from beginning to end for additional medical images.

In order to help in diagnosis, Jiji and Raj [59] suggest a content-based image retrieval approach for skin lesion images. The suggested architecture retrieves images and the name of the illness classification from an image database by extracting aspects from the image, such as form, texture, and color. The suggested technique retrieves extensive referencing materials for diagnostic purposes using a feature vector and a classification and regression tree. The studies produced high specificity of 97.25% and sensitivity of 91.24%, demonstrating the strong impact of the suggested design on a computer-aided diagnosis of skin lesions.

Sucharitha et al. [60] proposed a unique methodology to ameliorate medical image retrieval systems by addressing the semantic gap using a consolidation of texture and shape features constructed from Relative Directional Edge Binary Patterns (RDEBP) and complex Zernike moments. The DBSCAN algorithm is used to cluster the features, and images are acquired from the coterminous cluster using a correlation metric. Experiments on two databases depict that the proposed technique surpasses other state-of-the-art methods with a 2-5% increase in accuracy. They suggest a technique for retrieving biological images utilizing global and local properties of an image and a short feature vector. A novel approach called Local Directional Edge Binary Pattern (LDEBP), which analyzes inputs from every

conceivable direction for each pixel, is used to extract the local characteristics. Lower-order Zernike moments are utilized to extract both the global and form aspects. In benchmark datasets like Emphysema-CT and OASIS-MRI, the integration of shape and texture descriptors outperformed state-of-the-art methods in biomedical image retrieval.

Patil et al. [61] discovered a unique architecture for effective image retrieval from repositories utilizing shape characteristics as CBIR. The methodology uses Support Vector Machine (SVM) as a classifier and region-based descriptors like Hu's seven moments and Zernike moments for feature extraction. The effectiveness of several distance metrics, including Euclidean, Chebyshev, Cityblock, Canberra and Standardized Euclidean (Seuclidean), is assessed in this article on a medical database with six classes, each of which contains 100 images. The research confirms that the suggested approach successfully retrieves images from the database, and the outcomes are evaluated against those of alternative approaches using accuracy and recall criteria.

Pandian and Balasubramanian [62] discovered a texture fusion strategy for T1 and T2 weighted MRI scans. The method involves extracting texture and shape features from brain tumor images, selecting features using Genetic Algorithm and Particle Swarm Optimization (PSO), and classifying brain tumors using Deep Neural Network and Extreme Learning Machine (ELM). The method is efficient, and effective, and reduces retrieval time while improving retrieval accuracy. The study shows the optimal classification accuracy findings utilizing DiCom images.

Baji et al. [63] discuss the challenge of classifying brain tumors in MRI scans and propose a technique to improve accuracy. The authors use k-means clustering to isolate the images into clusters and then analyze the texture features of each cluster. The bilateral symmetry measure is then used to calculate which cluster comprises the tumor, and connected component labeling is used to confirm the target cluster. The proposed technique is tested on 30 MRI images and attains an accuracy of 87%.

Sut et al. [64] proposed a novel concept to accurately detect diseases of the adrenal glands, and a novel machine-learning technique has been created. Preprocessing, scaling, feature extraction utilizing a Center-Symmetric Local Binary Pattern (CS-LBP), and feature selection using Neighborhood Component Analysis (NCA) are all part of the suggested technique. To create the best-performing model, the chosen characteristics were categorized using the K-Nearest Neighbor (kNN), Support Vector Machine (SVM) and Neural Network (NN) classifiers. Using kNN, SVM, and NN classifiers, the suggested technique has accuracy rates of 99.87%, 99.21%, and 98.81%, respectively. The new technique may be utilized to test for many kinds of adrenal gland diseases using CT scans.

Ma et al. [65] enhanced the complementarity of multi-level heterogeneous features through the proposal of an adaptive multi-feature fusion approach for image retrieval. Initial correlation scores of the search image to the objective dataset are computed after numerous low-level and high-level semantic characteristics based on deep learning are retrieved. The suggested technique uses statistically elaborated regions derived from cross-entropy normalization to allocate the merging weights of each element in an adaptable manner. The suggested technique outperforms existing methods, according to tests on three publicly available benchmark datasets, increasing the metrics mAP and N-S.

III. DATA AND MATERIALS

This section describes the datasets that were used in the study. The repository is composed of three primary medical image modalities: chest x-ray from NIH, Skin lesion from ISIC and breast tumor from BreakHis.

A. DATASET

For the purposes of this investigation, chest X-ray datasets were used. The National Institutes of Health (NIHChestX-ray8)'s dataset, which includes approximately 112,000 frontal-view X-ray images taken between 1992 and 2015 from more than 30,000 individuals, was the first dataset to be utilized. Each X-ray image in the collection may contain more than one positive illness diagnosis since it is multi-label. There are 16,630 male and 14,175 female patients in the collection, which comprises 67,310 PA view images and 44,810 AP view images [66].

Around 9,000 histopathological images of breast cancer tissue samples, separated into benign and malignant categories, are included in the BreakHis collection. The images were obtained through a microscope at different magnification settings and developed by Professor Gloria Bueno and her colleagues at the University of Castilla-La Mancha in Spain [67].

Around 33,000 skin lesion images gathered from diverse sources make up the ISIC 2020 skin lesion dataset, which is a publicly accessible resource. It contains binary categorization labels of ''benign'' or ''malignant,'' as well as details particular to the lesion. The dataset was produced in order to brace the creation of computer-aided diagnosis algorithms for melanoma and other kinds of skin cancer [68].

B. TEXTURAL FEATURE EXTRACTION

Haralick features are a set of texture features that are used for feature extraction in medical image retrieval. These features are based on the gray-level co-occurrence matrix (GLCM), which is a matrix that captures the joint probability distribution of pairs of pixels with a specific spatial relationship in the image.

The Haralick features are calculated from the GLCM using the following equations:

- First, the GLCM is calculated by enumeration of the number of times each pair of pixel intensity values occurs with a specific spatial relationship. The GLCM

is defined by the equation:

$$P(i,j,d,\theta) = \sum_{x=1}^{N}\sum_{y=1}^{M} I(x,y)I(x+d\cos(\theta), y+d\sin(\theta)) \quad (1)$$

where *P(i,j,d,θ)* is the element of the GLCM at *(i,j)* for a spatial relationship defined by a distance *d* and an angle *θ*, *I(x,y)* is the pixel intensity at location *(x,y)*, and *N* and *M* are the dimensions of the image.

- Haralick features calculation: These features are calculated from the GLCM using various statistical measures. Some of the most commonly used Haralick features include:
- Angular Second Moment (ASM): analyze the homogeneity of the image

$$ASM = \sum_{i=1}^{N}\sum_{j=1}^{M} P(i,j)^2 \quad (2)$$

- Contrast: finds the local variations in the image

$$Contrast = \sum_{i=1}^{N}\sum_{j=1}^{M} (i-j)^2 P(i,j) \quad (3)$$

- Correlation: To quantify the linear dependency between the gray levels of neighboring pixels

$$Correlation = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M}(i-\mu)(j-\mu)P(i,j)}{\sigma^2} \quad (4)$$

where σ and μ are the standard deviation and mean of the GLCM.

- Energy (E): measures the total amount of gray-level co-occurrences in the image

$$E = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M} P(i,j)^2} \quad (5)$$

- Entropy (En): measures the randomness or complexity of the image

$$En = -\sum_{i=1}^{N}\sum_{j=1}^{M} P(i,j)\log_2(P(i,j)) \quad (6)$$

We used this Haralick features eqs. (1) to (6) for various medical image modality retrieval and compared using similarity metrics such as Euclidean distance or cosine similarity.

Zernike moments (ZMs) are a set of orthogonal complex polynomial functions that are used for shape representation and feature extraction in image processing. In the context of medical image retrieval, ZMs are utilized to extract shape-based characteristics from medical images such as X-rays, MRI scans, and CT scans.

