RESEARCH ARTICLE

Toward More Transparent and Accurate Cancer Diagnosis With an Unsupervised CAE Approach

ZAHRA TABATABAEI[1,2], **ADRIÁN COLOMER**[2], **JAVIER OLIVER MOLL**[1],
AND VALERY NARANJO[2], **(Member, IEEE)**

[1]Department of Artificial Intelligence, Tyris Tech S.L., 46021 Valencia, Spain
[2]Instituto Universitario de Investigación en Tecnología Centrada en el Ser Humano, HUMAN-Tech, Universitat Politècnica de València, 46022 Valencia, Spain

Corresponding author: Zahra Tabatabaei (elec.tabatabaei@gmail.com)

ABSTRACT According to the Global Cancer Observatory, 2020, breast cancer is the most prevalent cancer type in both genders (11.7%), while prostate cancer is the second most common cancer type in men (14.1%). In digital pathology, Content-Based Medical Image Retrieval (CBMIR) is a powerful tool for improving cancer diagnosis by searching for similar histopathological Whole Slide Images (WSIs). CBMIR empowers pathologists to explore similar patches to their query, enhancing diagnostic reliability and accuracy. In this paper, a customized unsupervised Convolutional Auto Encoder (CAE) was developed in the proposed Unsupervised CBMIR (UCBMIR) to replicate the traditional cancer diagnosis workflow, offering the potential to enhance diagnostic accuracy and efficiency by reducing pathologists' workload. Furthermore, it provides a more transparent supporting tool for pathologists in cancer diagnosis. UCBMIR was evaluated using two widely used numerical techniques in CBMIR, visual techniques, and compared with a classifier. Validation encompassed three data sets, including an external evaluation to demonstrate its effectiveness. UCBMIR achieved 99% and 80% top 5 recalls on BreaKHis and SICAPv2 with the first evaluation technique while using the second technique, it reached 91% and 70% precision for BreaKHis and SICAPv2, respectively. Moreover, UCBMIR displayed a strong capability to identify diverse patterns, yielding 81% accuracy in the top 5 predictions on an external image from Arvaniti. The proposed unsupervised CBMIR tool delivered 83% accuracy in retrieving images with the same cancer type.

I. INTRODUCTION

Cancer is a leading cause of death worldwide, with nearly 10 million deaths reported in 2020, as per the World Health Organization (WHO) [1]. In 2020, breast and prostate cancer affected 2.26 million and 1.41 million cases, respectively. Accurate cancer diagnosis is critical for effective treatment because each cancer type requires a specific treatment regimen. However, diagnostic errors are prevalent, affecting approximately 5.08% of cases, which translates to around 12 million adults in the United States [2]. This significant

percentage of human error in a large number of cancer cases poses significant drawbacks for society and the quality of human lives.

Moreover, there is a significant disparity in treatment availability between countries with varying income levels. In high-income countries, comprehensive treatment is available in over 90% of cases, but this figure drops to less than 15% in low-income countries [3]. In this context, there is an urgent need to develop reliable and accurate diagnostic tools that can assist medical professionals in making accurate and timely diagnoses, regardless of their location or income level.

Computer-Aided Diagnosis (CAD) models play a vital role in reducing the incidence of human errors and providing an

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

inclusive worldwide platform for individuals with varying incomes. CAD offers multiple approaches under the umbrella of "digital pathology" to enhance conventional cancer diagnosis. Digital pathology has garnered significant attention due to its ability to provide a definitive diagnosis at the pathology level, taking into account factors such as size, complexity, and color [4]. The challenges and opportunities presented in digital pathology are explained in [5]. Despite the challenges, digital pathology can serve as a bridge toward the discovery of histopathological imaging and enable more accurate prognostic predictions for disease aggressiveness and patient outcomes.

The following subsections cover a brief literature review of digital pathology on WSIs.

A. SEGMENTATION

Automatic detection of irrelevant regions of tissue may bring a more reliable prediction. In [6], they proposed a multi-scale model to detect invasive cancerous area patterns in WSIs of bladder cancer. Similarly, [7] focuses on detecting blood and damaged tissue as problematic artifacts in bladder tumors. In [8], the authors apply the DenseRes-Unet model to multi-organ histopathological images to segment overlapped/clustered nuclei. A binary threshold is set to detect the contour of the extracted nuclei in the images, as the morphological characteristics of the cells are critical to grading the cancers. Moreover, the two-stage nuclei segmentation strategy proposed in [9] based on watershed segmentation is used to distinguish between carcinoma and non-carcinoma recognition in the Bio-imaging 2015 data set. Additionally, [10] introduced a novel approach to detect nuclei in breast cancer histopathological images using a stacked sparse Auto Encoder (AE).

Segmentation techniques have been studied extensively to quantify cell nucleus form and dispersion, which may improve accuracy in classification and grading [11]. However, these methods do not offer direct benefits to pathologists. Though Deep Learning (DL) has shown promise in improving segmentation, it relies heavily on large annotated data sets [12], limiting its impact [13]. Innovative approaches are needed to develop new techniques that can benefit pathologists and improve disease diagnosis.

B. CLASSIFICATION

Classification of input images is a critical task in medical image analysis, where an optimal classifier is expected to provide accurate labels for each input [14]. This can significantly aid pathologists in their daily analysis of tissue grading. For instance, in the realm of cancer diagnosis, saliency maps computation has improved the diagnostic process, both in radiological [15] and histopathological [16] images.

Reference [17] validated an end-to-end pixel-level prediction of Gleason grades and scored the entire biopsy. Similarly, [18] proposed an AE using Siamese network aimed at learning image features by minimizing the distance between input and output. Another approach was proposed by [19], who aimed to decrease the rate of diagnostic errors by performing patch-based transfer learning. However, patch-level data sets extracted from Whole Slide Images (WSIs) often contain mislabeled patches, which may lead the classifiers to miss important information. To address this problem, [20] proposed DenseNet121-AnoGAN, which employs unsupervised anomaly detection with generative adversarial networks (AnoGAN) to prevent missing mislabeled patches. This approach has been successfully applied to classify the BreaKHis data set into benign and malignant.

The author in [21] fed an Inception Recurrent Residual Convolutional Neural Network (IRRCNN) model with two breast cancer data sets to have binary and multi-class classifiers. The authors in [22] conducted their experiments on DenseNet with SENet IDSNet and BreaKHis data set. They fine-tuned DenseNet-121 to propose an accurate classifier. A deep Feature Extractor (FE) from a pre-trained network and a classifier are used in [23] to classify BreaKHis. 16 pre-trained networks and 7 classifiers were tested in this paper.

In brief, classifier architectures have been proposed for use in diagnostic pathology to aid pathologists in making more accurate cancer diagnoses. Many studies have reported high classification accuracy, which has been validated in engineering laboratories. However, despite their potential importance, these measurements have not yet led to a significant change in diagnostic imaging.

While classification and segmentation have proven to be valuable tools, they have not drastically transformed the diagnostic process. This may be attributed to their inability to reduce ambiguity and boost the confidence of pathologists in their diagnoses. In essence, these methods do not provide any additional information to aid pathologists in their report writing during the diagnostic process. For instance, CAMs and saliency maps provide clinicians with information exclusively about the case under study. These are visual information from the input image that CNNs are paying attention to perform classification. While CBMIR goes deeper in providing transparency and reliability to pathologists. It is retrieving similar cases from the previously diagnosed cases. This approach is similar to the pipeline followed by pathologists who consult with reference books in pathology, such as Atlas. So, CBHIR is a digital intelligent Atlas book that can speed up the search process and be more accurate.

In regards to developing methods for specific applications, it is possible to achieve higher results for the intended objective. However, creating and implementing unique methods for each potential task of interest is impractical. As an alternative approach, CBMIR has established a reliable framework for quality control. While it may have poorer accuracy than an application-specific instrument, having a multipurpose general-purpose tool like CBMIR can still be useful.

C. CONTENT-BASED MEDICAL IMAGE RETRIEVAL (CBMIR)

Automated medical imaging has been growing dramatically to improve clinical treatment and intervention in medical diagnosis. This yields an exigent demand for developing highly effective CAD systems. CBMIR is an active area of research with significant applications in routine clinical diagnostic aid, medical education, and research. In CBMIR, the end user targets retrieving the most relevant images. So, pathologists will trust this outcome easier because not only will they have a second opinion on their tissue (label), but they can also look for the same patterns in the previous tissues. In the classification task, they can get a label for their new tissue without knowing the reason. But in CBMIR, they can see the similarity between their tissues and the retrieved ones. Moreover, the explainable nature of CBMIR allows clinicians to understand how the system arrived at a particular diagnosis or recommendation, promoting transparency and trust in AI-assisted medical decision-making. Most notably, CBMIR is pathologist-centric; in contrast to classification, it is essentially an attempt to make decisions on behalf of the pathologists.

In DL, similar patterns mean similar features and representations. Humans can properly describe and interpret image contents, while digital machines can provide fewer semantic words for the same image. Machines provide a numerical description of the images with a wide gap compared to the human interpretation of the same image. This gap is named "*semantic gap*," and this broadly limits the performance of retrieval tasks [24]. The semantic gap is the main reason CBMIR has not made it into the daily laboratories workflow, yet. Indeed, this is arguably the paramount challenge in adopting CBMIR into the laboratories' workflow. Pathologists face numerous challenges in the current diagnostic paradigm, with time being a common factor. However, the impact of these challenges extends beyond just medical professionals and patients; it can also affect society as a whole. This can lead to emotional distress and other adverse effects on the well-being of patients and their families. Digital pathology, through the use of CBMIR, can mitigate the impact of these changes and enhance the accuracy of diagnoses.

