Endowing a Content-Based Medical Image Retrieval System with Perceptual Similarity Using Ensemble Strategy

Marcos Vinicius Naves Bedo¹ · Davi Pereira dos Santos¹ · Marcelo Ponciano-Silva² · Paulo Mazzoncini de Azevedo-Marques³ · André Ponce de León Ferreira de Carvalho¹ · Caetano Traina Jr¹

Published online: 11 August 2015
© Society for Imaging Informatics in Medicine 2015

Abstract Content-based medical image retrieval (CBMIR) is a powerful resource to improve differential computer-aided diagnosis. The major problem with CBMIR applications is the semantic gap, a situation in which the system does not follow the users' sense of similarity. This gap can be bridged by the adequate modeling of similarity queries, which ultimately depends on the combination of feature extractor methods and distance functions. In this study, such combinations are referred to as perceptual parameters, as they impact on how images are compared. In a CBMIR, the perceptual parameters must be manually set by the users, which imposes a heavy burden on the specialists; otherwise, the system will follow a predefined sense of similarity. This paper presents a novel approach to endow a CBMIR with a proper sense of similarity, in which the system defines the perceptual parameter depending on the query element. The method employs ensemble strategy, where an extreme learning machine acts as a meta-learner and identifies the most suitable perceptual parameter according to a given query image. This parameter defines the search space for the similarity query that retrieves the most similar images. An instance-based learning classifier labels the query image following the query result set. As the concept implementation, we integrated the approach into a mammogram CBMIR. For each query image, the resulting

tool provided a complete second opinion, including lesion class, system certainty degree, and set of most similar images. Extensive experiments on a large mammogram dataset showed that our proposal achieved a hit ratio up to 10% higher than the traditional CBMIR approach without requiring external parameters from the users. Our database-driven solution was also up to 25% faster than content retrieval traditional approaches.

Keyword Content-based medical image retrieval · Computer-aided diagnosis · Similarity queries

Introduction

The increasing availability of radiological studies has enabled computer-aided diagnosis (CAD) to become part of the clinical routine for image diagnosis [1–3]. When a new case study is required, radiologists can opt to use CAD tools for detecting specific types of abnormalities [4]. Moreover, differential diagnosis CAD (CADx) can classify malignant and benign lesions by adding a potential “second opinion” [5]. The main problem with CADx systems is that their opinion-making process is usually unclear and enables no further discussions.

According to Doi et al. [1], an alternative CADx process would provide a set of similar past cases and the likelihood of malignancy (%). Content-based medical image retrieval (CBMIR) is a prospective technology through which a defined approach could recover past cases based on image patterns [6, 7]. In CBMIR systems, an image can be retrieved by ranking strategies or similarity queries. The similarity queries paradigm has been widely adopted by content-based image retrieval systems [8, 9] as it supports both handling and indexing of large datasets. Moreover, it relies on relational database management systems that can be straightforwardly

✉ Marcos Vinicius Naves Bedo
bedo@icmc.usp.br

¹ Institute of Mathematics and Computer Science, University of São Paulo, São Paulo, Brazil

² Institute of Education, Science and Technology of the Triângulo Mineiro, Uberaba, Brazil

³ Ribeirão Preto Medical School, University of São Paulo, São Paulo, Brazil

connected to other clinical systems (e.g., PACS) [10]. By using similarity queries, we model the content retrieval task in a metric space. A metric space is defined as $\langle \mathbb{I}, d \rangle$, where $\mathbb{I}$ is the domain of represented images and d is a metric distance function that enables the comparison of a pair of elements $\in \mathbb{I}$ [11]. Feature extractor methods are usually employed to represent the original content of an image in $\mathbb{I}$ [10, 12].

Thus, two of the tuning parameters of similarity queries are the feature extractor method and the distance function. In this study, they are referred to as perceptual parameters, as they impact on how the images are compared. The perceptual parameter's choice is not a simple task, as it depends on the semantics of domain and application, which generates a *semantic gap*. For instance, regarding the diagnostic hypothesis, what do the extracted features represent? Or, concerning the search space, which distance function allows the retrieval of the most similar images?

The number of perceptual parameters increases exponentially as new feature extractor methods and distance functions are developed. Manually setting these parameters is a heavy burden for experts and, sometimes, unfeasible. This factor contributes to the lack of CBMIR use in clinical practice [13, 14]. Therefore, to increase CBMIR accuracy and usability automatically endowing a CBMIR with proper perceptual parameters is a topic of major interest [7, 13]. This paper proposes a novel technique that combines supervised learning with similarity queries for the design of a parameter-free CBMIR. Although the method can be employed in any domain, the experiments conducted used mammogram images. CBMIR Higiiia [15] was updated to fulfill all steps of the proposal. Accordingly, the main contributions of this study have been summarized as follows:

- A strategy to define statistical relationships between the perceptual parameters and the previous diagnosed images.
- A method that automatically defines the feature extractor domain and the distance function for every image retrieval requisition.
- Implementation of the proposed method into a mammogram CBMIR and evaluation of accuracy gains and query performance.

Once the previous cases and the corresponding diagnoses are stored in a relational database, our proposal uses a two-stage classification model. The first employs a supervised strategy for automatically selecting the CBMIR perceptual parameters. The second stage employs the selected perceptual parameters as the similarity query settings and performs the image retrieval. Therefore, the most similar past cases are shown, and the query image is also classified following the instance-based learning technique (IBL) [16].

As a concept implementation, we updated the version of the Higiiia tool for a parameter-free image retrieval and

classification. To evaluate the overall impact of our method, we tested it against the traditional image retrieval approach over a mammogram dataset. Our method achieved an accuracy ratio up to 10% higher than the traditional CBMIR approach and was 25% faster.

The remainder of the paper is structured as follows: “Background and Related Works” section reviews the related work and background necessary for this study; “The Proposed Approach” section describes the proposed method; “Experimental Results” section provides the experimental results from the comparison between the proposed method and related approaches; finally, “Conclusions and Future Work” section addresses future works and conclusions.

Background and Related Works

This section briefly presents the main concepts required for the understanding of our proposal and discusses some of the related works.