The Zernike polynomials are defined over a unit circle or disk and have the form:

$$Z_{nm}(r,\theta) = R_{nm}(r)\exp(im\theta) \quad (7)$$

where n and m are non-negative integers with $n \geq m$, r is the radial distance from the center of the disk, θ is the polar angle, and i is the imaginary unit.

The radial part of the Zernike polynomial, Rnm(r), is given by:

$$R_{nm}(r) = \sum_{k=0}^{\frac{n-m}{2}} (-1)^k \frac{(n-k)!}{k!\left(\frac{n+m}{2}-k\right)!\left(\frac{n-m}{2}-k\right)!} r^{n-2k} \quad (8)$$

where *!* denotes the factorial function. To calculate the ZM coefficients of an image, the image is first converted to grayscale and normalized. A circular or elliptical Region of Interest (ROI) is then defined around the object of interest. The ZMs are calculated by projecting the normalized gray-level image onto a set of orthogonal Zernike polynomials that describe the shape of the ROI.

The ZM coefficients are given by:

$$A_{nm} = \frac{n+1}{\pi} \iint_{\text{ROI}} Z_{nm}(r,\theta) I(r,\theta) r\, dr\, d\theta \quad (9)$$

where *I(r,θ)* is the normalized intensity function of the ROI.

The ROI's shape information is represented by the computed ZM coefficients, which are characteristics for retrieving medical images. These qualities may be compared to features gathered from other medical images using similarity metrics like Euclidean distance or cosine similarity. Zernike moments equations employ orthogonal complex polynomial functions to describe an object's shape in a image. We employ eqs. (7) to (9) to retrieve and analyse shape-based data from medical photos.

The Global-Local Pyramid Pattern is a brand-new method for feature extraction in medical image analysis that was suggested by [69]. Building a multi-scale representation of a image is necessary, with each scale capturing a different degree of information, ranging from the overall organisation of the image to regional patterns and textures. Applying a sequence of spiral filters to the original image in escalating sizes is the core concept. The filters capture bigger and wider receptive fields at each iteration, enabling the network to learn properties that are more and more global in scope. But as the filters become bigger, their sensitivity to minute details decreases, making it challenging to record regional patterns and textures.

The Global-Local Pyramid Pattern additionally includes a series of "local" convolutions that are executed in parallel with the global spiral at each size to get around this restriction. Smaller receptive fields of these local spirals enable them to pick up more local patterns and fine-grained features. The Global-Local Pyramid Pattern is able to capture a comprehensive collection of characteristics that we employed for extracting various modalities from the medical image repository by combining these global and local features at various sizes.

The eq. (10) is a mathematical representation of the Global-Local Pyramid Pattern.

$$G_i = f_g(I, \theta_g^{(i)}) + f_l(I, \theta_l^{(i)}) \quad (10)$$

where G_i is the feature map at the i-th scale, f_g is the global convolution operation, f_l is the local convolution operation, I is the source image, $\theta_g^{(i)}$ and $\theta_l^{(i)}$ are the sets of learnable parameters for the global and local convolutions, respectively.

The global convolution operation $f_g(I, \theta_g^{(i)})$ applies a set of filters with a large receptive field to the input image I, which captures global features such as the overall structure of the image. The finding of this calculation is a feature map that captures these global features. The local convolution operation $f_l(I, \theta_l^{(i)})$ applies a set of filters with a smaller receptive field to the input image I, which captures local patterns and textures in the image. The outcome of this computation is also a feature map that captures these local features. The final feature map G_i at the i-th scale is obtained by adding together the feature maps from the global and local convolutions.

C. VISUAL FEATURE EXTRACTION

Convolutional Neural Networks (CNN) have demonstrated substantial success in visual feature extraction tasks for medical image retrieval. CNN is a deep learning architecture commonly used for building visual features from images. The CNN design, which comprises a number of convolutional layers succeeding by pooling layers, enables it to understand intricate information in a hierarchical fashion. Next, depending on their resemblance, these learned characteristics may be utilized to categorize images or fetch images.

Large datasets of healthcare images and their related labels, such as the NIH Chest X-ray, BreakHis for breast tumors and the ISIC skin lesion dataset, are used to train CNNs for medical image retrieval. In order to categorize or retrieve medical images, the CNN must learn to extract pertinent aspects from the images, such as texture, geometry, and edges. Using a distance measure, such as Euclidean distance or cosine similarity, the CNN may be trained to extract characteristics from medical images that can then be compared to other images.

The mathematical equation for the convolutional layer of a CNN is represented as eq. (11):

$$H_{i,j,k} = \sigma\Big(\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{l=1}^{L} W_{m,n,l,k} I_{i+m-1,j+n-1,l} + b_k\Big) \quad (11)$$

where $H_{i,j,k}$ is the outcome of the k^{th} feature map at position (i, j), $W_{m,n,l,k}$ is the weight parameter of the k^{th} feature map for the m^{th} row, n^{th} column, and l^{th} channel of the input image I, b_k is the bias term for the k^{th} feature map, σ is the activation function, and M, N, L are the dimensions of the filter. The finding of the convolutional layer is downsampled using a CNN's pooling layer. The widely used pooling procedure is known as max-pooling, which has the following mathematical representation as eq. (12).

$$H_{i,j,k} = \sum_{m=1}^{p}\sum_{n=1}^{q} H_{(i-1)p+m,(j-1)q+n,k} \quad (12)$$

where $H_{i,j,k}$ is the outcome of the k^{th} feature map at position (i, j) after max-pooling, p and q are the dimensions of the pooling filter. The fully connected layer of a CNN is used to produce the final output of the model. Mathematically represented as eq. (13).

$$y = \sigma\Big(\sum_{i=1}^{N} w_i x_i + b\Big) \quad (13)$$

where y is the outcome of the fully connected layer, N is the number of neurons in the layer, w_i and x_i are the weight and input of the i^{th} neuron, respectively, b is the bias term, and σ is the activation function.

IV. PROPOSED METHODOLOGY

The CBMIR model proposed in this study follows a series of developmental procedures as depicted in figs. 1 and 3. Initially, handcrafted features were extracted from Harlick, Zernike moments and Global-Local Pyramid Patter (GLPP) and deep features were also derived using Deep Convolutional Neural Networks (DCNN) containing Separable Convolution block as depicted in fig. 2 from the input image(s). The test image is utilized to determine related images using a similarity measure based on Hausdorff Distance and retrieve them. Subsequently, the retrieved images were processed using the SVM method to determine their corresponding class labels.

Haralick features are statistical measures that capture several facets of texture information in an image. They are also known as texture features or texture descriptors. In image processing and analysis, especially medical image analysis, these characteristics are often employed. Based on the gray-level co-occurrence matrix (GLCM) of an image, the Haralick features are calculated. The GLCM accurately depicts the spatial connection in the image between pixel pairs with respect to distance and orientation. Different pixel pairs are counted together with their respective locations. We use the following procedures to extract the relevant characteristic from various medical image modalities, such as histopathological images of breast cancer, MRI or CT scans of brain tumors, and dermoscopic images of skin lesions. Denoising and image enhancement techniques are used during preprocessing to change the quality of low-contrast or noisy images. The effect of noise or artifacts on the ensuing feature extraction procedure must be reduced as much as possible. Following preprocessing, we quantized the data to lower the number of grey levels, which aids in lowering the computational cost of the GLCM. It converts the pixel intensities into a more condensed range of distinct grey tones. We locate the GLCM to examine the spatial relationships between pairs of pixels in the quantized image. Each component of this square matrix, which indicates the frequency of a certain pair of pixels occurring at a given distance and direction, is a pair of pixels. The Haralick features are a set of statistical measurements that we compute from the GLCM. It records several texture information