CBMIR in virtual telepathology offers a reliable framework for achieving quality control through computational consensus-building, ensuring that diagnoses are accurate and consistent across different pathologists and healthcare institutions. By utilizing a vast database of reference images and advanced algorithms, CBMIR enhances the accuracy of diagnoses, potentially decreasing the need for additional studies and speeding up the diagnostic process. This can lead to better patient outcomes and a more efficient healthcare system. In recent years, CBMIR has gone through a renaissance with the promise of revolution. In a previous study [25], a CNN-based AE was applied to the BreaKHis data set with the aim of minimizing misinformation and evaluating the performance of CBMIR in a binary scenario. However, the reconstructed images produced by this method were found to be blurry, indicating that the extracted features by the AE were not robust enough to reconstruct the original image. In addition, the scope of this study was limited to detecting breast cancer using a two-class data set, without considering other diseases. These limitations highlight the need for further research to improve the quality of feature extraction in CBMIR systems. In [26], the CBMIR performance was improved in a supervised manner using a Hybrid feature-based ICNN model. The model was trained by adding three Fully Connected (FC) layers to accommodate the classification of cancer subtypes from TCGA. The researchers in [27] aimed to replicate the process of detecting morphological features used by pathologists in cancer diagnosis by incorporating different magnification levels into their CBMIR system. Specifically, they trained their system using a subset of TCGA data set in three magnification levels: $20\times$, $10\times$, and $5\times$. To address the differences in features that might exist at these different magnification levels, the last DenseNet-121 block [28] was re-trained using $10\times$ and $5\times$ magnification patches. This supervised approach improved the adaptability of the FE and resulted in better overall performance of the CBMIR system. KimiaNet reported two types of image search: horizontal search and vertical search. In the horizontal search, the query is applied to the entire data set to find similar whole slide images (WSI) with a self-supervised model [29], while in the study by Fashi et al. [30], the vertical search approach is designed to identify similar types of malignancies in a specific organ. This is achieved by utilizing pre-trained models with openly provided weights from the Keras library. The problem with supervised CBMIR is that it requires a large amount of labeled data, which can be time-consuming and costly to obtain. On the other hand, the problem with self-supervised CBMIR is that it may not perform as well as supervised methods and may require more complex models. Indeed, many researches [31], [32], [33], [34], [35], [36], [37], [38], [39], [40] have been dedicated to CBMIR, but the overall performance of the existing systems is not high enough due to the growing medical images and digital pathology.

The main contributions of this paper in proposing an Unsupervised CBMIR (UCBMIR) are:

- Proposing a new unsupervised approach for prostate and breast cancer gradation problem using CBMIR that achieves performance comparable to fully supervised methods.
- Extensively validating the proposed UCBMIR approach on three databases, including BreaKHis for a binary scenario and SICAPv2, which is the largest pixel-wise annotated prostate data set, and Arvaniti for multi-class grading problem, which is more challenging.
- Conducting an external evaluation to demonstrate the performance and generalization of UCBMIR, by training the model on SICAPv2 and testing it on the Arvaniti data set.

- A comprehensive evaluation of UCBMIR is presented, encompassing various numerical and visual performance metrics.

In addition, the paper addresses two major problems in traditional cancer diagnosis: inexperienced pathologists requiring more ancillary studies for diagnosis and the time-consuming process of differentiating between cancer grades. The UCBMIR model proposed in the paper provides a vast database of images that pathologists can use as a reference for diagnosis, allowing them to make more accurate diagnoses even if they are inexperienced. Additionally, the proposed tool enables pathologists to access annotated image databases instantly, leading to a faster diagnosis and skipping time-consuming reading and searching processes in "*Expert-Path*" and "*PathologyOutlines*" or an Atlas book. In order to reach the primary objective of this study we introduce an unsupervised image search tool for pathologists to facilitate the efficient retrieval of similar images from previous cases. The initial stage involves training a customized CAE that includes a skip layer between the encoder and the decoder, as well as an attention block in the bottleneck. This CAE is trained to reconstruct images and learn effective data representations while simultaneously ignoring the noise. The encoder with the bottleneck of the trained CAE serves as our FE in the search stage. We represent the complete training set of the data set as in previous cases and carry out patch-by-patch retrieval to obtain diagnosis-relevant patches for each query in the test set. Our ranking algorithm, which utilizes Euclidean distance, identifies the retrieved patches, which are then presented to the pathologists as the output of the proposed UCBMIR. Our study showcases the practicality of our approach in enhancing the efficiency and accuracy of image retrieval for pathologists and engineers. As a result, our method can accelerate cancer diagnosis for pathologists, and the deep layers in the custom-built CAE can learn image features in an unsupervised manner, circumventing the issue of insufficient training images.

II. MATERIAL

The study evaluates the performance of the UCBMIR on two of the largest labeled histopathological images in breast and prostate cancer, namely BreaKHis and SICAPv2, respectively. The Arvaniti data set is utilized as the third and an external data set in order to validate the model performance. These two cancers, prostate and breast cancer, are selected as they are prevalent in society.

BreaKHis: breast tissue biopsy slides were stained with Hematoxylin and Eosin (H & E) and labeled by pathologists at the P&D medical laboratory in Brazil [41]. This data set is composed of 7909 microscopic images of breast tumor tissues collected from patients using magnifying factors of $40\times$, $100\times$, $200\times$, and $400\times$ in the size of $224 \times 224 \times 3$. This binary data set contains 588 benign and 1232 malignant images in $400\times$.

SICAPv2: prostate samples were sliced, stained in H &E, and digitized at $40\times$ magnification. Images were divided into $512 \times 512 \times 3$ and down-sampled to $10\times$, which is commonly used for evaluating images. This multi-class data set contains 155 WSIs in total: 4417 non-cancerous patches, of which 1635 are labeled as Grade 3 (G3), 3622 as Grade 4 (G4), and 665 as Grade 5 (G5), Table 1. Images labeled by a group of expert urogenital pathologists at Hospital Clínico of Valencia. SICAPv2 is the largest publicly available data set that includes pixel-level annotations of Gleason grading, providing detailed information on the presence of cribriform patterns [17].

TABLE 1. SICAPv2 data set description.

Grades	NC	G3	G4	G5
WSIs	37	60	69	16
Patches	4417	1636	3622	655

In order to validate the generalization capability of the UCBMIR to find similar images, Arvaniti, an external data set containing pixel-level annotations of Gleason grades, is used.

Arvaniti: the data set was shared by Arvaniti et al. [42], which contains 625 patches of prostate histology images at $40\times$ magnification. Regarding a fair comparison with SICAPv2, in [17], some configurations were applied to re-sample images to 512×512 at $10\times$ magnification. To normalize the color distribution of Arvaniti, the author in [17] applied a histogram match to the re-sampled images and set the images in SICAP and Panda [43] as the reference images. These re-sampled images are used as the third data set and the external evaluations in this paper. Arvaniti is employed for performance evaluation alongside normalization by both Panda and SICAP, in addition to the external evaluation. This is discussed in detail in the following sections of this journal paper.

III. METHODOLOGY

A CBMIR contains four subsections: 1. training, 2. indexing and saving, 3. searching, and 4. evaluating. The search tool in CBMIR uses the contents within each pixel of the images instead of using annotations or metadata. Consequently, similar images are retrieved from a large data set that matches the contents of the queried image. Also, it is often impractical to manually annotate images in a large data set, thus an unsupervised FE is developed in this study to address this issue.

The four phases of the proposed UCBMIR are described in-depth in the following subsections with Figure 2 and Figure 1, in accordance with SICAPv2, as it is a complex multi-class data set.