Components of a CBMIR System

CBMIR systems address the problem of retrieving similar medical images. Distinct approaches such as ranking strategies, classification, or similarity queries [1, 17] can be used for content retrieval. According to similarity queries in metric spaces, a CBMIR architecture can be designed as interactions between offline and online modules [12, 9]. Figure 1 illustrates these modules and their interactions. The query process is triggered when the user poses one (or multiple) image as a query element.

The offline processing is carried out by the application of feature extractor methods over a set of diagnosed images. A relational database management system (RDBMS) enables this extraction process to be triggered after each insertion. Therefore, this architecture is designed to support an ever-increasing knowledge database. Once the image has been represented, it is also indexed. We highlight that one perceptual parameter is required for the indexing of an image regarding the metric space approach. Such index techniques, known as metric access methods (MAMs) employ the distance between represented images to partition the search space and improve the performance by orders of magnitude [18]. Analogously, this architecture enables the deletion of a single image without the need of reconstructing the entire database.

On the other hand, the online processing is triggered by an expert who submits an image as a query element. The user must also define the perceptual parameter employed in the query. The query element is represented by the specified feature extractor method and pairwise compared to the database elements according to the defined distance

Fig. 1 Generic pipeline for offline and online CBMIR stages

function. The elements are then grouped following the calculated distances and a given criterion. The k -NN criterion has been widely used for most CBMIR systems. It retrieves the k elements with the lowest distances to the query element [11, 6]. However, other criteria can also be used. For instance, range retrieves the elements that are at most at a given distance ξ , while the diversity criterion fetches the k most influential elements with the lowest distances to the query element [19]. More sophisticated criteria, as diversity, attempts to avoid the need for further relevance feedback cycles. Nevertheless, after visualizing the retrieved images, the expert may decide to explore the search space more deeply and provide some feedback. The CBMIR refines the query regarding the provided perceptual parameter, search criterion, and feedback parameters.

The perceptual parameters define the search space [13, 20, 21] and may lead to the (un)suitable performance of similarity queries. As our focus is on bridging this gap by automatically setting the perceptual parameters at the CBMIR online stage, hereafter, we summarize the main concepts used to model our approach. For a medical image domain $\mathbb{S}$, the active domain S ; $S \subset \mathbb{S}$, is the set of images stored in an RDBMS.

Definition 1 Feature Extractor Method (*FEM*): Given an extractor domain $\mathbb{F}$, a feature extractor method $FEM : \mathbb{S} \rightarrow \mathbb{F}$ is a computational function that represents any medical image $e \in \mathbb{S}$ as its summarization in $\mathbb{F}$.

Definition 2 Feature Vector (*FV*): Given an *FEM* F , the feature vector f_i is the concise representation of the original image s_i in $\mathbb{F}$, such that $F(s_i) = f_i$.

Definition 3 Metric Distance Function (*MDF*): A distance function $d : \mathbb{F} \times \mathbb{F} \rightarrow \mathbb{R}$ evaluates the distance, or dissimilarity, between any pairs of elements in $\mathbb{F}$. Given an *FEM* F , and three medical images $s_i, s_j, s_l \in \mathbb{S}$, d is also said to be a metric

distance function if satisfies the following three properties:

Symmetry: $d(F(s_i), F(s_j)) = d(F(s_j), F(s_i))$;

Non-negativity: $0 \leq d(F(s_i), F(s_j)) < \infty$;

Triangular inequality: $d(F(s_i), F(s_j)) = d(F(s_i), F(s_l)) + d(F(s_l), F(s_j))$;

Definition 4 Perceptual Parameters (*PP*): A perceptual parameter *PP* is a pair $PP = \langle F, d \rangle$ where F is an *FEM* and d is a distance function. Therefore, in a CBMIR system, $\mathbb{P} = \{PP_1, PP_2, \dots, PP_n\}$ is the set of all combinations between the available feature extractor methods and distance functions. Each perceptual parameter may provide semantics for either the application or the user.

Definition 5 k -Nearest-Neighbor Query (k -*NN_q*): Given a query element $s_q \in \mathbb{S}$, a perceptual parameter PP_q , a number of neighbors k , $k \in \mathbb{N}^+$, the k -nearest neighbor query answer is the subset of S such that k -*NN_q*(s_q, k, PP_q) = $\{s_n \in S \mid \forall s_i \in S; d(F(s_n), F(s_q)) \leq d(F(s_i), F(s_q))\}$. The k -*NN_q*(s_q, k, PP_q) is the search that uses the k -NN criterion.

Most CBMIR systems store the images in relational database management systems. It is a portable way to integrate CBMIR to other related tools, e.g., picture archiving and communication systems (PACS) [22]. However, the content retrieval operations are performed still at the application level, as similarity queries are not naturally supported by commercial RDBMS. On the other hand, recent studies have designed prototypes to extend such core routines [23, 10], where the user can straightforwardly pose similarity queries into a RDBMS using a proper relational language. The next section reviews the CBMIR Higiia that uses an extended structured query language to perform the operations of image storage, deletion, update, and querying.

The CBMIR Higiiia and the Semantic Gap Challenge

Higiiia is a CBMIR software [15] that supports perceptual retrieval, i.e., it enables the setup of perceptual parameters on-the-fly, adjusting the similarity evaluation to the specific image context. It is manually configured by an expert, who must know the semantics of the perceptual parameters regarding both system and image's domain. The major strengths of Higiiia are the native support to several file formats, including the standard DICOM [23], and the full integration of the retrieval engine over an RDBMS. Therefore, Higiiia can handle subsets of an enterprise's medical image database in an efficient and scalable manner, building distinct medical contexts through filters over DICOM or traditional attributes.

This CBMIR is basically composed of the modules of Fig. 1. In the offline stage, the feature extractor module extracts the visual features from the images, which are then indexed and stored in RDBMS, according to every distance function available. In the online stage, the perceptual parameters are embedded in a “query subsystem” that performs the retrieval and feedback relevance steps. The subsystem receives the queries from the user interface, constructs the extended-SQL statements, and invokes the RDBMS to execute them. The query types include standard queries (i.e., queries that employ traditional attributes as study number or patient name), DICOM metadata-based queries, and similarity queries. Whenever possible, the system employs MAM indexes to answer the similarity queries. For traditional attributes or DICOM metadata, Higiiia uses RDBMS indexes.