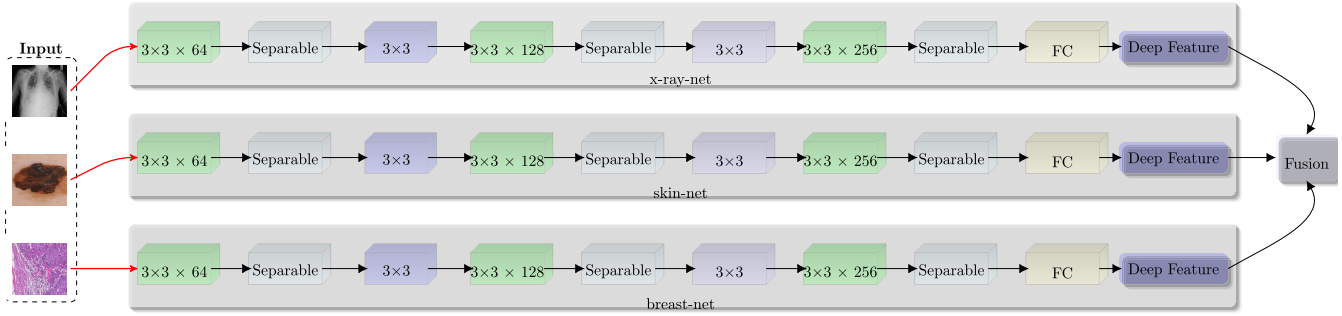

FIGURE 1. Feature extraction from multiple medical image modalities and fusion for classification.

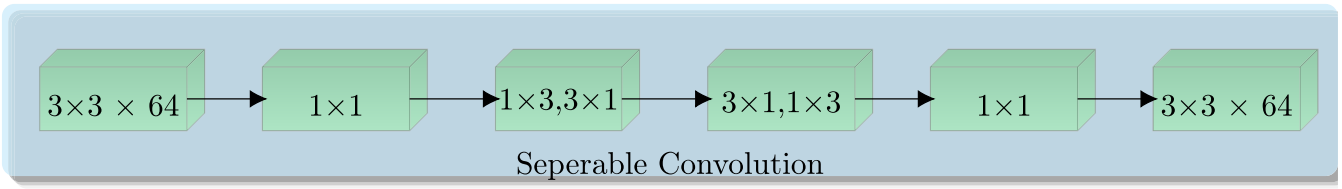

FIGURE 2. Separable Convolution block for our proposed CNN model.

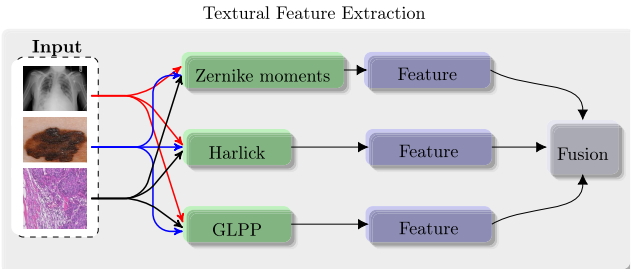

FIGURE 3. Integration of manual feature techniques from multiple medical image modalities for classification".

properties, including contrast, homogeneity, entropy, energy, and correlation. These features can gather crucial data on texture patterns from the spatial relationships between pixel pairs, which are helpful for a number of medical image processing tasks including tumor classification, tissue segmentation, or illness detection.

We employ Zernike moments in our second manual feature extraction strategy. The form and textural properties of objects in medical image modalities are described by a collection of mathematical attributes. It is founded on a set of orthogonal polynomials called Zernike polynomials, defined on a spherical domain. A full and orthogonal basis set for describing any shape or texture within a circular area is provided by these polynomials.

We use the same preprocessing phase to extract Zernike moments, which we previously explained while extracting Haralick features. We employ filtering, normalization, and other image-specific preparation methods. After preprocessing, we separate the lesion patch or other item or feature of interest from the background and extract the Region of Interest (ROI) containing it. In order to divide pixels or pixels depending on their intensity values, thresholding, region growing, and level set methods are used. The border of the segmented object is recovered using edge detection algorithms and other contour extraction techniques during the segmentation of skin lesions or foot ulcers. The retrieved boundary points are normalized to fit inside a unit circle throughout the normalization phases, which is necessary for the computation of Zernike moments. It guarantees that the features are scale, rotation, and translation invariant. After normalization, the Zernike moments are revealed. By integrating the intersection of the Zernike polynomials over the circular domain with the normalized boundary points. Each Zernike moment represents a distinct form or texture feature and correlates to a certain sequence and repetition. The results of Zernike moments provide a feature vector (1D) that describes the lesion or tumor's shape and texture in the medical image. In order to effectively extract features for tasks like classification, segmentation, registration, and anomaly detection in medical images, it provides a concise representation of shape and texture information.

We employ our suggested technique, Global-Local Pyramid Pattern (GLPP), to examine the spatial pattern of pixel intensities in a small neighborhood. Using clockwise and anticlockwise rotation, GLPP is able to extract important characteristics from many medical image modalities. By evaluating each pixel's intensity value in relation to its immediate surroundings, we get a binary code for each pixel in an image. When creating the code, a value of 1 is assigned if the neighbor's intensity is more than or equal to the center pixel and a value of 0 if it is lower. The identical process is carried out anticlockwise, and the binary codes are then translated to decimal values and the higher number is chosen to create the GLPP image. We must specify parameters such as the size of the local neighborhood (neighbor), the number of neighboring pixels (often written as P), and the central value for GLPP computation when extracting relevant

features using GLPP. We extract relevant information from the produced GLPP images using a variety of methods, including histogram analysis and statistical metrics (such as mean and variance). These characteristics can help with tasks like tissue characterization, lesion identification, or disease classification by offering insights into the patterns, abnormalities, and textural aspects that are present in medical images.

A. FEATURE FUSION

A highly effective approach for thorough and reliable feature representation in medical image analysis uses the combination of manual features such as Haralick features, Zernike moments, and Global-Local Pyramid Pattern features with auto-features generated by Convolutional Neural Networks (CNN). It makes it possible to combine manually created features that capture certain traits with automatically discovered features that collect more intricate and abstract data. In order to measure the texture aspects of an image and capture statistical data regarding pixel intensity connections, we first compute Haralick features. By examining the fluctuation in pixel intensities inside circular areas during the calculation of Zernike moments, it is possible to capture shape and texture information. Extraction of relevant features utilizing the global-local pyramid pattern, which uses pyramidal decomposition and analysis to gather information on several scales and levels of texture. We preprocess the medical images (e.g., normalization, scaling) for compliance with the CNN model after manually extracting the features. We present a custom-designed model to automatically extract features from the medical images and use pre-trained CNN architectures like VGG, ResNet, Inception, Xception, and DenseNet for auto-feature extraction. In order to get high-level feature representations, we freeze the final classification layers from the CNN. We integrate the manual features (Haralick, Zernike moments, Global-Local Pyramid Pattern) with the auto-features acquired from the CNN after feature extraction through CNN. As depicted in figs. 1 and 3, we concatenate or combine the feature vectors to provide a single feature representation for each image. We employ the fused feature representation as source to a machine learning method, such as Support Vector Machines (SVM), K-Nearest Neighbour (KNN), and Decision Tree for classification and analysis tasks on the fused feature vectors. These classifiers are trained using labeled training data, and their effectiveness is assessed using test data. To evaluate the results and efficiency of the fusion technique, we use relevant assessment measures including Precision, Recall, F1-Score, Sensitivity, and Specificity.