A. TRAINING

The only training part of the proposed UCBMIR is training the proposed CAE. CAE aims to reconstruct the output as equal to the input. CAE could learn effective features with unlabeled data in an unsupervised manner. Using a CAE approach offers several benefits, such as its ability to capture spatial information through convolutional layers,
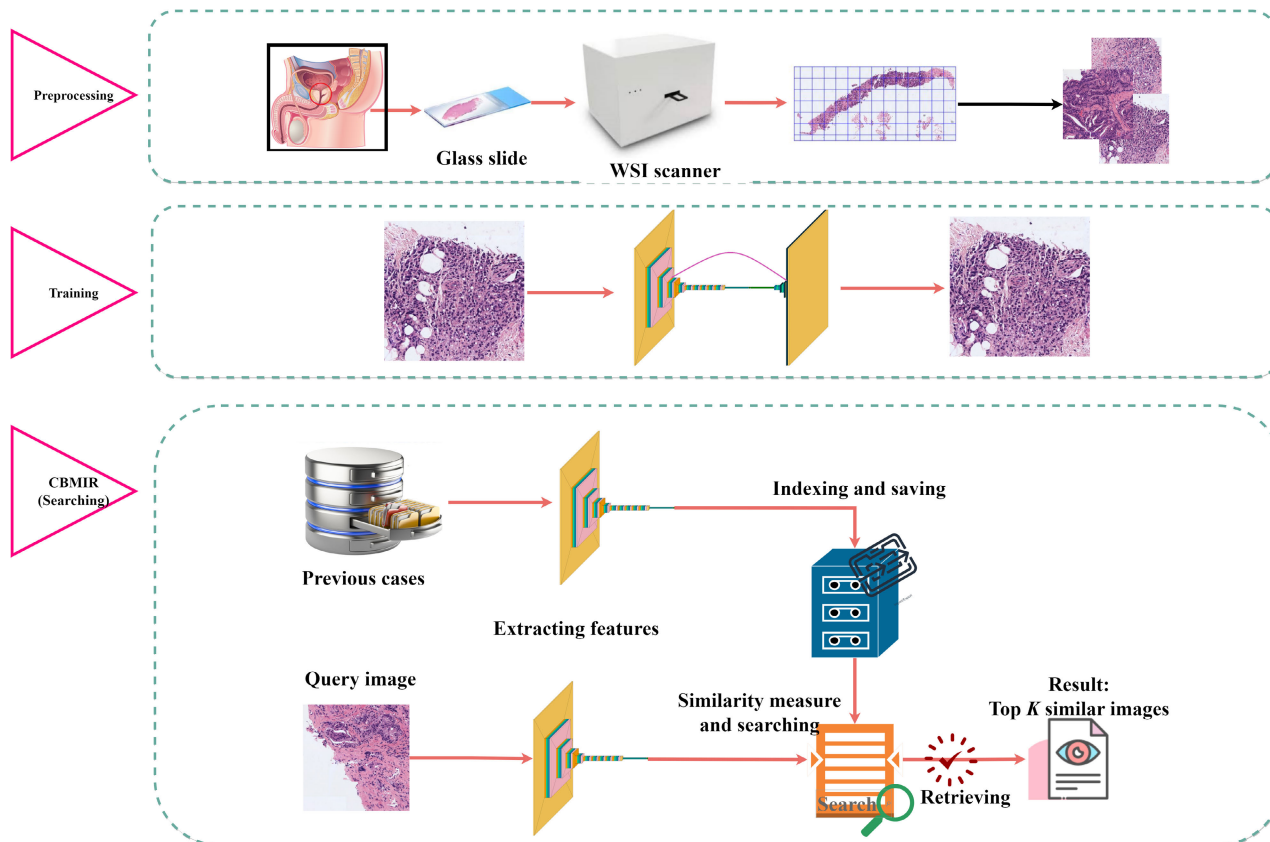
FIGURE 1. An overview of the UCBMIR for retrieving similar cases to a given query. The preprocessing stage involves extracting tissue from the patient's body and dividing the whole slide images (WSIs) into patches of a specific size [17]. In the training stage, these patches are used to train the proposed unsupervised CAE to extract feature representations. The trained encoder and bottleneck layers are then used to extract feature embeddings FEs that are used in the CBMIR section. In the CBMIR stage, the search engine computes the embedding features of the training set and stores them in a dictionary. When a query image is selected from the test set, the FE computes the embedding of that query and compares it with those in the dictionary. The model then returns the K most similar patches based on the pathologists' needs.

which are well-suited for processing image data. Moreover, CAE employs multiple layers to extract advanced image representations, resulting in higher-level feature recognition. These multi-layered models have fewer free parameters, making them simpler and faster to train, reducing the cost and resources required for training. Figure 2 exhibits an illustration of the proposed CAE architecture. It contains three main parts:

- **Encoder**: it captures the structural attributes of the input images across a feature vector per image with 200 elements. The sizes of the convolutional filters in the encoder are [16, 32, 64, 128, 256]. Histopathological images are highly detailed; using operations such as pooling layers will cause them to lose lots of information. Because of that, in the proposed CAE, the size of the images decreases by passing through convolutional layers without any pooling layer.
- **Bottleneck**: it contains 200 extracted features per image. As can be seen in Figure 2, a residual block in the bottleneck contains four filters in the size of [64, 32, 1, 256].
- **Decoder**: it reconstructs the input from its 200 intermediate feature vectors. Consequently, to the encoder,
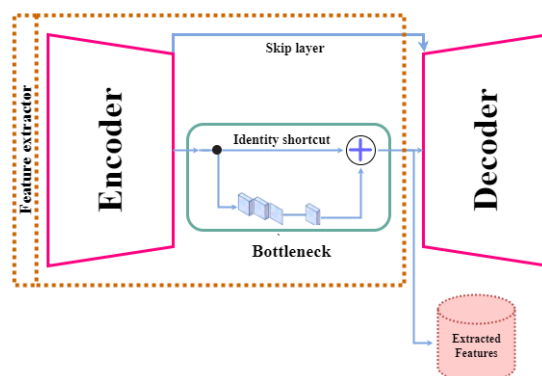

FIGURE 2. Proposed CAE architecture with kernel size of 3 throughout the model, the stride of 2 in the encoder and decoder, and 1 in the bottleneck layer.

the filters in the decoder part are in the size of [128, 64, 32, 16, 3].

The main objective of the proposed CAE is to find the most discriminative feature vectors to describe the images without supervision. Briefly, it compresses input image patches (of dimensions width $\times$ height $\times$ channels) into a fixed-length vector. The performance of the FE is directly related to the

depth of the learning model, as deeper and more complex models might result in overfitting. To address this issue, the proposed CAE employs a residual block in the bottleneck to increase the network depth and improve end-to-end mapping.

To get better performance gain, inspired by highway networks [44] and deep residual networks [45], we add a skip connection between two corresponding convolutional and deconvolutional layers. When the network goes deeper, image details can be lost, making deconvolution weaker at recovering the input. Skip connections benefit by back-propagating the gradient to the bottom layers, making training a deeper network much more accessible. In other words, skip connections pass gradients backward, which helps find a better local minimum.

Mean Square Error (MSE) is the loss function that updates the network weights during the training phase. The minimum amount of MSE means more similarity between the input and the output. The greater the similarity between the input and the reconstructed images, the more meaningful features are in the bottleneck. In practice, we find that using Adam with a learning rate $5 \times 10^{(-5)}$ can train the model in 10 epochs. Then, the decoder part of the trained CAE is discarded, and the remaining sections, including the encoder and the bottleneck, play a role as an FE.

B. INDEXING AND SAVING

The indexing and saving stage is a crucial step in CBMIR as it enables efficient storage and organization of extracted features. This, in turn, enables fast and accurate retrieval of relevant medical images during the search stage, thus improving the diagnosis and treatment of medical conditions. In our study, we used n images from both the validation and train sets of each data set as input to the FE, resulting in n feature vectors that represent each image in a $200-$dimensional latent space. These n feature vectors were then stored in a dictionary, $D_i = [F_1, F_2, \ldots, F_n]$, where each F_i contains the features for a single image.

During the retrieval stage, we utilized this dictionary as a reference for comparison with the query image. For the SICAPv2 data set, there were 2122 query images. By organizing and storing the extracted features in this manner, we aimed to improve the efficiency and accuracy of our CBMIR system. This approach enabled us to retrieve medical images that were relevant to a given query, thereby aiding in faster and more accurate diagnosis and treatment of medical conditions. Figure 1 illustrates the process of organizing and storing the extracted features for efficient retrieval of medical images during the search stage.

The extracted feature vectors of the database and the index of the related image were stored in a pickle file with the columns of ("*indexes''* : *indexes*, "*features''* : *features* by importing the pickle library in Python. Since extracting, indexing, and storing need high computational power,

we implemented the work on GPU with the *NVIDIA GeForce RTX 3090*.

C. SEARCHING

The searching process in CBMIR involves three key steps: similarity calculation, ranking and retrieval, and visualization and presentation. During similarity calculation, the search engine uses similarity measures such as Euclidean distance, cosine, Manhattan, and Haversine to calculate the similarity between the query image and other images in the database. The images in the database are then ranked based on their similarity to the query image, and the top-ranked images are retrieved and presented to the user for further analysis.

In this paper, we experimented with both Cosine and Euclidean distances, and based on our results, we concluded that the Euclidean distance was the more suitable choice. We use Euclidean Distance to measure the similarity of two feature vectors. Specifically, we calculate the distance by each query feature F_Q with all the feature vectors in D_i, and the smaller Euclidean value corresponds to more similar images. Our experimental findings suggest that Euclidean Distance is an effective metric for measuring similarity in CBMIR systems. By accurately measuring the similarity between images, the search engine can more effectively retrieve relevant medical images, leading to improved diagnosis and treatment outcomes.

D. EVALUATION

It is worth considering what "accuracy" means in the context of a CBMIR. The accuracy of CBMIR depends on what we are looking for and what is displayed by the search engine. The use case determines whether the search is looking for images with the same stain, comparable stain intensity, same histologic feature, or similar grade; hence, this objective is ambiguous. To address this lack of awareness of the intent of the search engine, top K score at retrieving images of the same histologic features and Gleason grades engaged in the prior research to determine the performance of their experiments. To the best of the author's knowledge, there are two most-used strategies for calculating the top K score described in the recent articles:

1) If there is only one correct retrieved image, this has been shown as a correct answer [46]. In this paper, we set $K = 3, 5, 7$, which evaluates the performance of our model to correctly present at least one correct result in the top K retrieved images. In this paper, we name this method as "EV 1" regarding the report of the results in the following tables.

$$ACC@K = \frac{1}{N} \sum_{i}^{N} \varepsilon(\alpha_i, TOP(ans[:K])) \qquad (1)$$

In this equation, N denotes the number of query patches, and α_i represents the label of the i-th query patch. The function $TOP(ans_i[:K])$ retrieves the top k most similar results for the query and outputs 1 if any of these results

match with the query and 0 otherwise. In other words, if $TOP(ans_i[:K])$ belongs to the set of labels of the i-th query, denoted by α_i, the function $\varepsilon()$ returns 1.