The Higiiia also provides a relevance feedback module as the query refinement option. It collects information on the users' interactions, which is then employed in a tuned query. It is also stored as the user profile [15]. The user-friendly interface was designed to support mammogram images; therefore, the expert can visualize the associated diagnosis, the BI-RADS, and density and subtlety regarding each retrieved image. To enable a query refinement, the user may discard the images judged as non-relevant as well as highlight the retrieved instance judged as relevant. The expert may request as many relevance feedback cycles as necessary. Finally, Higiiia requests a user diagnosis containing BI-RADS, density, and subtlety, which are stored in the database.

Although Higiiia provided few feature extractor methods and distance functions in its first instantiation, it can support the addition of new feature extractor methods and distance functions, as it relies on an extended RDBMS solution to perform similarity queries. As new image extraction techniques are developed, they can be straightforwardly integrated to Higiiia. However, it still requires that the expert select the perceptual parameter for the search retrieval, which increases the semantic gap and reduces the tool usability [13].

According to the ontology proposed by Deserno et al. [24], semantic gap is a situation in which the system does not follow the radiologists' sense of similarity. For CBMIR systems based on the metric space approach, the similarity queries are always performed following the perceptual parameters [10, 23]. Therefore, non-similar medical images are recovered mainly due to the incorrect setting of perceptual parameters. Thus, a previous deep knowledge of perceptual parameters and the target image domain would be required from the user.

Bridging this gap by merely leaving the task of setting computer parameters to the health specialists also increases the *usability gap* and contributes to the CBMIR lack of use in clinical environments [25]. The relation between a perceptual parameter and a diagnosis class and/or visual disease characteristics is not intuitive [26, 4] and requires a critical expert-charging process. Moreover, the number of possible combinations quickly scales up whenever new FEMs and MFDs are added to the system. Therefore, for a wide collection of FEMs and MDFs, manually setting the query parameters becomes unfeasible.

Combining CAD and CBMIR for Analyses of Mammograms

Some modules of CBMIR architecture are also present in most of CAD definitions [27, 17]. While the primary goal of the CBMIR tool is to retrieve medical images regarding content [24, 7], CAD applications are commonly designed to perform image classification [4, 37, 28]. Table 1 summarizes the state-of-the-art CAD and CBMIR applied over mammograms. This set of studies was compiled according to the evaluation of the DDSM¹ database, except for the last two reviewed works.

We highlight here some common features for the reviewed literature strategies. Most of them are based on *black-box* classification strategies, in which no deep clues about similarity values between every pair of images. Moreover, no scalability experiments were provided for the testing of the robustness of the approaches for larger datasets. The majority of studies were also designed specifically to handle a particular pathology, e.g., mass, which implies that the solution space is bounded according to a diagnosis hypothesis. The work of Bosis [29] relies on a semi-automated segmentation process, where the resulting images are represented through texture features and then classified by an ensemble of artificial neural networks (ANNs) [30], achieving varying results for evaluated “fatty” and “dense” classes. In an interesting way, the work of Bovis presented the important result that employing ensemble strategy boosts the overall image classification, a premise that is used in our hypothesis in the next section.

¹ <http://marathon.csee.usf.edu/Mammography/Database.html>

Table 1 Summary of reviewed CADx and CBMIR techniques for mammograms

Author	Classes	Classifier	Images	Source	Results
Bovis and Singh 2002	4	ANN	377	DDSM	40–71 %
Elter and Haßlmeier 2008	2	IBL	360	DDSM	85 % (ROC)
Mazurowski et al. 2011	2	RMHC	1852	DDSM	49–82 %
Wei et al. 2011	2	SVM	2563	DDSM	72–74 %
Tao et al. 2011	2	IBL	415	Private	75 % (ROC)
Deserno et al. 2012	20	SVM	900	IRMA	78 %

The work of Elter [28] combines content and clinical data to calculated weights used in a classification according to decision rules. Although the approach achieved a high ROC curve value, it employed only a few elements of DDSM. Moreover, the steps strongly depend on previously defined parameters, as similarity weights and the thresholds of decision rule. The approach of Mazurowski [31] defines a template-based strategy to recognize malignant masses. It depends on previous constructed templates, employed for the definition of likelihood maps.

On the other hand, the proposal of Wei [32] proposes to capture users' similarity perception through support vector machines (SVMs) [33]. The classifier is used in a relevance feedback strategy for the analysis of the users' information necessary for the retrieval of similar past cases. Two important characteristics of this approach are calcification and mass are handled alike and the solution is based on hierarchical similarity measurement architecture. The proposal achieved a precision ratio up to 70 %. However, this technique is heavily biased through the relevance feedback interactions, which determine the similarity by updating a probability function.

The last two approaches in Table 1 aim at combining CAD and CBMIR strategies as a single tool. The study of Tao [34] proposes an architecture where masses are previously segmented and represented using shape features. To lobulated and irregular shape features, curvature scale descriptors and radial length were used, while texture features were extracted from mass margin. The similarity is obtained by the locally linear embedding. The last proposal [14] employs the principal component analysis (PCA) over 128×128 pixel patches extracted from each mammogram region of interest and then classified by two SVMs. The query image posed by the user is used in the calculation of the distances to hyper planes that divide the stored data into relevant or not relevant. This two-class problem enables the retrieval of most relevant images.

Our approach differs from the reviewed ones as it does not depend on users' interference in the CBMIR initial query and automatically sets the query parameters in the online stage. Moreover, it can handle distinct pathologies alike in the same database and employs structures designed for the metric space approach to ensure scalability for larger datasets. Our solution also extends the CBMIR as a CADx tool by using ensemble strategy.

Extreme Learning Machines and Instance-Based Learning

To combine CBMIR and CAD strategies, we must employ a scalable labeling process for both approaches. In this paper, we resort to two classifiers, the extreme learning machines (ELMs) and IBL. Extreme learning machines have drawn considerable attention because of their simplicity and performance in comparison with other supervised classifiers. The main strengths of the ELM technique are (i) they have only one parameter to adjust, which is the hidden layer size L , and (ii) they are trained in a single step [35, 36].