B. CLASSIFICATION OF FUSED FEATURES

Depending on the various medical image modalities, the classification outcomes utilizing conventional machine learning classifiers like SVM, KNN, and Decision Tree on fused features (Haralick, Zernike moments, GLPP) with auto-features generated by CNN might vary. We employ traditional machine learning classifiers like SVM, KNN, and Decision Tree to categorize these fused features. The SVM classifier is a well-liked one because it can handle high-dimensional data and nonlinear decision limits. SVM can successfully learn the intricate patterns and connections between the manual and automatic features when applied to fused features. The objective is to identify the best hyperplane in the feature space that best separates the various classes. The results of classification using SVM on fused features can provide an optimal separation and accurate classification, especially when the classes are well-discriminated in the feature space. Another popular classification algorithm is KNN. It is a non-parametric classifier that assigns a class label to a data point based on the labels of its nearest neighbors in the feature space. We apply to the fused features, KNN can capture the local structure and relationships between the manual and auto-features. The results of classification using KNN on fused features may heavily depend on the choice of the number of neighbors (K=3) and the distance metric (Manhattan or Euclidean) used. It performs well when the classes exhibit distinct clusters in the feature space. The third classifier is the decision Tree, it is a tree-based classifier that builds a hierarchical structure of decisions based on the features. We apply to the fused features, Decision Tree can capture the interdependencies between the manual and auto-features and learn a set of rules for classification. The results of classification using a Decision Tree on fused features can provide interpretable decision boundaries and rule-based classification. We experiment with different classifiers, tune their hyperparameters and analyze the performance using appropriate evaluation measurements such as accuracy, precision, recall, or F1-score to determine the best classifier for a specific classification task.

V. RESULTS AND DISCUSSIONS

To discuss the results of different classifiers, we highlight several key aspects to consider. We analyze the performance of each classifier using accuracy and other relevant performance measurements such as Accuracy depicted in eq. (14), Precision, Recall depicted in eq. (15), F1-score, or Area Under the ROC curve (AUC) and Mean Average Precision (MAP) depicted in eq. (16).

$$\text{Top-N Accuracy} = \frac{\text{Number of relevant images among top N}}{N} \tag{14}$$

$$P = \frac{\text{Number of relevant retrieved images}}{\text{Total number of retrieved images}} \tag{15}$$

$$R = \frac{\text{Number of relevant retrieved images}}{\text{Total number of relevant images}} \tag{16}$$

$$\text{MAP} = \frac{1}{Q}\sum_{q=1}^{Q} \text{AP}_q \tag{17}$$

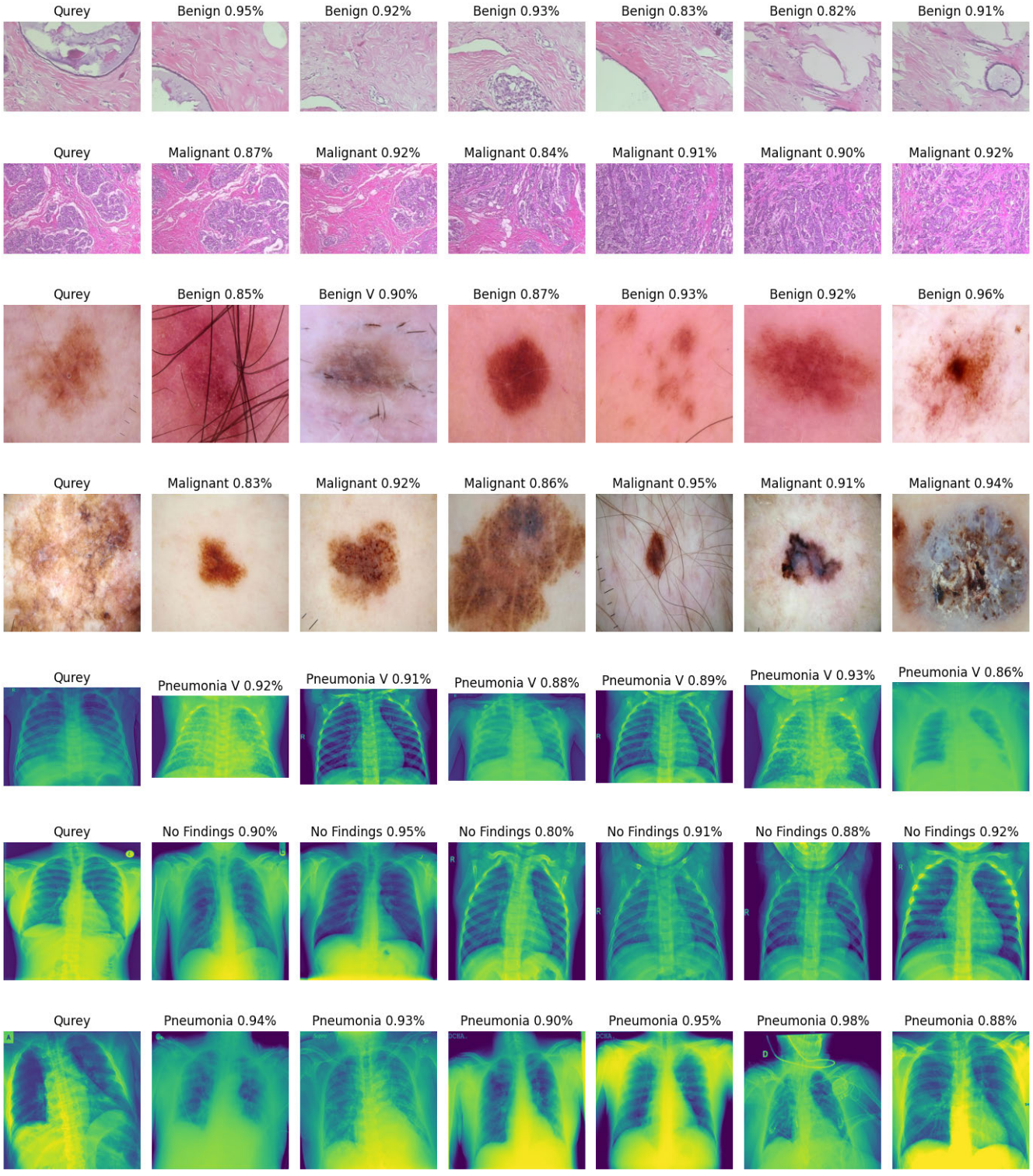

FIGURE 4. Retrieval results for the different medical image modalities: Query image is depicted in the first column and Retrieved images are depicted in remaining columns based on class prediction percentage.

Compare the performance of the classifiers in terms of these metrics to determine which one performs better on the given task. fig. 7 and fig. 9 depicts the findings of different classifiers and SVM gives us plausible results as compared to other classifiers such as KNN and Decision Tree. We explain the interpretability and explainability of the classifiers. Decision trees are inherently interpretable, as they can provide insights into the decision-making process.