2) Precision (2) and recall (3) are the two selected indicators to evaluate the results. In this study, this is termed "EV 2".

$$Precision = \frac{R_v}{n} \qquad (2)$$

$$Recall = \frac{R_v}{M} \qquad (3)$$

The relevancy of the query and retrieved images has to be measured by considering the provided labels for each patch in the ground truth. Herein, R_v denotes the set of retrieved images that are considered relevant, while n signifies the total number of captured images. Moreover, the number of relevant images present in the data set is explicitly annotated as M. The proposed UCBMIR is evaluated based on both top-ranking image retrieval strategies to mimic the standard search process.

It is worth noting that this is the only place where the grand truth provided by expert pathologists was used. In other words, the labels were exclusively used to assess how effectively the model could retrieve images with similar histopathological patterns.

IV. DISCUSSION AND RESULTS

Matching pairs to the image is the core of any search engine, in which an image is compared to a database to determine similarities. Numerous studies have been conducted on CBMIR in a binary manner, as it is more challenging with multi-class data sets.

Breast cancer is a prevalent malignancy affecting women globally. In the domain of CAD, the BreaKHis data set is a popular choice for evaluating the performance of algorithms in CBMIR. In this study, we employed the BreaKHis to assess the efficacy of our UCBMIR approach. Our method demonstrated superior performance in matching image pairs, as evidenced by the results presented in Table 2. Specifically, our approach outperformed two previously reported methods, namely [25] and [47], with precision scores of 92% and 91%, respectively, for both evaluation criteria (EV1 and EV2). These results suggest that our approach is highly effective in accurately identifying patterns in breast cancer images. Besides, according to the obtained results in Table 2, EV 1 works better in assessing the performance of a CBMIR tool since it succeeds in reporting the recall while EV 2 suffers from a notable decrease in measuring the recall. So, in the following experiments, EV 1 evaluates the CBMIR performance for prostate cancer as the multi-class data set.

In order to evaluate the effectiveness of our CBMIR method on a multi-class data set, we utilized SICAPv2 and Arvaniti data sets, both containing four classes. Given the global prevalence of prostate cancer, we selected this type of cancer for our experiments and used SICAPv2 as the largest

TABLE 2. Comparative results on BreaKHis 400× at $k = 5$. We measure the precision and recall with both EV 1 and EV 2.

Type of evaluation	Method	Precision	Recall
EV 1	UCBMIR	**0.92**	**0.99**
EV 2	UCBMIR	**0.91**	**0.50**
	Minarno [25]	0.70	0.31
	Gu [47]	0.63	-

TABLE 3. Model quality results on SICAPv2 with top 5 retrieved images. The reported results are obtained by EV 1. The metrics are precision, recall, and accuracy. All other studies reported their results at the top 5 images.

Method	Data set	Precision	Recall	Accuracy
UCBMIR	SICAPv2	0.79	**0.80**	0.79
UCBMIR	Patches normalized Arvaniti (SICAP)	0.71	0.75	**0.80**
UCBMIR	Patches normalized Arvaniti (Panda)	**0.80**	0.68	0.78
Hegde [48] *(SMILY)*	TCGA	-	-	0.73
Hegde [48] *(SIFT)*	TCGA	-	-	0.62
VGG16 *(ImageNet))*	SICAPv2	**0.80**	**0.80**	**0.80**

pixel-wise annotated data set. The Arvaniti data set was re-sampled by referencing Panda and SICAP, as stated in [17], and was used in two experiments of this paper to demonstrate the robustness of our methodology. Table 3 presents the results obtained using our approach with $K = 5$ and EV1 as the evaluation criteria. To demonstrate the efficacy of our methodology, we conducted two experiments using the Arvaniti data set. In the first experiment, to ensure a fair comparison, we trained the model using Arvaniti normalized based on both SICAP and Panda. The results of these two trained models, obtained by conducting the entire training and searching steps, are reported in Table 3. These findings demonstrate the superior performance of our approach in accurately identifying and classifying prostate cancer images in multi-class data sets, thereby potentially contributing to the development of improved diagnostic tools and clinical decision-making processes.

In a study by Hegde [48], Scale-Invariant Feature Transform (SIFT) [49] was used as a traditional FE, along with SMILY, to report the accuracy of retrieving images with the correct Gleason patterns from prostate specimens in TCGA. Our UCBMIR, as shown in Table 3, achieved an accuracy of 80%, surpassing SMILY's accuracy of 73%.

To provide an interpretive perspective for the quantitative results, we incorporated a pre-trained VGG16 [50] (ImageNet) as a backbone to extract histological features from the images. We added a GlobalMaxPooling2D (GMP) and two dense layers [200, 4] to train the model as a classifier in a fully-supervised manner. After training the model, we removed the last layer (Dense (4)) and used the remaining
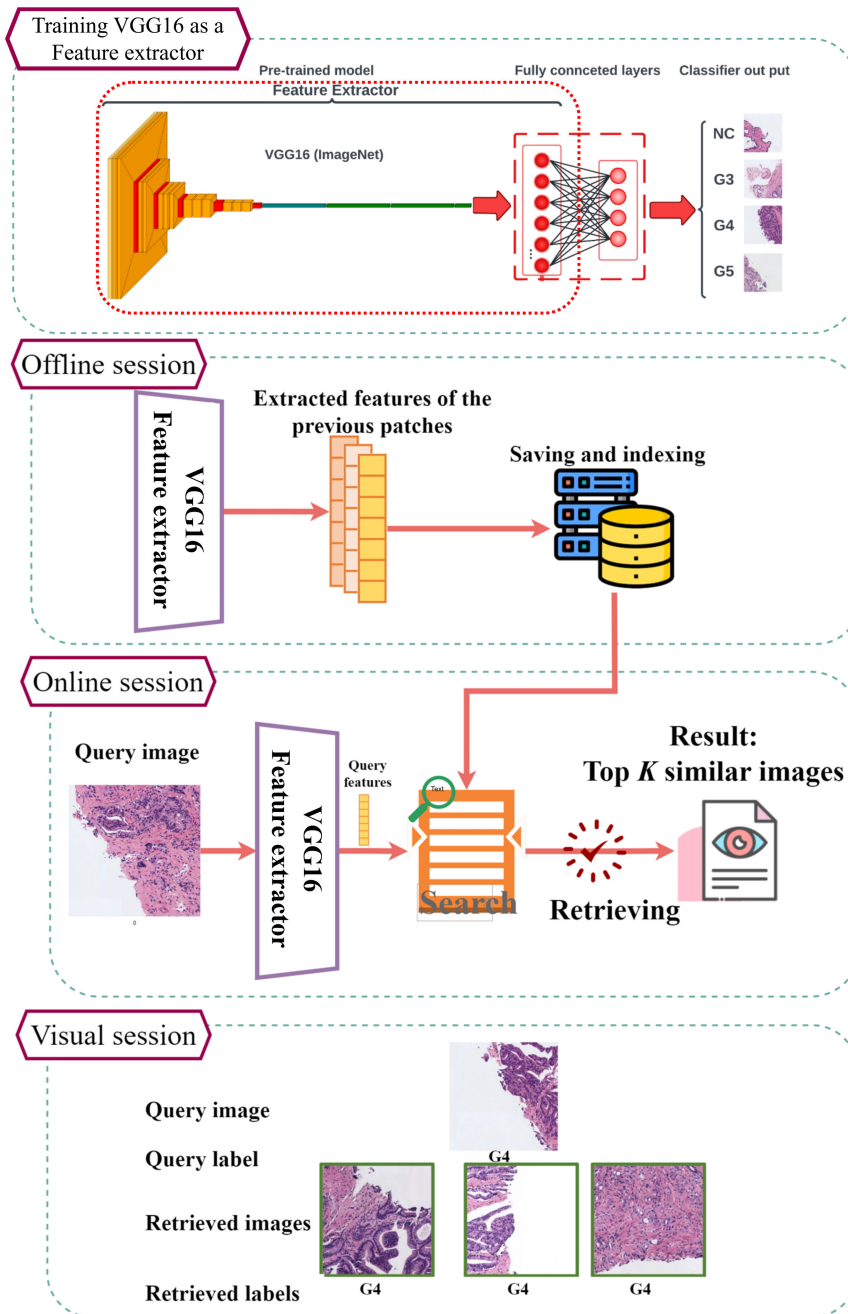
FIGURE 3. The structure of training VGG as a multi-class classification on SICAPv2 and delivering 200 features per image to the following CBMIR steps. The Conv2D layers are shown in "orange", MaxPooling2D in "red", Dense layers in "green", and flatten in "teal". By discarding the last dense layer of the well-trained VGG16, the FE can be moved to the next sessions to extract the features of the images. This is just used as a fully supervised baseline method to compare the performance of the proposed unsupervised FE in extracting features with respect to the one extracted from the classification-task VGG training.