Internally, an ELM models the classification problem as a set of linear equations, regarding the instances (medical images) and objective class (pathology). The set of equations composes a matrix, where each row corresponds to a feature vector from the training set and one equivalent single-row matrix contains the mapped classes. Therefore, the goal is to calculate the weights that provide the best overall approximation for the set of equations. Calculating such weights is equivalent to training the classifier. This procedure can be quickly solved by the pseudo-inverse strategy. The ELM extension, known as on-line sequential ELM (OS-ELM) [37], is of particular interest, as it enables the fast ELM updating regarding new incoming images.

The IBL classifier straightforwardly applies the concept of similarity search in metric spaces for classification purposes [16]. IBL is based on three main components: (i) a similarity function, which calculates the (dis)similarity between two instances for ranking them; (ii) a classification function, which receives the retrieved instances and classifies them based on a previous rule and considering the majority of similar instances; (iii) a concept description updater, which maintains the record of classification performances and decides on whether the new information should be added (i.e., the set of instances employed to drive further classifications).

Components (i) and (ii) are naturally provided by the k - NN_q procedure. Therefore, IBL variations can employ the concept description for the final classification. The widely employed IBL variation is IB1, which labels the query image as the majority class regarding retrieved elements. Another IBL variation is IB2 strategy, which performs the same steps of IB1, although it has a policy for treating misclassified

instances. Finally, IB3 employs a “wait and see” strategy to determine whether representative instances can be stored and used in further classifications. The application of IB1 as a classifier offers several advantages, such as robustness and representation of both probabilistic and overlapping concepts. IB1 approach is also simple and follows the natural experience-based human reasoning. It also enables further analysis to complement the experts’ intuition.

The Proposed Approach

To the best of our knowledge, the reviewed CBMIR systems run using the definition of one unique predetermined perceptual parameter for image retrieval. In other words, such systems employ a single similarity perspective to compare images of the same domain, i.e., essentially employing a fixed feature extractor and a distance function and/or evaluator. This traditional approach, however, limits the images’ comparison to a single search space and disregards the notorious existence of distinct visual patterns between the classes of images of the same domain, caused by variations in the lesion types or even pathological severity levels.

This one-way comparison contributes toward increasing the semantic and usability gaps in CBMIR tools. It ignores the variations of the “perceptual similarity,” which are distinct ways to compare images of the same domain regarding defined pathological features. Our proposal is based on the hypothesis that the semantic gap can be bridged by automatically choosing such perceptual parameters. Therefore, the similarity query would depend on the similarity notion applied to each class of pathology.

Figure 2a summarizes the Fig. 1 query procedure regarding traditional CBMIR systems. Figure 2b shows our proposed solution to the challenge of using the suitable “notion of similarity” for the same search problem. Our rationale is generic, regardless of the image domain. Therefore, given a

query image of a valid domain, our procedure answers the two following questions:

- Q1. Regarding the similarity query, what perceptual parameter must be set to define the search space?
- Q2. Regarding the defined search space, what is the system classification to the query image concerning retrieved elements?

The offline and online stages of CBMIR architecture must be updated so that such questions can be answered. “Offline Training Step: Associating Perceptual Parameters and Pathological Classes in CBMIR” section describes an algorithm that updates the offline stage (Fig. 2b proposed training step offline)—it associates a perceptual parameter to a pathological class according to the training dataset. “Online CBMIR Query: an Ensemble-Based Strategy for Similarity Queries” section shows how to employ ensemble strategy in two phases for the updating of the online stage (Fig. 2b proposed CBMIR query online). The first phase automatically determines the most suitable perceptual parameter and defines the search space. The second stage performs the content retrieval over the defined search space. “Using CBMIR as a CADx” section addresses the integration of CBMIR and CADx as the result of the application of our approach. Using the IB1 classifier, the CBMIR also classifies the query element regarding the pathological class and severity levels. Finally, “The New Version of CBMIR Higiiia” section describes the update of the CBMIR Higiiia using the proposed concepts.

Offline Training Step: Associating Perceptual Parameters and Pathological Classes in CBMIR

The processing described in Algorithm 1 gathers statistics from the stored and diagnosed images. It uses the entire training set to determine the perceptual parameters that enable the highest hit ratio for every pathological class. It relies on the

Fig. 2 Summarized content retrieval procedure by (a) traditional CBMIR approach and (b) our proposal through perceptual parameters

IB1 classifier by setting the search space for every experimented feature extractor method and distance function.

Data: Pathological class c , Image $\text{pastCases}[]$, int k

Result: The most suitable perceptual parameter according to IB1 precision

```

PerceptualParameter pps[][];
PerceptualParameter pp;
foreach extractor method fem do
    foreach distance function df do
        countHit ← 0;
        for i = 0 to pastCases.size() do
            q ← pastCases[i];
            if q.class = c then
                setSearchSpace(fem, df);
                labIB1 ← IB1(pastCases - q, q, k);
                if c = labIB1 then
                    countHit ← countHit + 1;
                end
            end
        end
        pps[fem][df] = countHit;
    end
end
pp ← PerceptualParameterMaxHit(pps);
return pp;

```

Algorithm 1 Associating a pathological class with a perceptual parameter

The *PerceptualParameterMaxHit()* routine executes the search in the matrix of feature extractor methods and distance functions and returns the perceptual parameter with the highest hit ratio. The underlying hypothesis is the most precise the perceptual parameter in the classification, the most suitable to the similarity search. In our implementation and experiments, we used the mammogram domain regarding image classification in terms of the BI-RADS standard² summarized in Table 2.

Following this BI-RADS first edition, the mass and calcification findings were classified according to four categories, namely “2”, “3”, “4”, and “5”. For simplification purposes, we derived four classes from the BI-RADS standard. Images containing mass and BI-RADS 4, 5 were labeled as “malignant mass” and those containing mass with BI-RADS 2, 3 were labeled as “benignant mass”. The calcification images with BI-RADS 4, 5 were labeled as “malignant calcification,” while calcification images with BI-RADS 2, 3 were defined as “benignant calcification”.