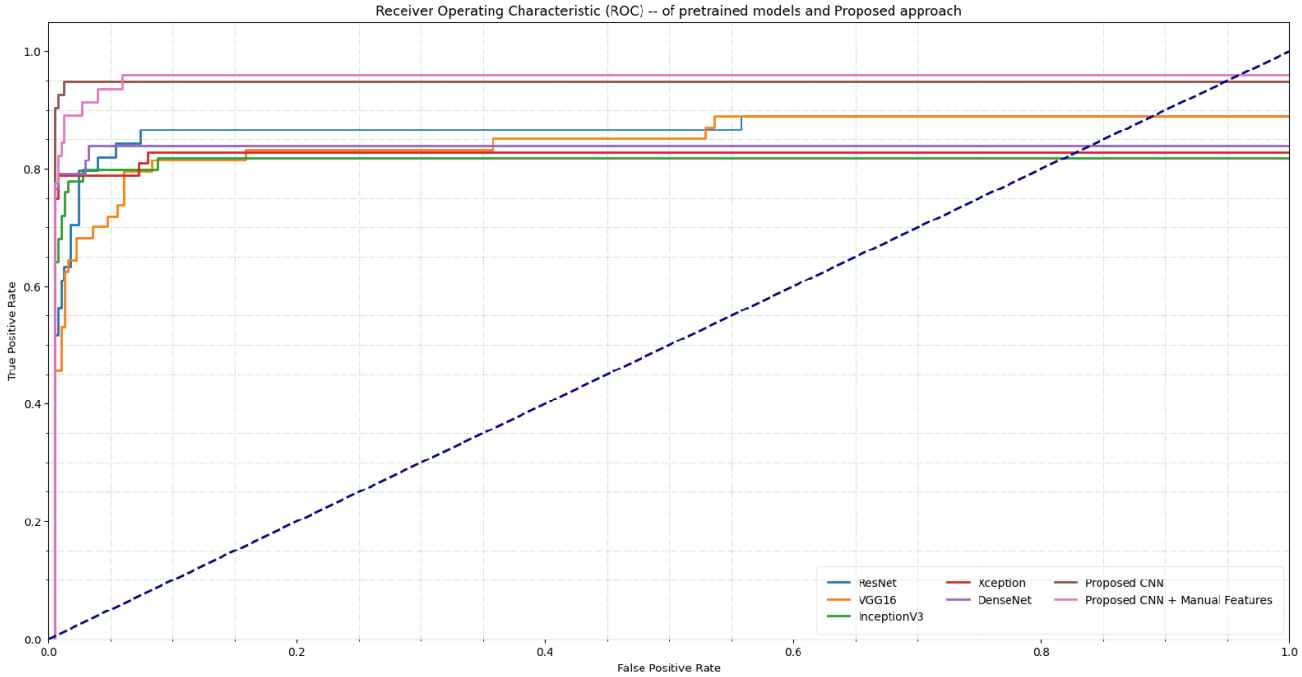

FIGURE 5. Receiver Operating Characteristic (ROC) curves of different pre-trained CNN models compared with our proposed model.

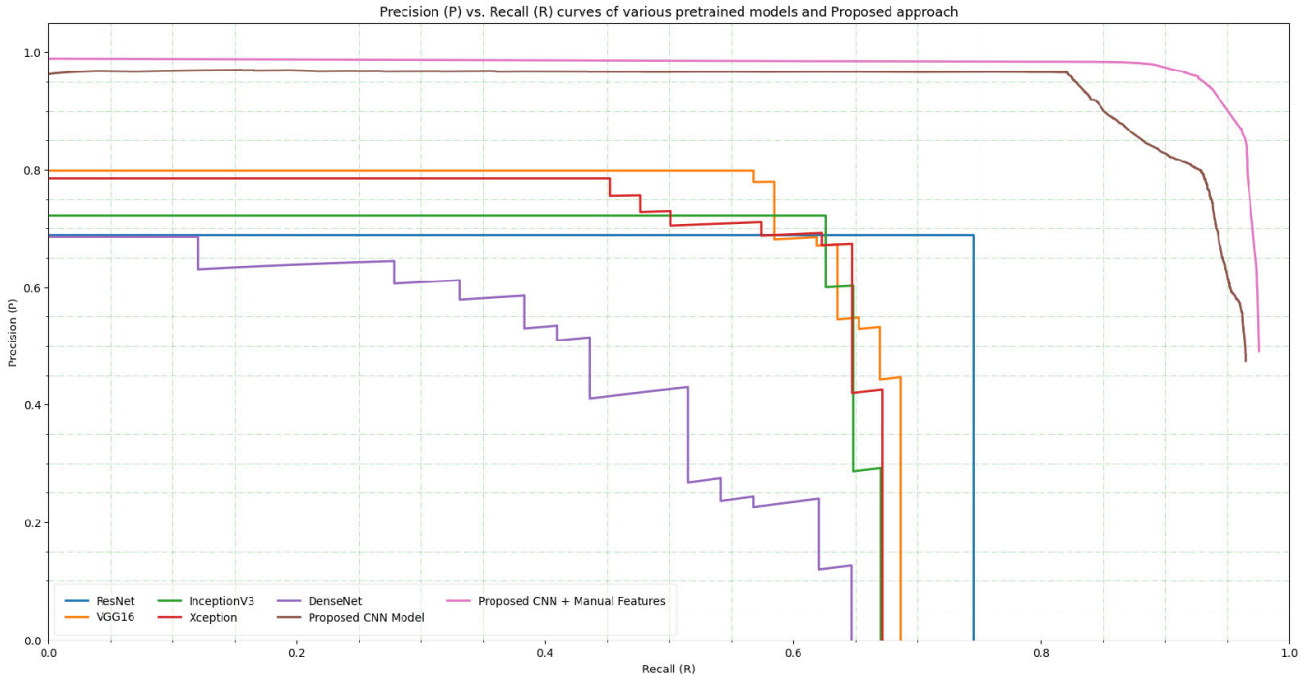

FIGURE 6. Precision and Recall comparison of different pre-trained CNN models with our proposed approach.

On the other hand, SVM and KNN may be less interpretable but can provide accurate predictions. We highlight the trade-off between interpretability and achievement based on the specific requirements of medical image retrieval. We ensure the robustness of the classifiers to noisy and outlier data points. Decision Trees, for instance, is sensitive to noise and may over-fit the training data. SVM, with its ability to find an optimal hyperplane, may handle noisy data better. Discuss the performance of each classifier in the presence of noise or outliers.

Three different runs were averaged to provide the stated results for the test set. fig. 6, which shows precision and recall at various rankings, respectively. The discounted cumulative gain is shown in fig. 8, the ROC-AUC is shown in fig. 5, and

TABLE 1. Training accuracy, F1-Score, Specificity, Precision and Sensitivity on different medical images modalities such as skin lesion, chest X-ray and Breast histopathological images using our proposed approach and different pretrained CNN models such as VGG, ResNet, Inception, Xception and DenseNet.

Model	Sensitivity	Specificity	Precision	Accuracy	F1-Score
ResNet	88.49	88.15	88.74	88.56	88.67
VGG16	86.34	85.97	86.18	86.23	86.75
InceptionV3	87.34	87.93	87.64	88.04	87.34
Xception	85.21	85.74	86.06	85.57	86.01
DenseNet	89.78	90.07	90.41	89.69	90.21
Proposed CNN	93.34	94.38	93.58	93.32	93.76
Proposed CNN+ Manual Features	96.31	95.84	95.89	96.08	95.74

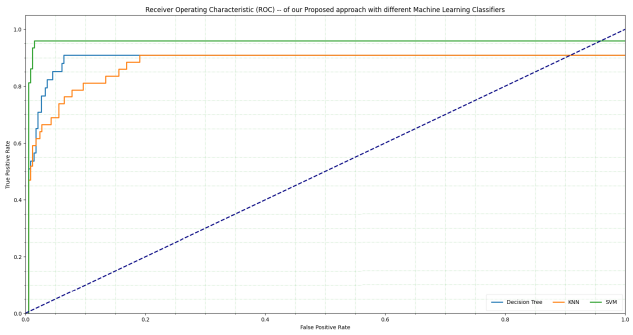

FIGURE 7. Receiver Operating Characteristic (ROC) curves of our Proposed approach with different Machine Learning Classifiers such as Decision Tree, KNN and SVM.

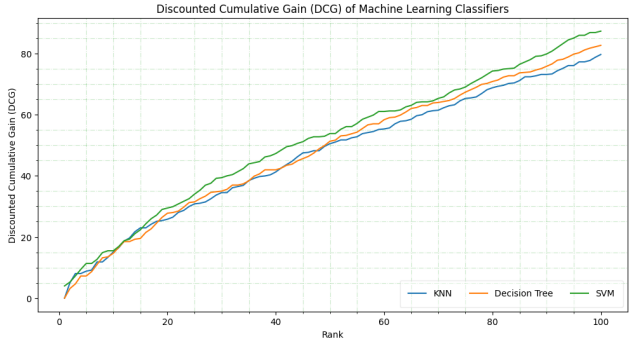

FIGURE 8. Discounted Cumulative Gain (DCG) of traditional Machine Learning Classifiers such as Decision Tree, KNN and SVM.

the MAP score is shown in fig. 9. With whiskers extending to points within 1.5 times the interquartile range of the lower and upper quartile, the boxplot displays the results' quartile values, including extreme values.