layers as an FE to extract 200 features per image, which were then fed into the search engine component of UCBMIR. Figure 3 illustrates how we integrated the pre-trained VGG16 into our CBMIR. So, as can be seen in Figure 3, first, the VGG16 trained as a classifier by adding two dense layer with 200 and 4 nodes. Subsequently, the well-trained VGG16, in conjunction with the dense layer with 200 features, was

moved to the offline and online sessions of the proposed CBMIR platform to extract the features of the data set and the query. Following the retrieval of the top K images, it is time to display and evaluate the performance of the CBMIR platform with a supervised FE (VGG16). Comparing the results shown in Table 3, UCBMIR achieved a comparable performance as the supervised method with EV1.
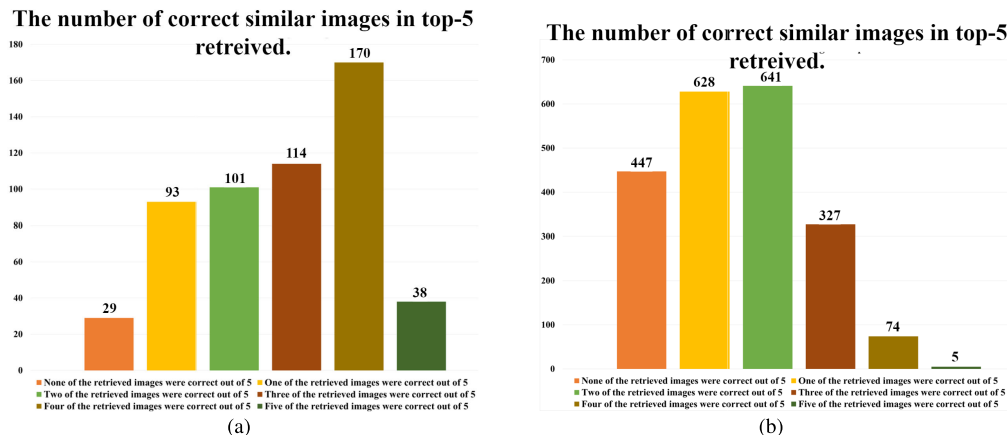
FIGURE 4. Evaluation of the UCBMIR at k = 5 on BreaKHis 4a and SICAPv2 4b. From 2122 query images in SICAPv2, for 447 cases, the model could not find at least one correct similar image according to their labels, while it retrieved two similar images at 5 top for 641 cases.

We conducted experiments using SICAPv2 and Arvaniti data sets to evaluate and compare the effectiveness of our UCBMIR with respect to the fully-supervised VGG16 in identifying and classifying prostate cancer images. As shown in Tables 3 our unsupervised method achieved similar results to the supervised VGG16, with a slightly lower precision score in both EV 1 (by 0.01). This suggests that our unsupervised method is a promising approach for CBMIR in the context of prostate cancer, potentially reducing the need for manual annotation and supervision.

In this study, Figure 4a and Figure 4b were used to present the results of the experiments. The bar charts were used to depict the number of similar images out of $K = 5$ retrieved images for BreaKHis and SICAPv2, respectively. For the BreaKHis data set, 545 images in the test set were used as query images. Based on Figure 4a, the model failed to find at least one similar image for 29 queries, while it could find three and four similar images for 114 and 170 queries, respectively. In the case of SICAPv2 data set, the model could retrieve one similar image among the top K for 628 image queries, according to the results shown in Figure 4b. The model was able to retrieve two images out of 5 in the same class label as the query in 101 cases of BreaKHis. In the case of SICAPv2, the model could retrieve two images with the same class label for 628 queries out of the top 5 retrieved images. So, the proposed UCBMIR approach can correctly retrieve at least one sample belonging to the same category as the query in 78.9% of cases for the SICAPv2 data set. In the BreaKHis data set, this retrieval success rate is even higher at 94.7%. To the best of the authors' knowledge, these results are very promising in the field of automatic atlases.

Due to the well-known variability between pathologists in Gleason grading and variations in histology sample preparation, it is a difficult challenge to distinguish between different grades of prostate cancer. These factors may contribute to the differences in results. Differentiating between G3 and G4 in prostate cancer requires highly experienced pathologists, takes time, and has limited inter-pathologist repeatability.

However, Figure 5 demonstrates the impressive ability of our UCBMIR to identify similar patterns between G3 and G4. Each row and column in Figure 5 corresponds to three different values of K in Arvaniti (Panda), SICAPv2, and Arvaniti (SICAP), respectively.

These results highlight the potential of our approach to aid in the accurate identification and CBMIR of prostate cancer images, thereby facilitating diagnosis and improving patient outcomes. Further research is needed to validate these findings on larger and more diverse data sets. Due to this, our model is also verified on an external data set with the intention of evaluating the trained model's capacity for generalization.

A. VALIDATION ON AN EXTERNAL DATA SET
In order to validate the performance of our trained model on an external data set, we utilized the SICAPv2-trained model to make predictions on the re-sampled Arvaniti data set (normalized by SICAP). The results of this evaluation are reported in Table 4. The obtained accuracy and precision results are slightly better than those obtained from the test set on SICAPv2, while the recall is slightly lower by 0.07. It is important to mention that this validation process is crucial in demonstrating the generalization ability of our UCBMIR beyond the original training data set. These results further confirm the robustness of our approach and its capability to provide accurate retrieval results across multiple data sets.

TABLE 4. Results of Arvaniti (normalized by SICAP) as data set for the external experiments with top K = 3, 5, and 7, EV 1.

K	Precision	Recall	Accuracy
3	0.66	0.58	0.66
5	0.80	0.73	0.81
7	0.89	0.86	0.91

B. VISUAL EVALUATION
We have included three figures, Figure 6, Figure 7, and Figure 8, which showcase the results of our experiments.
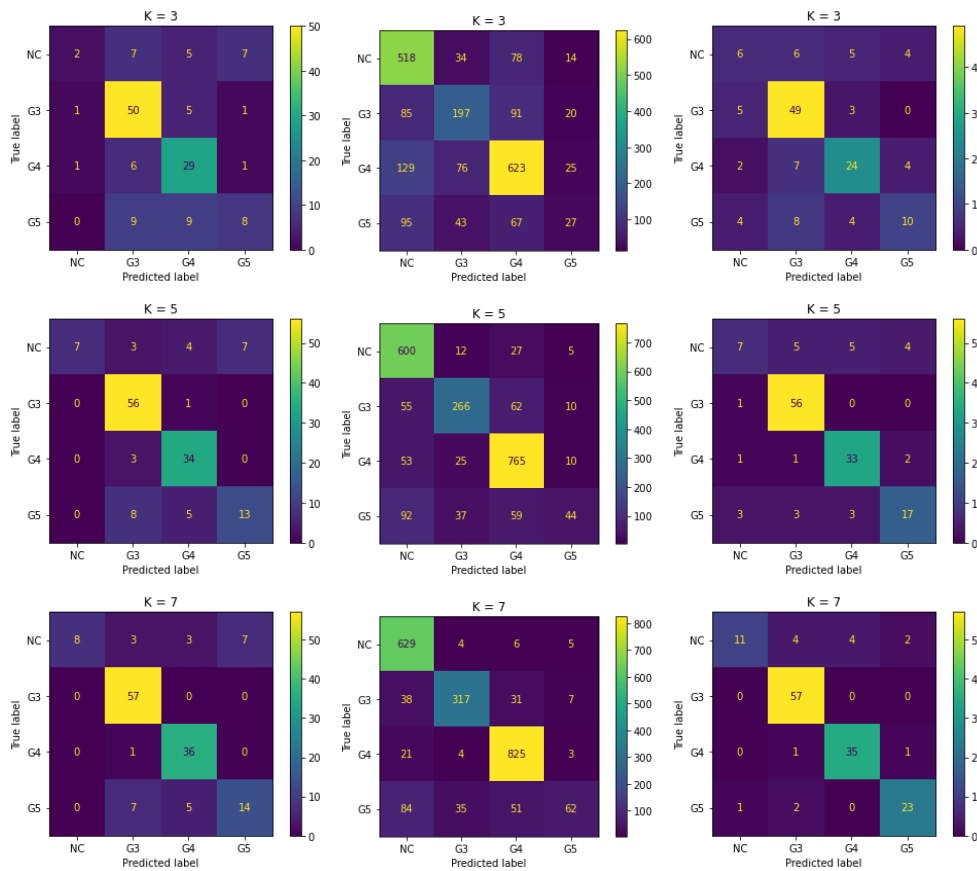
FIGURE 5. Confusion matrix of UCBMIR on the different test cohorts at $K = 3, 5, 7$. a. ARVANITI (Panda), b. SICAPv2, c. ARAVNITI (SICAP).
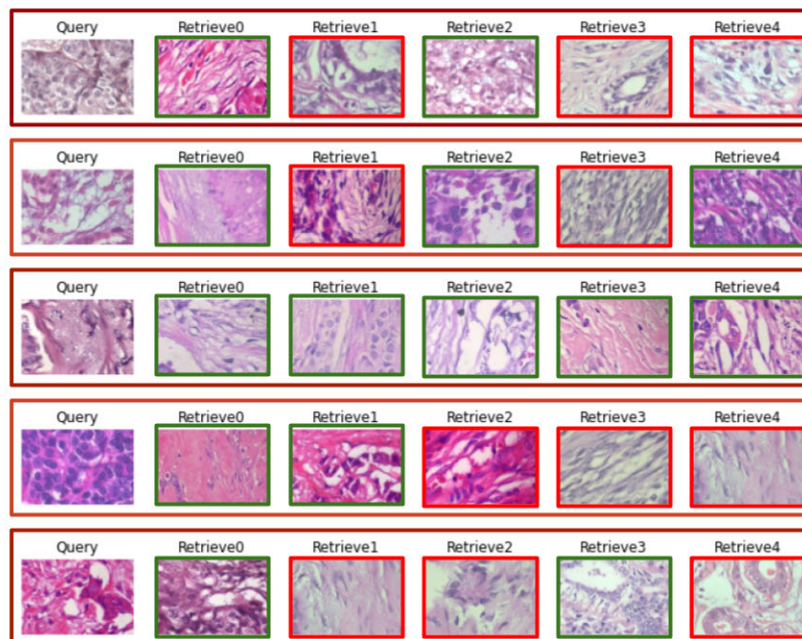

FIGURE 6. The top 5 images retrieved from the BreakHis data set for five randomly selected query images, with the true and false retrieval results depicted in green and red boxes, respectively.