In our proposal, the diagnoses of the mammogram dataset follow this four classes model. In the offline step, Algorithm 1 associates the most precise perceptual parameter with each of the four classes, according to the IB1 classifier. A variation of Algorithm 1 is to obtain the precision regarding the pathological classes whenever new perceptual parameters are added to the system, i.e., when a new feature extractor method or a distance function is inserted. This variation considers only the new perceptual parameter in the inner loops of Algorithm 1. This routine is particularly useful for keeping the system up-to-date as

Table 2 First edition BI-RADS patterns

Category	Assessment
2	Benign finding(s)
3	Probably benign finding(s)
4	Suspicious abnormality
5	Highly suggestive of malignancy

new representative extractors and/or distances are developed for the CBMIR domain. Finally, we highlight that the innermost loop of Algorithm 1 can be performed using extended SQL. It is a desirable feature that allows the algorithm to be triggered whenever perceptual parameters are inserted and/or updated.

Online CBMIR Query: an Ensemble-Based Strategy for Similarity Queries

This section details the proposed online module that answers questions Q1 and Q2. Let an undiagnosed image be the query element; the first step is to determine the search space. We propose the use of ensemble strategy in two stages, as illustrated in Fig. 3. In the first stage, the OS-ELM acts as a meta-learner, while IB1 uses the perceptual parameter predicted by the OS-ELM to define the search space and perform the final classification based on the most similar images.

In step 1, all feature extractor methods available are used to represent the query element and their feature vectors are concatenated into a single feature vector. Additionally, a feature selector method (e.g., the SymmetricalUncertAttributeEval [38]) reduces the dimensionality of the single image representation. The OS-ELM classifies this feature vector as one pathological class. However, the OS-ELM output is not the system’s final labeling.

The previous training stage (“Offline Training Step: Associating Perceptual Parameters and Pathological Classes in CBMIR” section) provides a direct mapping between the pathological classes and the perceptual parameters. We take advantage of this association to determine the perceptual parameter to define the most suitable search space to the second step of our ensemble. Therefore, question Q1 is answered regarding the similarity notion the CBMIR has *learned* from the historical database following the statistical data distribution.

In step 2, the ensemble employs the mapped perceptual parameter to define the search space. Therefore, the query element is represented only through the feature extractor method defined by the perceptual parameter. Whenever possible, a metric access method is used to speed-up the k -nearest neighbor query. The results of the similarity query are the most similar elements to the undiagnosed image.

Figure 3 summarizes the ensemble strategy pipeline. When the user provides a query element (i.e., an undiagnosed image) and the number k of similar objects to be retrieved, the system

² <http://www.acr.org>

Fig. 3 Online CBMIR and CADx pipeline according to our proposal

automatically performs the two ensemble steps in sequence, with no human interference. In step 1, the most suitable perceptual parameter is determined. In step 2, the CBMIR performs the similarity search according to the search space defined by the perceptual parameter. Finally, the query element is classified based on the most similar images (step 3). The next section describes how step 3 is used to answer Q2.

Using CBMIR as a CADx

The IB1 procedure relies on similarity queries; therefore, we use the results of step 2 to label the undiagnosed image. This additional feature enables the use of CBMIR as a CADx system. Following the IB1 premise, most query element nearest-neighbors are likely to pertain to the same class of the undiagnosed image. Therefore, IB1 uses their diagnosis to label

the query element. For instance, the diagnosis stores information, such as lesion type and BI-RADS regarding the mammogram domain. We used the IB1 labeling as the final step of our proposal, illustrated as step 3 in Fig. 3.

By bounding the search space in the earlier stages, we expect to boost the overall IB1 classification. Therefore, the system's prediction error can be expressed as a ratio between the IB1 and OS-ELM predictions, where the IB1 classification is the most critical. The ensemble strategy limits the final classification error according to the OS-ELM hit ratio, which determines the search space.

The ensemble classifiers are independent; therefore, let e_1 be the ELM error ratio when choosing a perceptual parameter and e_2 the IB1 error ratio when classifying an image. The overall image prediction error is expressed as the OS-ELM error plus the conditioned IB1 error, as given by Eq. 1.

$$\text{Error} = e_1 + ((1-e_1)*e_2) \quad (1)$$

The OS-ELM also enables incremental trainings and reduces eventual costs as new images are inserted into the RDBMS. We used this incremental property to define the ELM re-training policy. Considering the CBMIR periodically runs Algorithm 1, we propose the following rule: if Algorithm 1 determines that a new perceptual parameter becomes more representative than the previous one for any class, a new re-training is required. According to the OS-ELM incremental property, only the set of inserted images are used to update the classifier. On the other hand, if no mapping change occurs due to Algorithm 1 running, the OS-ELM does not need to be updated. We claim that such a policy avoids unnecessary and costly classifier updates.

The New Version of CBMIR Higiiia

In its first version, Higiiia enabled the addition of new feature extractor methods and distance functions through the extended-RDBMS architecture. Therefore, to perform the content retrieval operation, the user needed to manually set the perceptual parameters to define the search space.

The new version of CBMIR Higiiia (Higiiia v2) includes our proposed ensemble strategy for similarity queries. The interface has also been updated to present the new analysis to be used as a detailed second opinion to the user. Former system features, as relevant feedback cycles, image windowing, and image annotation are still available.

By answering questions Q1 and Q2, Higiiia v2 provides three main outputs: (i) the set of the most similar images regarding the query element; (ii) the image classification according to the search space and the retrieved elements; (iii) the classification certainty degree, which is bounded according to the search space.

Unlike the reviewed CBMIR approaches, Higiiia v2 enables queries on distinct types of lesion, i.e., calcification and mass, and uses no previously defined perceptual parameter to perform content retrieval. The tool also does not require the expert to provide any query and/or training parameter. Rather, for a given set of possible diagnoses, the system is expected to adopt a proper notion of similarity.

We have added four basic feature extractor methods, namely histogram, Haralick [39], Haar Wavelets, and Daubechies Wavelets [40] to Higiiia v2. The histogram extractor is the 256-dimension normalized histogram of image gray-tone pixels. The Haralick extractor is composed of texture relationships, such as entropy, variance, angular second moment, generating a 24-dimension representation. Daubechies and Haar wavelets are composed of features extracted from the Daubechies and Haar wavelet transforms, representing features such as energy, entropy, and creating a representation of 16 dimensions.