Using our suggested method, we discuss the findings and outcomes of our study comparing several pre-trained CNN models for CBMIR. The retrieved samples have a defined range and threshold, and any samples outside of this range are handled differently and independently. Several examples of images that were retrieved using various pre-trained CNN models and our suggested method are shown in fig. 4. It implies that the suggested method is applied to get pertinent images depending on the given query. The fig. 4 reported in this study demonstrates that the discovered architecture surpassed the various pre-trained CNN models in the experiment. This suggests that, when compared to

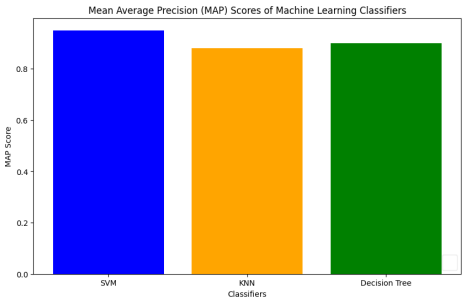

FIGURE 9. Mean Average Precision (MAP) scores of machine learning classifiers.

pre-trained models, our suggested strategy is more successful in retrieving pertinent images.

In this study, we have shown that obtaining diagnostically comparable images using Convolutional Neural Network-based CBMIR, a functional tool for radiologists, is possible. Not all CBMIR techniques, it is crucial to remember, are suitable for analyzing medical images. In this specific instance, the similarity of the recovered images must take into account both the lesions' visual qualities and their closeness to the F1 score. We compare our proposed approach, which outperforms various pre-trained CNN models by including visual appearance and textural features extracted via Zernike moments, Haralick, and GLPP features, to pre-trained CNN models, which traditionally emphasize visual appearance. fig. 4 shows how, by obtaining more pertinent image samples, the proposed approach beats the pre-trained CNN models. These two ways of learning are what set them apart from one another. While the pre-trained models may be able to infer this information indirectly from the medical images, the proposed approach training method expressly takes advantage of the diagnostic similarities. The CBMIR outcome is used to create a classifier for predicting the diagnosis of the pathology displayed in the image, the performance gap between the two approaches is further demonstrated. The diagnostic prediction ROC-AUC values are shown in fig. 5 and it demonstrates an increase from 0.52 to 0.69 of VGG, and then to 0.76. The outcomes show how our proposed approach may be expanded to include and integrate different manual features, which improves performance more than the single auto-extracted features. Notably, the ROC-AUC increased from 0.88 to 0.97 comparing the without integrating manual features using

only the chest X-ray and integrating manual features with our proposed approach.

In order to assess our proposed deep learning approach for medical image retrieval, we conducted a comparison with existing systems used for this task. However, a direct comparison was not feasible due to the unavailability of a standard medical dataset that could serve as a benchmark for the retrieval system. Therefore, we employed two criteria for the comparison. The first criterion involved classification accuracy, average precision, and average recall for classification, as presented in table 1. The second criterion was mean average precision (mAP) for retrieval, as shown in table 1. While our proposed approach achieved a higher mAP value, it is important to note that our approach focused solely on different modalities. In contrast, pre-trained CNN models are designed to handle multimodal data, which enhances their applicability and versatility.

VI. CONCLUSION

To improve Content-Based Medical Image Retrieval (CBMIR) architecture, we introduce a novel CNN model based on Fusion that combines manually created and automatically generated features. This adaptable framework anticipates and finds images that are diagnostically important across several medical modalities. It is quite versatile and works for a variety of diagnostic imaging activities. In conclusion, fusion-based CBMIR provides a workable method for combining data sources and foretelling comparable images in medical applications. By adding attributes from the classification outcomes of the retrieved images, integrating a Computer-Aided Diagnosis (CAD) system with CBMIR improves classifiers and facilitates AI interpretability. Future research should evaluate the clinical efficacy of CBMIR for radiologists through observer studies. The idea is also used to medical image segmentation, utilizing CBMIR findings to improve deep learning frameworks. The accuracy of picture segmentation can be improved by this fusion.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

DECLARATIONS

Conflict of interest No potential conflict of interest was reported by the authors.

REFERENCES

[1] C. Chen, M. Y. Lu, D. F. K. Williamson, T. Y. Chen, A. J. Schaumberg, and F. Mahmood, "Fast and scalable search of whole-slide images via self-supervised deep learning," *Nature Biomed. Eng.*, vol. 6, no. 12, pp. 1420–1434, Oct. 2022.

[2] P. Tschandl, C. Rinner, Z. Apalla, G. Argenziano, N. Codella, A. Halpern, M. Janda, A. Lallas, C. Longo, J. Malvehy, and J. Paoli, "Human–computer collaboration for skin cancer recognition," *Nature Med.*, vol. 26, no. 8, pp. 1229–1234, 2020.

[3] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Commun.*, vol. 5, no. 1, p. 4006, Jun. 2014.

[4] E. Pinho, J. Figueira Silva, and C. Costa, "Volumetric feature learning for query-by-example in medical imaging archives," in *Proc. IEEE 32nd Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2019, pp. 138–143.

[5] M. F. Sohan and A. Basalamah, "A systematic review on federated learning in medical image analysis," *IEEE Access*, vol. 11, pp. 28628–28644, 2023.

[6] Z. Li, X. Zhang, H. Müller, and S. Zhang, "Large-scale retrieval for medical image analytics: A comprehensive review," *Med. Image Anal.*, vol. 43, pp. 66–84, Jan. 2018.

[7] Z. Tabatabaei, A. Colomer, K. Engan, J. Oliver, and V. Naranjo, "Residual block convolutional auto encoder in content-based medical image retrieval," in *Proc. IEEE 14th Image, Video, Multidimensional Signal Process. Workshop (IVMSP)*, Jun. 2022, pp. 1–5.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[9] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. W. L. Aerts, "Artificial intelligence in radiology," *Nature Rev. Cancer*, vol. 18, no. 8, pp. 500–510, 2018.

[10] L. Zheng, Y. Yang, and Q. Tian, "SIFT meets CNN: A decade survey of instance retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 5, pp. 1224–1244, May 2018.

[11] F. Prior, K. Smith, A. Sharma, J. Kirby, L. Tarbox, K. Clark, W. Bennett, T. Nolan, and J. Freymann, "The public cancer radiology imaging collections of the cancer imaging archive," *Sci. Data*, vol. 4, no. 1, pp. 1–7, Sep. 2017.

[12] S. Kastryulin, J. Zakirov, N. Pezzotti, and D. V. Dylov, "Image quality assessment for magnetic resonance imaging," *IEEE Access*, vol. 11, pp. 14154–14168, 2023.

[13] E. Turro et al., "Whole-genome sequencing of patients with rare diseases in a national health system," *Nature*, vol. 583, no. 7814, pp. 96–102, 2020.

[14] S. Zhang and D. Metaxas, "Large-scale medical image analytics: Recent methodologies, applications and future directions," *Med. Image Anal.*, vol. 33, pp. 98–101, Oct. 2016.

[15] C. B. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, "Content-based image retrieval in radiology: Current status and future directions," *J. Digit. Imag.*, vol. 24, no. 2, pp. 208–222, Apr. 2011.

[16] H. Yaacob, F. Hossain, S. Shari, S. K. Khare, C. P. Ooi, and U. R. Acharya, "Application of artificial intelligence techniques for brain–computer interface in mental fatigue detection: A systematic review (2011–2022)," *IEEE Access*, vol. 11, pp. 74736–74758, 2023.

[17] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, and C. Roux, "Fast wavelet-based image characterization for highly adaptive image retrieval," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1613–1623, Apr. 2012.

[18] X. Zhang, W. Liu, M. Dundar, S. Badve, and S. Zhang, "Towards large-scale histopathological image analysis: Hashing-based image retrieval," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 496–506, Feb. 2015.

[19] R. Lan, S. Zhong, Z. Liu, Z. Shi, and X. Luo, "A simple texture feature for retrieval of medical images," *Multimedia Tools Appl.*, vol. 77, pp. 10853–10866, Jan. 2018.