These three figures illustrate the interpretability of the proposed UCBMIR.

The purpose of these figures is to enhance comprehension of the comparison by utilizing visual aids. Each of these
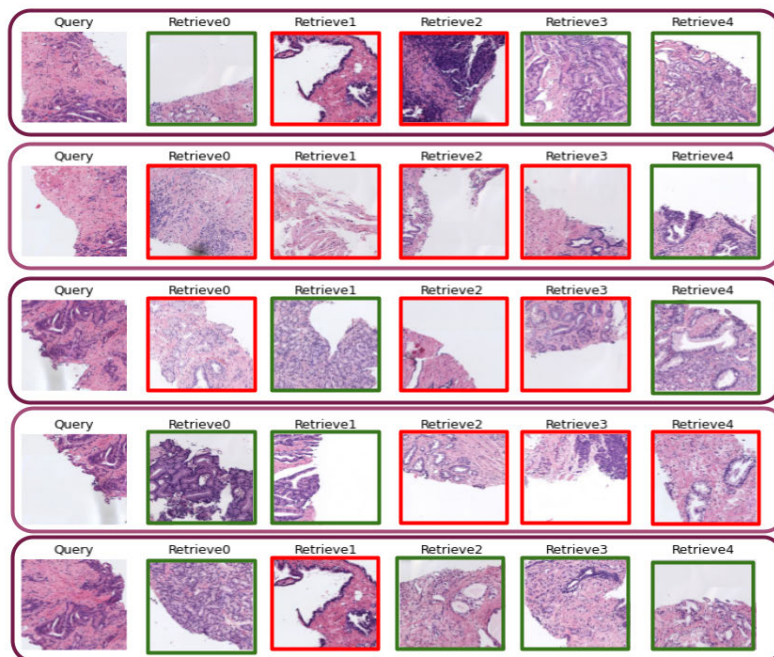
FIGURE 7. The top 5 retrieved images from the SICAPv2 data set for five query images selected at random, where true and false retrieval results are respectively indicated with green and red boxes.

figures comprises rows that correspond to a random query from the test set of BreaKHis×400, SICAPv2, and Arvaniti, respectively. The subsequent images in each row exhibit the top 5 retrieved images from the training set of the relevant data set. We have implemented a color-coding scheme to facilitate the interpretation of the results. In particular, a green border surrounds the correct retrieved image, which possesses the same label as the query, whereas a red border highlights the mis-retrieved images that have different labels than the query.

Figure 7 demonstrates that one of the challenges we encountered in our experiments with SICAPv2 was the presence of a white background in the images. To determine whether the patches in SICAPv2 contained meaningful patterns for pathologists to analyze, we enlisted the help of an expert pathologist to review them. Our pathologist confirmed that despite the presence of a white background in the images, there was still enough tissue for pathologists to evaluate and compare the patterns in the query tissue with the retrieved patches. There are some bad cases as shown in line 2 of Figure 7 which most of the retrieved images are not highlighted with the red border. This means that the tool had challenges in retrieving similar patches for this query. The reason is mainly due to the high similarity of histopathological features of G3 and G4 of prostate cancer.

In addition to validating the approach of UCBMIR using an external data set and demonstrating the generalization capability of our method, we selected the Arvaniti data set for another reason: it does not have a white background. Figure 8 shows the top 5 retrieved images resulting from our external

validation experiment, where a well-trained model with SICAPv2 was used to retrieve images for five random queries from the Arvaniti data set. Our external validation experiment not only validates our proposed UCBMIR for use with external data sets but also demonstrates the generalization capability of our method.

Through our visual evaluation, we aim to present a clearer understanding of the effectiveness of our approach. Observing the retrieved images alongside their labels can be useful to evaluate the performance of our method and assess its strengths and limitations. These figures are an essential component of our evaluation and will contribute significantly to understanding our methodology. Overall, this evaluation can provide valuable insights into the performance of our approach and make informed judgments regarding its effectiveness.

C. COMPARING UCBMIR WITH A CLASSIFIER
CBMIR and classification are two different approaches in medical image analysis. CBMIR aims to retrieve similar images from a database based on the content features of a query image, while classification aims to categorize images into pre-defined classes or labels. The only mutual output of the classification tool and CBMIR is the output labels corresponding to the output patches. In the above sections, it is mentioned that UCBMIR achieved comparable results with supervised CBMIR techniques, proven by the reported accuracies in Table 3. Regarding comparing the predicted labels in terms of retrieving similar images belonging to the same cancer type, we provide Table 5. This table compares
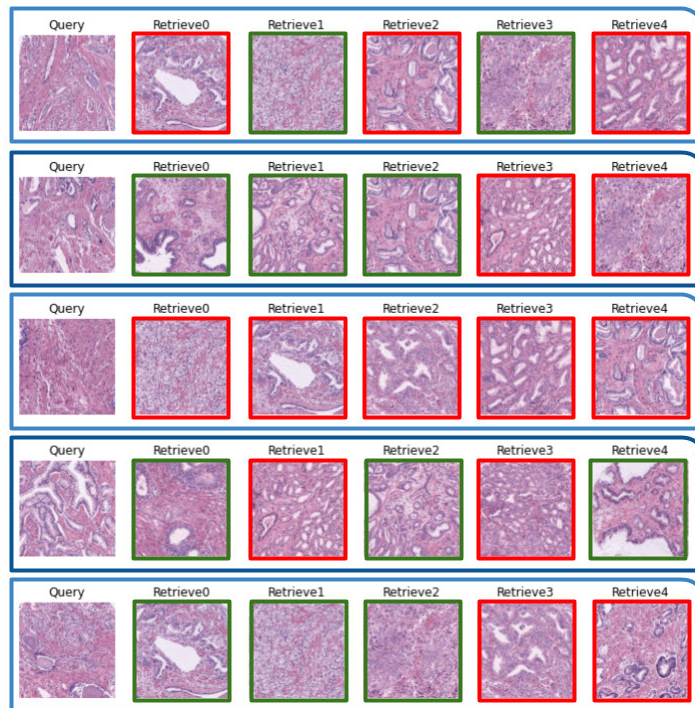
FIGURE 8. The top 5 images retrieved from the Arvaniti data set, which have been normalized by SICAP, for five query images that were randomly selected. The figure visually demonstrates the results of the external validation, utilizing a well-trained model with the SICAPv2 data set and applying it in the search stage of Arvaniti. The green and red boxes, respectively, indicate the true and false retrieval results.
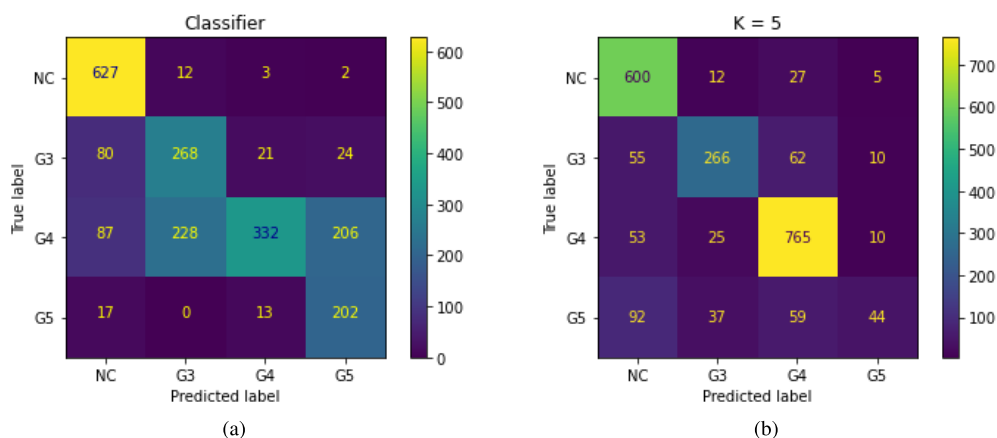

FIGURE 9. The confusion matrix presented in [17] is displayed in Figure 9a, while the matrix for the retrieved labels is shown in Figure 9b. It can be observed that the UCBMIR model results in less conflict between the challenging grades (G3 and G4) compared to the classifier.

the accuracy of UCBMIR with a classifier in [17] in both the validation and the test set of SICAPv2.