We have also added five distinct distance functions, namely City-Block, Euclidean, Chebyshev, Jeffrey Divergence, and Canberra. All of them can be used to measure similarity. The first three are members of the Minkowski family [11] and metric distance functions, as well as Canberra distance. Jeffrey Divergence is not a metric distance function because it does not comply with the triangular inequality. Such feature extractor methods and distance functions can be freely combined to create perceptual parameters.

Experimental Results

This section addresses the experiments performed with our technique for a differential computer-aided diagnosis. We ran two experiments over Higiiia v2 to determine (i) the accuracy of our strategy in comparison to the traditional CBMIR approach and (ii) the scalability regarding the training and retrieval stages. Basic feature extractor methods and distance functions were employed to illustrate how the technique can boost the retrieval and classification process.

The experiments were conducted over DDSM, a public dataset of nearly 2500 complete studies. Each study is composed of four mammogram projections, information related to breast features and diagnosis using the BI-RADS standard. The DDSM dataset was also employed as basis of studies reviewed in “Combining CAD and CBMIR for Analyses of Mammograms” section.

A set of images containing regions of interest were cropped from the DDSM dataset by researchers of our university. Each region of interest (ROI) was constructed cropping existing lesions with size up to 5×5 cm. The images that have not fulfilled the 5×5 cm constraint were excluded from the dataset. An excerpt of those images is shown in Fig. 4.

The total number of ROIs employed was 2919. We labeled the ROIs following our four-class model. Therefore, 615 images were labeled as benignant calcification, 574 as malignant calcification, 906 as benignant mass, and 824 as malignant mass. The training dataset has included only images with abnormalities. However, the same reasoning here could be applied to any other mammogram classes.

Associating a Pathological Class and Perceptual Parameters

We ran Algorithm 1 over the stored dataset to determine the associations among the four possible pathological classes and the perceptual parameters. In our experiments, we construct the perceptual parameters using extractors histogram, Haralick, Haar Wavelets, and Daubechies Wavelets and distance functions City-Block, Euclidean, Chebyshev, Canberra, and Jeffrey divergence.

Fig. 4 Examples of ROI images from each class of the experimented dataset. (a) Benignant calcification, (b) malignant calcification, (c) benignant mass, and (d) malignant mass

We randomly inserted two thirds of the dataset and collected statistics regarding the perceptual parameters and the four possible classes. The remaining one third of dataset images were used as query elements in the accuracy experiments. This procedure was repeated ten times. Figure 5 illustrates the average IB1 hit ratio concerning the search spaces employed. The resulting associations among perceptual parameters and pathological classes were the following:

- Histogram and Chebyshev for benignant calcification class—Fig. 5a.

- Histogram and Jeffrey Divergence for malignant calcification class—Fig. 5b.
- Haar Wavelets and City-Block for benignant mass class—Fig. 5c.
- Daubechies Wavelets and Jeffrey Divergence for malignant mass class—Fig. 5d.

The OS-ELM training was performed with two thirds of the dataset. The images were represented by the 32 most relevant features according to the SymmetricalUncertAttributeEval feature selector over the representation generated by the four

Fig. 5 Average precision regarding Algorithm 1 and grouped by (a) benignant calcification, (b) malignant calcification, (c) benignant mass, and (d) malignant mass

Fig. 6 The main screen of Higii v2 CBMIR/CADx tool

feature extractor methods available. The representations were provided to the supervised ELM classifier alongside with their respective labels using the four classes' model. The OS-ELM incremental learning property was evaluated in the scalability experiments.

Querying with the Automatically Selected Perceptual Parameter

In the training stage, we mapped the perceptual parameters with possible classes and trained the OS-ELM to classify mammogram images. Therefore, in the online stage, the ensemble strategy was used to provide a CBMIR and CADx output. The ensemble represented the images posed as query elements using the most prominent features according to the available feature extractor methods and labeled them using the OS-ELM.

The perceptual parameter associated with the OS-ELM output was used to define the similarity query search space. Thus, in this stage, the CBMIR is endowed with a proper

sense of similarity according to the selected perceptual parameter. The retrieved images are used by the IB1 classifier, which labels the query elements. The only required parameter is the number k of images to be retrieved.

This entire process can be triggered in Higii v2 interface illustrated in Fig. 6. Figure 6b provides the most similar k images according to the search space defined by the perceptual parameter, as shown in Fig. 6a. Figure 6c displays the classification suggested for the labeling of the query element. Figure 6d shows the system certainty degree according to Eq. 1. Other desirable CBMIR and image processing features, such as relevance feedback strategies and windowing, are also available in the same interface (Fig. 6e).

Accuracy Experiments

We compared the prediction accuracy of our proposal to the traditional CBMIR approach that employs only one search space to perform image retrieval. Therefore, in our

Fig. 7 Accuracy ratio achieved by our proposal and traditional content retrieval

experiments, we used the perceptual parameter with the best average precision regarding the four mammogram classes as the traditional CBMIR setting.

According to the measures of “Associating a Pathological Class and Perceptual Parameters” section -Fig. 5, the perceptual parameter with the best average precision was the (Histogram, City-Block). To perform the accuracy comparison, we used two thirds of the dataset were employed in the training stage and the remaining one third as query elements. The dataset was randomly shuffled ten times, and the methods’ hit ratio is provided in Fig. 7. The best and worst perceptual parameter selection are the theoretical upper and lower bounds, respectively.

As it can be seen in Fig. 7, for the malignant mass and malignant calcification classes, our proposal achieved an

average hit ratio 14% higher than the traditional approach. For benignant calcification class, the accuracy gain was 3.5 %. For the images of benignant mass class, the proposed method was 6.5 % more accurate than the competitor. On average, our method was 10.3 % more accurate than the traditional CBMIR. In comparison to the methods reviewed in “Combining CAD and CBMIR for Analyses of Mammograms” section, our method achieved an equal or higher hit ratio, particularly when used to mass classification.

Figures 8 and 9 show examples of query elements used in the experiment and the result sets according to the compared approaches. Figure 8a, b and Fig. 9a, b present query examples for the benignant calcification, malignant calcification, benignant mass, and malignant mass classes, respectively.