[20] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[21] Y. Anavi, I. Kogan, E. Gelbart, O. Geva, and H. Greenspan, "A comparative study for chest radiograph image retrieval using binary texture and deep learning classification," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 2940–2943.

[22] A. Shah, S. Conjeti, N. Navab, and A. Katouzian, "Deeply learnt hashing forests for content based image retrieval in prostate MR images," *Proc. SPIE*, vol. 9784, pp. 302–307, Jan. 2016.

[23] X. Liu, H. R. Tizhoosh, and J. Kofman, "Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2016, pp. 2872–2878.

[24] S. Conjeti, A. G. Roy, A. Katouzian, and N. Navab, "Hashing with residual networks for image retrieval," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2017, pp. 541–549.

[25] Z. Chen, R. Cai, J. Lu, J. Feng, and J. Zhou, "Order-sensitive deep hashing for multimorbidity medical image retrieval," in *Proc. 21st Int. Conf. Med. Image Comput. Comput. Assist. Intervent.* Cham, Switzerland: Springer, 2018, pp. 620–628.

[26] K. K. Kumar and T. V. Gopal, "A novel approach to self order feature reweighting in CBIR to reduce semantic gap using relevance feedback," in *Proc. Int. Conf. Circuits, Power Comput. Technol.*, Mar. 2014, pp. 1437–1442.

[27] N. A. M. Zin, R. Yusof, S. A. Lashari, A. Mustapha, N. Senan, and R. Ibrahim, "Content-based image retrieval in medical domain: A review," *J. Phys., Conf. Ser.*, vol. 1019, Jan. 2018, Art. no. 012044.
[28] H. Müller and D. Unay, "Retrieval from and understanding of large-scale multi-modal medical datasets: A review," *IEEE Trans. Multimedia*, vol. 19, no. 9, pp. 2093–2104, Sep. 2017.
[29] S. M. Alizadeh, M. S. Helfroush, and H. Müller, "A novel Siamese deep hashing model for histopathology image retrieval," *Expert Syst. Appl.*, vol. 225, Sep. 2023, Art. no. 120169.
[30] X. Zhang, W. Liu, and S. Zhang, "Mining histopathological images via hashing-based scalable image retrieval," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 1111–1114.
[31] D. J. Foran, L. Yang, W. Chen, J. Hu, L. A. Goodell, M. Reiss, F. Wang, T. Kurc, T. Pan, A. Sharma, and J. H. Saltz, "ImageMiner: A software system for comparative analysis of tissue microarrays using content-based image retrieval, high-performance computing, and grid technology," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 4, pp. 403–415, Jul. 2011.
[32] A. Kumar, S. Dyer, J. Kim, C. Li, P. H. W. Leong, M. Fulham, and D. Feng, "Adapting content-based image retrieval techniques for the semantic annotation of medical images," *Computerized Med. Imag. Graph.*, vol. 49, pp. 37–45, Apr. 2016.
[33] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 157–166.
[34] A. Qayyum, S. M. Anwar, M. Awais, and M. Majid, "Medical image retrieval using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 8–20, Nov. 2017.
[35] Z. Çamlica, H. R. Tizhoosh, and F. Khalvati, "Autoencoding the retrieval relevance of medical images," in *Proc. Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Nov. 2015, pp. 550–555.
[36] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 18661–18673.
[37] P. Yang, Y. Zhai, L. Li, H. Lv, J. Wang, C. Zhu, and R. Jiang, "Liver histopathological image retrieval based on deep metric learning," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2019, pp. 914–919.
[38] P. Yang, Y. Zhai, L. Li, H. Lv, J. Wang, C. Zhu, and R. Jiang, "A deep metric learning approach for histopathological image retrieval," *Methods*, vol. 179, pp. 14–25, Jul. 2020.
[39] G. Wei, P. Huang, C. Xu, L. Chen, X. Ju, and X. Du, "Experimental study on the radiative properties of open-cell porous ceramics," *Sol. Energy*, vol. 149, pp. 13–19, Jun. 2017.
[40] J. Cheng, W. Yang, M. Huang, W. Huang, J. Jiang, Y. Zhou, R. Yang, J. Zhao, Y. Feng, Q. Feng, and W. Chen, "Retrieval of brain tumors by adaptive spatial pooling and Fisher vector representation," *PLoS ONE*, vol. 11, no. 6, Jun. 2016, Art. no. e0157112.
[41] X. Wang, Y. Du, S. Yang, J. Zhang, M. Wang, J. Zhang, W. Yang, J. Huang, and X. Han, "RetCCL: Clustering-guided contrastive learning for whole-slide image retrieval," *Med. Image Anal.*, vol. 83, Jan. 2023, Art. no. 102645.
[42] T. S. Reddy, K. Sanjeevaiah, S. Karthik, M. Kumar, and D. Vivek, "Content-based image retrieval using hybrid DenseNet121-BiLSTM and Harris hawks optimization algorithm," *Int. J. Softw. Innov.*, vol. 11, no. 1, pp. 1–15, Dec. 2022.
[43] G. Singh, V. Violi, and M. Fisichella, "Federated learning to safeguard patients data: A medical image retrieval case," *Big Data Cognit. Comput.*, vol. 7, no. 1, p. 18, Jan. 2023.
[44] L. Gao, Y. Wu, T. Yang, X. Zhang, Z. Zeng, C. K. D. Chan, and W. Chen, "Research on image classification and retrieval using deep learning with attention mechanism on diaspora Chinese architectural heritage in Jiangmen, China," *Buildings*, vol. 13, no. 2, p. 275, Jan. 2023.
[45] X. Zhang, C. Bai, and K. Kpalma, "OMCBIR: Offline mobile content-based image retrieval with lightweight CNN optimization," *Displays*, vol. 76, Jan. 2023, Art. no. 102355.
[46] N. Hashimoto, Y. Takagi, H. Masuda, H. Miyoshi, K. Kohno, M. Nagaishi, K. Sato, M. Takeuchi, T. Furuta, and K. Kawamoto, "Case-based similar image retrieval for weakly annotated large histopathological images of malignant lymphoma using deep metric learning," *Med. Image Anal.*, vol. 85, Apr. 2023, Art. no. 102752.
[47] S. M. Mahmoud, H. A. S. Al-Jubouri, and T. E. Abdoulabbas, "Chest radiographs images retrieval using deep learning networks," *Bull. Electr. Eng. Informat.*, vol. 11, no. 3, pp. 1358–1369, Jun. 2022.
[48] S. Agrawal, A. Chowdhary, S. Agarwala, V. Mayya, and S. Kamath, "Content-based medical image retrieval system for lung diseases using deep CNNs," *Int. J. Inf. Technol.*, vol. 2022, pp. 1–9, Jan. 2022.
[49] R. F. Mansour, "Multimodal biomedical image retrieval and indexing system using handcrafted with deep convolution neural network features," *J. Ambient Intell. Humanized Comput.*, vol. 10, pp. 1–10, May 2023.
[50] J. Pradhan, A. K. Pal, and H. Banka, "Medical image retrieval system using deep learning techniques," in *Deep Learning for Biomedical Data Analysis*. Switzerland: Springer, 2021, pp. 101–128.
[51] D. Raja and M. Karthikeyan, "Content based image retrieval using reptile search algorithm with deep learning for agricultural crops," in *Proc. 7th Int. Conf. Commun. Electron. Syst. (ICCES)*, Jun. 2022, pp. 1038–1043.
[52] N. F. Haq, M. Moradi, and Z. J. Wang, "A deep community based approach for large scale content based X-ray image retrieval," *Med. Image Anal.*, vol. 68, Feb. 2021, Art. no. 101847.
[53] L. Xu, X. Zeng, B. Zheng, and W. Li, "Multi-manifold deep discriminative cross-modal hashing for medical image retrieval," *IEEE Trans. Image Process.*, vol. 31, pp. 3371–3385, 2022.
[54] Y. Liu, J. Song, K. Zhou, L. Yan, L. Liu, F. Zou, and L. Shao, "Deep self-taught hashing for image retrieval," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2229–2241, Jun. 2019.
[55] C. Deng, E. Yang, T. Liu, J. Li, W. Liu, and D. Tao, "Unsupervised semantic-preserving adversarial hashing for image search," *IEEE Trans. Image Process.*, vol. 28, no. 8, pp. 4032–4044, Aug. 2019.
[56] Q. Feng, P. Li, Z. Lu, Z. Zhou, Y. Wu, J. Weng, and F. Huang, "DHAN: Encrypted JPEG image retrieval via DCT histograms-based attention networks," *Appl. Soft Comput.*, vol. 133, Jan. 2023, Art. no. 109935.
[57] D. Li, D. Dai, J. Chen, S. Xia, and G. Wang, "Ensemble learning framework for image retrieval via deep hash ranking," *Knowl.-Based Syst.*, vol. 260, Jan. 2023, Art. no. 110128.
[58] J. Fang, M. Zeng, X. Zhang, H. Liu, Y. Zhao, P. Zhang, H. Yang, J. Liu, H. Miao, Y. Hu, and J. Liu, "Deep metric learning with mirror attention and fine triplet loss for fundus image retrieval in ophthalmology," *Biomed. Signal Process. Control*, vol. 80, Feb. 2023, Art. no. 104277.
[59] G. W. Jiji and P. S. J. D. Raj, "Content-based image retrieval in dermatology using intelligent technique," *IET Image Process.*, vol. 9, no. 4, pp. 306–317, Apr. 2015.
[60] G. Sucharitha, N. Arora, and S. C. Sharma, "Medical image retrieval using a novel local relative directional edge pattern and Zernike moments," *Multimedia Tools Appl.*, vol. 82, pp. 1–21, May 2023.
[61] D. Patil, S. Krishnan, and S. Gharge, "Medical image retrieval by region based shape feature for CT images," in *Proc. Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon)*, Feb. 2019, pp. 155–159.
[62] A. A. Pandian and R. Balasubramanian, "Fusion of contourlet transform and Zernike moments using content based image retrieval for MRI brain tumor images," *Indian J. Sci. Technol.*, vol. 9, no. 29, pp. 1–8, Aug. 2016.
[63] F. S. Baji, S. B. Abdullah, and F. S. Abdulsattar, "K-mean clustering and local binary pattern techniques for automatic brain tumor detection," *Bull. Electr. Eng. Informat.*, vol. 12, no. 3, pp. 1586–1594, Jun. 2023.
[64] S. K. Sut, M. Koc, G. Zorlu, I. Serhatlioglu, P. D. Barua, S. Dogan, M. Baygin, T. Tuncer, R.-S. Tan, and U. R. Acharya, "Automated adrenal gland disease classes using patch-based center symmetric local binary pattern technique with ct images," *J. Digit. Imag.*, vol. 36, pp. 1–14, Jan. 2023.
[65] W. Ma, T. Zhou, J. Qin, X. Xiang, Y. Tan, and Z. Cai, "Adaptive multi-feature fusion via cross-entropy normalization for effective image retrieval," *Inf. Process. Manage.*, vol. 60, no. 1, Jan. 2023, Art. no. 103119.
[66] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3462–3471.
[67] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016.
[68] V. Rotemberg et al., "A patient-centric dataset of images and metadata for identifying melanomas using clinical context," *Sci. Data*, vol. 8, no. 1, p. 34, Jan. 2021.
[69] S. Iqbal and A. N. Qureshi, "A heteromorphous deep CNN framework for medical image segmentation using local binary pattern," *IEEE Access*, vol. 10, pp. 63466–63480, 2022.