The proposed unsupervised CBMIR model was found to be highly effective in distinguishing between different cancer grades, especially between the challenging Gleason grades G3 and G4. This observation was evident from the confusion matrix shown in Figure 9. The proposed method's success can be attributed to its ability to identify and utilize subtle features and patterns in the images that may be missed by human observers or conventional supervised models.

D. LIMITATIONS

The limitation of the pre-processing step and the medical session can directly affect the final results of the DL-based UCBMIR. For instance, noisy images and low-quality scanners can diminish the accuracy of the retrieval task. However, CAE was chosen to tackle the noisy images while training; the low quality of images might still affect the final results. Color variation as a result of different staining processes, different scanners, and laboratory conditions might fool the DL-based models. In this paper, adding a

TABLE 5. Shows a comparison between the performance of UCBMIR with EV 1 and the classification introduced in [17]. SICAPv2 is the data set under study.

Data set	Model	Accuracy
Validation set	UCBMIR	**0.83**
	Classification [17]	0.76
Test set	UCBMIR	**0.79**
	Classification [17]	0.67

histogram equalization technique during the image loading process, could tackle this issue, and as can be seen in Figure 9b and Figure 6, the proposed framework is robust against this issue. Due to the computational limitation, large-scale retrieval is a limitation that hinders the performance of the CBHIR tools in searching through a vast database of biopsies. The proposed tool can retrieve the images at high speed. In the SICAPv2 experience, it can retrieve the top 5 images in almost 0.28 seconds.

Although some challenges have been solved by the proposed tool, some challenges still remain such as semantic gap, scalability, cross-domain retrieval, privacy and security, etc. For instance, medical images need high data protection due to the personal information. This affects extending the CBMIR domain as a worldwide tool across different hospitals. Another important limitation is the semantic gap between the extracted features of the low-level histopathological features and high-level semantic concepts. To address this issue, FEs have to train effectively to extract the most representative features of the patches, fast and accurately. However, limitation or inconsistent annotations makes the well-training of the DL-based FEs difficult. Despite all these limitations, the CBHIR tool provides a promising improvement in cancer diagnosis, research, and treatment. Several pieces of research aim to address many of these issues and further enhance the capability of the CBHIR tool and provide more transparency, robustness, and trustworthiness.

V. CONCLUSION

In summary, this paper introduces a highly qualified Unsupervised CBMIR (UCBMIR) model that can be used for both binary and multi-class data sets. The model was evaluated on three different data sets, as well as an external validation set. Using the two most-used evaluation techniques, the proposed method achieved 79% precision in EV 1 on SICAPv2 as a multi-class data set. Notably, the unsupervised method was able to differentiate between challenging Gleason grades of prostate cancer. In addition to numerical evaluation, visual assessments were conducted to demonstrate the effectiveness of the UCBMIR. The results show that UCBMIR has good generalizability and can be effectively applied to other types of cancer. UCBMIR has the potential to improve laboratory productivity, increase pathologists' diagnostic confidence, and contribute to the advancement of cancer diagnosis and treatment.

The UCBMIR model not only addresses the needs and challenges of pathologists but also addresses the problem of engineers who face a lack of sufficient images for training models. Future research in this field could build on these findings and further enhance the performance of CBMIR models for cancer diagnosis.

VI. FUTURE WORK

In the world of CBMIR, there is a vast range of possibilities for enhancing and optimizing laboratory productivity. With a large archive of diagnosed patients and corresponding data, including images and treatment and monitoring reports, it should be possible to identify and retrieve images that are either anatomically or pathologically similar to the biopsy sample of the patient being examined, as well as the annotated data for each case. CBMIR has the potential to be applicable to many types of cancer, which would increase its utility.

Furthermore, pathologists' reports contain the medical knowledge of many other pathologists for similar cases, making them a treasure trove of high-quality diagnostic information. In the future of CBMIR, it may be possible to make the raw information directly available to the pathologist or to merge the important information in retrieved reports. This would make the diagnosis process more efficient, accurate, and informative for both the pathologist and the patient. Additionally, expanding the use of CBMIR to other types of medical imaging and diagnostic data could provide valuable insights for a range of medical specialties.

In order to integrate CAD tools into daily clinician routines, reliability, trustworthiness, and transparency are the most critical needs. By incorporating eXplainable AI (XAI) methodologies like filter activations, the decision of the DL-based tools can be demystified. In regard to integrating these tools with the clinical workflows, it is necessary to harmonize DL-based tools alignment with the diagnostic precision and the quality of patient care.

AUTHOR CONTRIBUTIONS STATEMENT

Zahra Tabatabaei was responsible for the conception of the experiments, analysis of results, validation, writing of the original draft, and study design. At the same time, all other authors reviewed the manuscript.

REFERENCES

[1] M. Piñeros, L. Mery, I. Soerjomataram, F. Bray, and E. Steliarova-Foucher, "Scaling up the surveillance of childhood cancer: A global roadmap," *J. Nat. Cancer Inst.*, vol. 113, no. 1, pp. 9–15, Jan. 2021.

[2] H. Singh, A. N. D. Meyer, and E. J. Thomas, "The frequency of diagnostic errors in outpatient care: Estimations from three large observational studies involving U.S. adult populations," *BMJ Quality Saf.*, vol. 23, no. 9, pp. 727–731, Sep. 2014.

[3] *Global Action Plan on Physical Activity 2018–2030: More Active People for a Healthier World*, World Health Org., Geneva, Switzerland, 2019.

[4] S. Kalra, H. R. Tizhoosh, C. Choi, S. Shah, P. Diamandis, C. J. V. Campbell, and L. Pantanowitz, "Yottixel—An image search engine for large archives of histopathology whole slide images," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101757.

[5] H. R. Tizhoosh and L. Pantanowitz, "Artificial intelligence and digital pathology: Challenges and opportunities," *J. Pathol. Informat.*, vol. 9, no. 1, p. 38, Jan. 2018.

[6] S. Fuster, F. Khoraminia, U. Kiraz, N. Kanwal, V. Kvikstad, T. Eftestøl, T. C. M. Zuiverloon, E. A. M. Janssen, and K. Engan, "Invasive cancerous area detection in non-muscle invasive bladder cancer whole slide images," in *Proc. IEEE 14th Image, Video, Multidimensional Signal Process. Workshop (IVMSP)*, Jun. 2022, pp. 1–5.

[7] N. Kanwal, S. Fuster, F. Khoraminia, T. C. M. Zuiverloon, C. Rong, and K. Engan, "Quantifying the effect of color processing on blood and damaged tissue detection in whole slide images," in *Proc. IEEE 14th Image, Video, Multidimensional Signal Process. Workshop (IVMSP)*, Jun. 2022, pp. 1–5.

[8] I. Kiran, B. Raza, A. Ijaz, and M. A. Khan, "DenseRes-unet: Segmentation of overlapped/clustered nuclei from multi organ histopathology images," *Comput. Biol. Med.*, vol. 143, Apr. 2022, Art. no. 105267.

[9] H. Hu, S. Qiao, Y. Hao, Y. Bai, R. Cheng, W. Zhang, and G. Zhang, "Breast cancer histopathological images recognition based on two-stage nuclei segmentation strategy," *PLoS ONE*, vol. 17, no. 4, Apr. 2022, Art. no. e0266973.

[10] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 119–130, Jan. 2016.

[11] W. Li, J. Li, K. V. Sarma, K. C. Ho, S. Shen, B. S. Knudsen, A. Gertych, and C. W. Arnold, "Path R-CNN for prostate cancer diagnosis and Gleason grading of histological images," *IEEE Trans. Med. Imag.*, vol. 38, no. 4, pp. 945–954, Apr. 2019.

[12] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. 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.

[13] W. Li, J. Li, J. Polson, Z. Wang, W. Speier, and C. Arnold, "High resolution histopathology image generation and segmentation through adversarial training," *Med. Image Anal.*, vol. 75, Jan. 2022, Art. no. 102251.

[14] Z. Tabatabaei, A. Colomer, K. Engan, J. Oliver, and V. Naranjo, "Self-supervised learning of a tailored convolutional auto encoder for histopathological prostate grading," in *Proc. 31st Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2023, pp. 980–984.

[15] F. Prinzi, M. Insalaco, A. Orlando, S. Gaglio, and S. Vitabile, "A yolo-based model for breast cancer detection in mammograms," *Cognit. Comput.*, pp. 1–14, Aug. 2023.

[16] J. Li, W. Li, A. Sisk, H. Ye, W. D. Wallace, W. Speier, and C. W. Arnold, "A multi-resolution model for histopathology image classification and localization with multiple instance learning," *Comput. Biol. Med.*, vol. 131, Apr. 2021, Art. no. 104253.

[17] J. Silva-Rodríguez, A. Colomer, M. A. Sales, R. Molina, and V. Naranjo, "Going deeper through the Gleason scoring scale: An automatic end-to-end system for histology prostate grading and cribriform pattern detection," *Comput. Methods Programs Biomed.*, vol. 195, Oct. 2020, Art. no. 105637.

[18] M. Liu, Y. He, M. Wu, and C. Zeng, "Breast histopathological image classification method based on autoencoder and Siamese framework," *Information*, vol. 13, no. 3, p. 107, Feb. 2022.

[19] N. Ahmad, S. Asghar, and S. A. Gillani, "Transfer learning-assisted multi-resolution breast cancer histopathological images classification," *Vis. Comput.*, vol. 38, no. 8, pp. 2751–2770, Aug. 2022.