Fig. 8 Query examples for $k=5$, regarding query images with calcifications

Figure 10 shows the precision vs. recall graphs for the mammogram classes. According to them, our strategy improved the similarity queries precision in all cases. Regarding the recall values, our technique achieved a precision gain up to 13 % for malignant calcification class, while for malignant mass, our approach achieved a notable precision gain up to 15.2 %. We emphasize that users usually ask for small numbers of similar images; therefore, the low values of recall are the most relevant ones.

Our technique improved the CBMIR accuracy even for a basic set of feature extractor methods and distance functions. Although limited by the perceptual parameter semantics, our approach boosted our mammogram CBMIR endowing the

system with a proper similarity sense. Nevertheless, we highlight more powerful feature extractor methods, and distance functions can be easily incorporated into our method without loss of generality. The method's main strength is the search space accurate definition. It bounds the precision error and enables the retrieval of relevant images.

Scalability Experiments

In the last experiment, we evaluated the time required to perform the similarity queries using our proposal. The values were obtained as the average of 700 queries posed over the entire dataset. The experiments were performed over an Intel

Fig. 9 Query examples for $k=5$, regarding query images with masses

Fig. 10 Precision vs. recall graphs for the four classes model: (a) benignant calcification, (b) malignant calcification, (c) benignant mass, and (d) malignant mass

Core i5 processor with 2 GB memory running at 2.67 GHz under Ubuntu 12.04.

Two measurements were taken: the time spent on the OS-ELM training according to our predefined policy and the time spent on the similarity queries with automatically selected perceptual parameters. Both measurements were expected to increase in function of the size of the database. We populated the database starting with 700 images and incrementally provided *chunks* of 100 images until all 2919 regions of interest had been inserted.

According to our proposed policy, the OS-ELM must be updated only when one perceptual parameter associated with any defined class has changed. Table 3 shows the measurements regarding the OS-ELM training. The second column of Table 3 shows the number of re-training when the proposed policy was employed. The third column shows the number of re-trainings for updating the OS-ELM after the insertion of each data chunk. The last column shows the system's accuracy with and without the policy.

Table 3 Time spent to keep OS-ELM updated

Chunks	Policy	No policy	Accuracy
7–12	2.7	4	−0.13 %
13–16	2.3	4	+0.1 %
17–20	2.3	4	−0.17 %
21–24	2.1	4	−0.12 %
25–29	1.8	4	−0.1 %
Total	11.2	20	−0.97 %

Our policy has required shorter processing time. In the aggregate time, our proposal was 40 % times faster, at the cost of being only a little (0.97 %) less accurate. On average, our proposed policy required eight less re-trainings. Finally, we measured the average time spent on the 700 similarity queries varying the dataset size. Figure 11 shows the average time required by both our proposal and the traditional approach to answer the same query. Notice whenever possible, our system resorted to an MAM to speed-up the *k*-nearest neighbor queries. Our approach was 25 % faster on average when compared to content retrieval that used no metric access method.

Fig. 11 Average time to perform a similarity query over the experimented dataset

Conclusions and Future Work

Differential diagnosis CAD tools can potentially help radiologists as they provide a “second opinion.” An opened question for this technology is how to bridge the *semantic gap* between the application and the users. The suitable choice of perceptual parameters may help to significantly reduce this gap. Manually setting such parameters may be unfeasible due to the large amount of feature extractor methods and distance functions available in the medical image domain.

This paper has presented an approach that enables the system to query a medical image according to its own sense of similarity. The approach is based on ensemble strategy, in which a supervised OS-ELM is used as a meta-learner to define the most suitable perceptual parameter. Such a parameter is used to delimit the search space employed in the similarity query and image classification.

This strategy enables the system to provide complete results, including suggested medical image labeling, system certainty degree, and set of most similar images regarding the similarity notion adopted by the system. The strategy was fully integrated into a mammogram CBMIR named Higiiia v2. We used a basic set of feature extractor methods and distance function to demonstrate how the approach can boost the image retrieval and classification in a scalable manner. Experiments over a large mammogram dataset showed that the proposal increased a traditional CBMIR accuracy and required no external parameter from the users.

Our proposal can be extended to several other medical domains, as it is based on a database-centric solution. Previously proposed feature extractor methods and distance function can be also incorporated in our approach without loss of generality. As future work, we will focus on the variations of ensemble-based strategies and how to create search spaces that include image content and annotations.