SAEED IQBAL received the Ph.D. degree from the University of Central Punjab, Lahore. He is currently an accomplished computer science Researcher and an Assistant Professor with the University of Central Punjab. He has published numerous articles in reputable journals and presented his work at esteemed international conferences, making significant contributions to cutting-edge technologies and applications in the field. His research interests include medical image analysis, deep learning, computer vision, image processing, and information retrieval.

ADNAN N. QURESHI received the Ph.D. degree from the Institute for Research in Applicable Computing (IRAC), University of Bedfordshire, U.K. With more than 19 years of teaching experience in higher education institutions, both in Pakistan and U.K., along with developing software solutions, he has been a part of MyHealthAvatar and CHIC projects, funded by the EU, during his job with the Centre for Computer Graphics and Vision (CCGV), University of Bedfordshire, U.K. His research contributions include peer-reviewed conference and journal publications as an Independent Researcher and coauthor. His research interests include artificial intelligence, bio-medical image and signal processing, computer vision, machine learning, and autonomous systems.

MUSAED ALHUSSEIN received the B.S. degree in computer engineering from King Saud University (KSU), Riyadh, Saudi Arabia, in 1988, and the M.S. and Ph.D. degrees in computer science and engineering from the University of South Florida, Tampa, FL, USA, in1992 and 1997, respectively. Since 1997, he has been with the Faculty of the Computer Engineering Department, College of Computer and Information Science, KSU, Riyadh, Saudi Arabia. Recently, he has been successful in winning a research project in the area of AI for healthcare, which is funded by the Ministry of Education, Saudi Arabia. He is the Founder and the Director of the Embedded Computing and Signal Processing Research (ECASP) Laboratory. He is currently a Professor with the Department of Computer Engineering, College of Computer and Information Sciences, KSU. His research interests include typical computer architecture and signal processing topics with an emphasis on big data, machine/deep learning, VLSI testing and verification, embedded and pervasive computing, cyber-physical systems, mobile cloud computing, big data, eHealthcare, and body area networks.

IMRAN ARSHAD CHOUDHRY is currently an Assistant Professor and a Doctoral Researcher with the University of Central Punjab, Lahore, Pakistan. He has published a number of publications in reputable peer-reviewed journals throughout the world. His research interests include deep neural networks (DNN), convolutional neural networks (CNN), computer vision, and image processing.

KHURSHEED AURANGZEB (Senior Member, IEEE) received the B.S. degree in computer engineering from the COMSATS Institute of Information Technology Abbottabad, Pakistan, in 2006, the M.S. degree in electrical engineering (system on chip design) from Linköping University, Sweden, in 2009, and the Ph.D. degree in electronics design from Mid Sweden University, Sweden, in June 2013. He is currently an Associate Professor with the Department of Computer Engineering, College of Computer and Information Sciences, King Saud University (KSU), Riyadh, Saudi Arabia. He has obtained more than 15 years of excellent experience, as a Instructor and Researcher of data analytics, machine/deep learning, signal processing, electronics circuits/systems, and embedded systems. He has authored and coauthored more than 90 publications including IEEE/ACM/Springer/Hindawi/MDPI journals, and flagship conference papers. He has been involved in many research projects, as a Principal Investigator and a Co-Principal Investigator. His research interests include the diverse fields of embedded systems, computer architecture, signal processing, wireless sensor networks, communication, camera-based sensor networks with an emphasis on big data and machine/deep learning with applications in smart grids, precision agriculture, and healthcare.

TARIQ M. KHAN (Member, IEEE) received the Ph.D. degree from Macquarie University. He is currently a Senior Research Associate with the School of Computer Science and Engineering (CSE), UNSW Sydney, Australia. Prior to joining UNSW Sydney, in 2021, he studied and worked at various universities, including COMSATS University Islamabad Pakistan, as an Assistant Professor; and Deakin University, as a Research Fellow. He has published more than 80 peer-reviewed journal and conference papers. His research interests include computer vision and machine learning, encompassing image segmentation, image classification, medical image analysis, resource-constrained neural networks, and deep neural networks. He is currently working on developing new computer vision and machine learning methods, particularly deep learning, for automated quantitative analysis of biomedical imaging data, and other industrial applications. According to SciVal data from the last five years, he is ranked number 1 in the world for research in the topics of retina image, retina blood vessels, and hypertension retinopathy.