[20] R. Man, P. Yang, and B. Xu, "Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks," *IEEE Access*, vol. 8, pp. 155362–155377, 2020.

[21] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," *J. Digit. Imag.*, vol. 32, no. 4, pp. 605–617, Aug. 2019.

[22] X. Li, X. Shen, Y. Zhou, X. Wang, and T.-Q. Li, "Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet)," *PLoS ONE*, vol. 15, no. 5, May 2020, Art. no. e0232127.

[23] M. R. Abbasniya, S. A. Sheikholeslamzadeh, H. Nasiri, and S. Emami, "Classification of breast tumours based on histopathology images using deep features and ensemble of gradient boosting methods," 2022, *arXiv:2209.01380.*

[24] F. Taheri, K. Rahbar, and P. Salimi, "Effective features in content-based image retrieval from a combination of low-level features and deep Boltzmann machine," *Multimedia Tools Appl.*, vol. 82, no. 24, pp. 37959–37982, Oct. 2023.

[25] A. E. Minarno, K. M. Ghufron, T. S. Sabrila, L. Husniah, and F. D. S. Sumadi, "CNN based autoencoder application in breast cancer image retrieval," in *Proc. Int. Seminar Intell. Technol. Appl. (ISITIA)*, Jul. 2021, pp. 29–34.

[26] P. Mazaheri, A. A. Bidgoli, S. Rahnamayan, and H. R. Tizhoosh, "Ranking loss and sequestering learning for reducing image search bias in histopathology," *Appl. Soft Comput.*, vol. 142, 2023, Art. no. 110346

[27] M. Rasoolijaberi, M. Babaei, A. Riasatian, S. Hemati, P. Ashrafi, R. Gonzalez, and H. R. Tizhoosh, "Multi-magnification image search in digital pathology," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 9, pp. 4611–4622, Sep. 2022.

[28] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "DenseNet: Implementing efficient ConvNet descriptor pyramids," 2014, *arXiv:1404.1869.*

[29] S. Kalra, H. R. Tizhoosh, S. Shah, C. Choi, S. Damaskinos, A. Safarpoor, S. Shafiei, M. Babaie, P. Diamandis, C. J. V. Campbell, and L. Pantanowitz, "Pan-cancer diagnostic consensus through searching archival histopathology images using artificial intelligence," *npj Digit. Med.*, vol. 3, no. 1, pp. 1–15, Mar. 2020.

[30] P. A. Fashi, S. Hemati, M. Babaie, R. Gonzalez, and H. R. Tizhoosh, "A self-supervised contrastive learning approach for whole slide image representation in digital pathology," *J. Pathol. Informat.*, vol. 13, 2022, Art. no. 100133.

[31] L. Zheng, A. W. Wetzel, J. Gilbertson, and M. J. Becich, "Design and analysis of a content-based pathology image retrieval system," *IEEE Trans. Inf. Technol. Biomed.*, vol. 7, no. 4, pp. 249–255, Dec. 2003.

[32] X. Shi, F. Xing, K. Xu, Y. Xie, H. Su, and L. Yang, "Supervised graph hashing for histopathology image retrieval and classification," *Med. Image Anal.*, vol. 42, pp. 117–128, Dec. 2017.

[33] L. Pantanowitz, P. Michelow, S. Hazelhurst, S. Kalra, C. Choi, S. Shah, M. Babaie, and H. R. Tizhoosh, "A digital pathology solution to resolve the tissue floater conundrum," *Arch. Pathol. Lab. Med.*, vol. 145, no. 3, pp. 359–364, Mar. 2021.

[34] J. C. Caicedo, F. A. González, and E. Romero, "Content-based histopathology image retrieval using a kernel-based semantic annotation framework," *J. Biomed. Informat.*, vol. 44, no. 4, pp. 519–528, Aug. 2011.

[35] X. Qi, D. Wang, I. Rodero, J. Diaz-Montes, R. H. Gensure, F. Xing, H. Zhong, L. Goodell, M. Parashar, D. J. Foran, and L. Yang, "Content-based histopathology image retrieval using CometCloud," *BMC Bioinf.*, vol. 15, no. 1, pp. 1–17, Dec. 2014.

[36] A. Sridhar, S. Doyle, and A. Madabhushi, "Content-based image retrieval of digitized histopathology in boosted spectrally embedded spaces," *J. Pathol. Informat.*, vol. 6, no. 1, p. 41, Jan. 2015.

[37] J. T. Kwak, S. M. Hewitt, A. A. Kajdacsy-Balla, S. Sinha, and R. Bhargava, "Automated prostate tissue referencing for cancer detection and diagnosis," *BMC Bioinf.*, vol. 17, no. 1, pp. 1–12, Dec. 2016.

[38] S. Zhu, Y. Li, S. Kalra, and H. R. Tizhoosh, "Multiple disjoint dictionaries for representation of histopathology images," *J. Vis. Commun. Image Represent.*, vol. 55, pp. 243–252, Aug. 2018.

[39] M. Babaie, S. Kalra, A. Sriram, C. Mitcheltree, S. Zhu, A. Khatami, S. Rahnamayan, and H. R. Tizhoosh, "Classification and retrieval of digital pathology scans: A new dataset," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 760–768.

[40] 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.

[41] 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.

[42] E. Arvaniti, K. S. Fricker, M. Moret, N. Rupp, T. Hermanns, C. Fankhauser, N. Wey, P. J. Wild, J. H. Rüschoff, and M. Claassen, "Automated Gleason grading of prostate cancer tissue microarrays via deep learning," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, Aug. 2018.

[43] W. Bulten et al., "Artificial intelligence for diagnosis and Gleason grading of prostate cancer: The PANDA challenge," *Nature Med.*, vol. 28, no. 1, pp. 154–163, 2022.

[44] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Training very deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[46] 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.

[47] Y. Gu and J. Yang, "Multi-level magnification correlation hashing for scalable histopathological image retrieval," *Neurocomputing*, vol. 351, pp. 134–145, Jul. 2019.

[48] N. Hegde, J. D. Hipp, Y. Liu, M. Emmert-Buck, E. Reif, D. Smilkov, M. Terry, C. J. Cai, M. B. Amin, C. H. Mermel, P. Q. Nelson, L. H. Peng, G. S. Corrado, and M. C. Stumpe, "Similar image search for histopathology: SMILY," *npj Digit. Med.*, vol. 2, no. 1, pp. 1–9, Jun. 2019.

[49] N. Mehta, R. S. Alomari, and V. Chaudhary, "Content based sub-image retrieval system for high resolution pathology images using salient interest points," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2009, pp. 3719–3722.

[50] S. Liu and W. Deng, "Very deep convolutional neural network based image classification using small training sample size," in *Proc. 3rd IAPR Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2015, pp. 730–734.

ZAHRA TABATABAEI received the bachelor's degree in electronic engineering from Bu-Ali Sina University, in 2017, and the M.Sc. degree in electronic engineering from the Hamedan University of Technology, Hamedan, Iran, in 2019. She is currently pursuing the Ph.D. degree with the Polytechnic University of Valencia, supervised by Prof. Valery Naranjo and Dr. Adrián Colomer. During the master's thesis, she focused on the clustering and classification of hyperspectral images. Her Ph.D. is a part of the European Marie-Curie Program (CLARIFY) (http://www.clarify-project.eu/), focused on creating a robust diagnostic environment for digital pathology. Additionally, she is also a Researcher with Tyris Software Company under the supervision of Dr. Javier Oliver Moll. She has experience as a Lecturer and a Teacher Assistant with the Hamedan University of Technology and Bu-Ali Sina University, Hamedan. Her research interests include image processing, deep learning, information retrieval, and computer science.

ADRIÁN COLOMER received the degree in telecommunications engineering, the master's degree in telecommunication technologies, systems, and networks, and the Ph.D. degree in telecommunications, in 2013, 2014, and 2018, respectively. He coordinates research with CVBLab and directs the deep learning applied to signal & image analysis course with the Polytechnic University of Valencia (UPV), where he is currently a Telecommunications and Computer Sciences Professor, authorizing more than 50 international publications.

JAVIER OLIVER MOLL received the M.Sc. degree in telecommunications and the Ph.D. degree in image processing from the Technical University of Valencia, Spain, in 2006 and 2015, respectively. He is currently a Board Member of the Telecommunications Society COITCV. Since 2006, he has been working for different industries and research centers in the field of image processing and artificial intelligence. In 2012, he co-founded MyBrana, a tech startup that added augmented reality to social networks. In 2014, he co-founded the technological firm Tyris, where he is also the CEO. During these years, he has conducted more than 100 projects in digital transformation using artificial intelligence, mainly in the fields of manufacturing, media, and health. He combines his work with Tyris with academics as a part-time Professor with VIU and Nebrija University.

VALERY NARANJO (Member, IEEE) received the Graduate and Ph.D. degrees in telecommunications from the Polytechnic University of Valencia (UPV), in 1995 and 2002, respectively.

She currently holds the position of a Full Professor in signal theory with UPV, the Director of the Research and Innovation in Bioengineering Institute (I3B), and the Academic Head of the Telecommunications Degree. Over the years, she has taken the lead on numerous European and national projects focused on image analysis and machine learning. She holds the role of a Coordinator of the Clarify Project. One of its main lines of research is the development of computer-aided systems applied to different fields, such as medicine, biology, and quality control in industrial applications.

• • •