References

- Doi K: Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Comput Med Imaging Graph* 31:198–211, 2007
- Astley S: Computer-based detection and prompting of mammographic abnormalities. *Br J Radiol* 77:194–200, 2014
- Pereira R, Azevedo-Marques P, Honda M, Kinoshita S, Engelmann R, Muramatsu C, Doi K: Usefulness of Texture Analysis for Computerized Classification of Breast Lesions on Mammograms. *J Digit Imaging* 20:248–255, 2007
- Nakayama R, Abe H, Shiraishi J, Doi K: Evaluation of Objective Similarity Measures for Selecting Similar Images of Mammographic Lesions. *J Digit Imaging* 24:75–85, 2011
- Jalalian A, Mashohor S, Mahmud H, Saripan M, Ramli A, Karasfi B: Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clin Imaging* 37:420–426, 2013
- Azevedo-Marques P, Rangayyan T: Content-based Retrieval of Medical Images: Landmarking, Indexing, and Relevance Feedback. *Synthesis Lectures on Biomedical Engineering*. Morgan & Claypool, New Jersey, NJ, 2013
- Bugatti P, Kaster D, Ponciano-Silva M, Traina Jr, C, Azevedo-Marques P, Traina A: PROSPER: Perceptual similarity queries in medical CBIR systems through user profiles. *Comput Biol Med* 45:8–19, 2014
- Alto H, Rangayyan R, Desautels J: Content-based retrieval and analysis of mammographic masses. *J Electronic Imaging* 14: 023016-1-17, 2007
- Kohli M, Warnock M, Daly M, Toland C, Meenan C, Nagy P: Building Blocks for a Clinical Imaging Informatics Environment. *J Digit Imaging* 27:174–181, 2014
- Bedo M, Traina A, Traina Jr, C: Seamless Integration of Distance Functions and Feature Vectors for Similarity-Queries Processing. *J Inf Data Manag* 5:308–320, 2014
- Zezula P, Amato G, Dohnal V, Batko M: Similarity Search - The Metric Space Approach. *Advances in Database Systems*. Springer, Heidelberg, GE, 2006
- Guelde M, Thies C, Fischer B, Lehmann T: A generic concept for the implementation of medical image retrieval systems. *Int J Med Inform* 76:2–3, 2007
- Ponciano-Silva M, Souza J, Bugatti P, Bedo M, Kaster D, Braga R, Belucci A, Azevedo-Marques P, Traina-Jr C, Traina A: Does a CBIR system really impact decisions of physicians in a clinical environment? *Comput Based Med Syst*, 2013.
- Deserno T, Soiron M, Oliveira J, Araujo A: Computer-aided diagnostics of screening mammography using content-based image retrieval. *Proc SPIE Med Imaging*, 2013. doi:10.1117/12.912392
- Bedo M, Ponciano-Silva M, Kaster D, Bugatti P, Traina A, Traina Jr. C: Higiiia: A Perceptual Medical CBIR System Applied to Mammography Classification. *Demo and Applications Session - Symposium on Databases*, 2012
- Aha D, Kibler D: Instance-based learning algorithms. *Mach Learn* 6:37–66, 1991
- Akgul C, Rubin D, Napel S, Beaulieu C, Greenspan H, Acar B: Content-Based Image Retrieval in Radiology: Current Status and Future Directions. *J Digit Imaging* 24:208–222, 2011
- Traina Jr. C, Traina A, Faloutsos A, Seeger B: Fast Indexing using Slim-Trees. *IEEE Trans Knowl Data Eng*, 2002
- Santos L, Bedo M, Ponciano-Silva M, Traina A, Traina Jr, C: Being Similar is Not Enough: How to Bridge Usability Gap Through Diversity in Medical Images. *Comput Based Med Syst*, 2014. doi:10.1109/CBMS
- Kinoshita S, Azevedo-Marques P, Pereira R, Rodrigues J, Rangayyan R: Content-based Retrieval of Mammograms Using Visual Features Related to Breast Density Patterns. *J Digit Imaging* 20:172–190, 2007
- Muramatsu C, Nishimura K, Oiwa M, Shiraiwa M, Endo T, Doi K, Fujita H: Correspondence among Subjective and Objective Similarities and Pathologic Types of Breast Masses on Digital Mammography. *Breast Imaging*, Springer, 2012.
- Traina Jr, C, Traina A, Araujo M, Bueno J, Chino F, Razente H, Azevedo-Marques P: Using an image-extended relational database to support content-based image retrieval in a PACS. *Comput Methods Prog Biomed* 80:71–83, 2005
- Kaster D, Bugatti P, Ponciano-Silva M, Traina A, Paulo M, Azevedo-Marques, Santos A, Traina Jr, C: MedFMI-SiR: A Powerful DBMS Solution for Large-Scale Medical Image Retrieval. *Inf Technol Biomed Med Inf*, 2011. doi:10.1007/978-3-642-23208-4_2
- Deserno T, Antani S, Long L: Ontology of Gaps in Content-Based Image Retrieval. *J Digit Imaging* 22:202–215, 2009

25. Naqa I, Yang Y: The Role of Content-Based Image Retrieval in Mammography CAD. *Comput Intell Biomed Imaging*, 2013. doi:[10.1007/978-1-4614-7245-2_2](https://doi.org/10.1007/978-1-4614-7245-2_2)
26. Town C: Content-Based and Similarity-Based Querying for Broad-Usage Medical Image Retrieval. *Stud Comput Intell*, 2013. doi:[10.1007/978-3-642-37137-0_8](https://doi.org/10.1007/978-3-642-37137-0_8)
27. Kumar A, Jinman K, Cai W, Fulham M, Feng D: Content-Based Medical Image Retrieval: A Survey of Applications to Multidimensional and Multimodality Data. *J Digit Imaging* 26: 1025–1039, 2013
28. Elter M, Hasslmeyer E: A knowledge-based approach to the CADx of mammographic masses. *Proc SPIE*, 2008. doi:[10.1117/12.770135](https://doi.org/10.1117/12.770135)
29. Bovis K and Singh S: Classification of Mammographic Breast Density Using a Combined Classifier Paradigm. *Int Work Digital Mammography*, 2002
30. Dietterich T: *Ensemble Methods in Machine Learning. Multiple Classifier Systems*. Springer, Berlin, GE, 2000
31. Mazurowski M, Lo J, Harrawood B, Tourassi G: Mutual information-based template matching scheme for detection of breast masses: From mammography to digital breast tomosynthesis. *J Biomed Inform* 44:815–823, 2011
32. Wei C, Li Y, Huang P: Mammogram Retrieval Through Machine Learning Within BI-RADS Standards. *J Biomed Inform* 44:607–614, 2011
33. Frenay B, Verleysen M: Using SVMs with randomised feature spaces: an extreme learning approach. *European Symposium on Artificial Neural Networks, Comput Intell Mach Learn*, 2010
34. Tao Y, Lo S, Hadjiski L, Chan H, Freedman M: BI-RADS guided mammographic mass retrieval. *Proc SPIE*, 2011. doi:[10.1117/12.878274](https://doi.org/10.1117/12.878274)
35. Huang G, Zhou H, Ding X, Zhang R: Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans Syst Man Cybern*, 2012. doi:[10.1109/TSMCB.2011](https://doi.org/10.1109/TSMCB.2011)
36. Liang N, Huang G, Saratchandran P, Sundararajan N: A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks. *IEEE Trans Neural Netw* 17:1411–1423, 2006
37. Huang G, Zhou H, Xiaojian D, Zhang R: Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Trans Syst Man Cybern B* 42:513–529, 2012
38. Yu L, Liu H: Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution, *Proceedings of the 20th International Conference on Machine Learning*, 2003
39. Haralick R, Shanmugam K, Dinstein I: Textural Features for Image Classification. *IEEE Trans Syst Man Cybern* 6:610–621, 1973
40. Daubechies I: *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992