Text-based Medical Image Retrieval using Convolutional Neural Network and Specific Medical Features

Nada Souissi, Hajer Ayadi and Mouna Torjmen-Khemakhem
Research Laboratory on Development and Control of Distributed Applications (ReDCAD), Department of Computer Science and Applied Mathematics, National School of Engineers of Sfax, University of Sfax, Tunisia

Keywords: Text-based Image Retrieval, Convolutional Neural Network, Specific Medical Image Features, Word2vec.

Abstract: With the proliferation of digital imaging data in hospitals, the amount of medical images is increasing rapidly. Thus, the need for efficient retrieval systems, to find relevant information from large medical datasets, becomes high. The Convolutional Neural Network (CNN)-based models have been proved to be effective in several areas including, for example, medical image retrieval. Moreover, the Text-Based Image Retrieval (TBIR) was successful in retrieving images with textual description. However, in TBIR, all queries and documents are processed without taking into account the influence of certain medical terminologies (Specific Medical Features (SMF)) on the retrieval performance. In this paper, we propose a re-ranking method using the CNN and the SMF for text-medical image retrieval. First, images (documents) and queries are indexed to specific medical image features. Second, the Word2vec tool is used to construct feature vectors for both documents and queries. These vectors are then integrated into a neural network process and a matching function is used to re-rank documents obtained initially by a classical retrieval model. To evaluate our approach, several experiments are carried out with Medical ImageCLEF datasets from 2009 to 2012. Results show that our proposed approach significantly enhances image retrieval performance compared to several state of the art models.

1 INTRODUCTION

The increasing amount of available medical images causes a difficulty in managing and querying these large databases. Thus, the need for systems providing efficient researches becomes high. However, few works investigate the impact of CNN-based models on the Text-Based Image Retrieval (TBIR) performance.

To improve the performance of the TBIR approach, authors (Ayadi et al., 2017a) and (Ayadi et al., 2018) proposed a thesaurus which is composed of a set of Specific Medical Features (SMF) such as image modality, image dimensionality and image color. In fact, the SMF have shown their effectiveness on medical query classification (Ayadi et al., 2013) and (Ayadi et al., 2017b) and medical image retrieval (Ayadi et al., 2017a) and (Ayadi et al., 2018). In this paper, we propose a new re-ranking model based on CNN and SMF (Ayadi et al., 2017b). Thus, the main contribution of this paper is the exploration of SMF in a CNN model (CSMF) for medical image re-ranking. In this work, we represent queries and documents as a set of SMF. We propose to use the popular Word2Vec model (Mikolov et al., 2013) to generate vector representations for SMF-based document and SMF-based queries. The resulting vectors are the input of the CSMF model, and are used to get a new semantic representation to improve the medical image retrieval accuracy.

The remainder of this paper is organized as follows: Section 2 describes the background of our work. Section 3 summarizes the related work. Section 4 describes the proposed CSMF model. Experiments are presented and discussed in Section 5. Finally, Section 6 concludes the paper and gives some perspectives.

2 BACKGROUND

In this section, we present the SMF set proposed in (Ayadi et al., 2013).

Authors in (Ayadi et al., 2017b) and (Ayadi et al., 2013) proposed SMF to predict the best retrieval model for a given query and to retrieve images (Ayadi et al., 2018). These features are manually defined by a

Souissi, N., Ayadi, H. and Torjmen-Khemakhem, M.
Text-based Medical Image Retrieval using Convolutional Neural Network and Specific Medical Features.
DOI: 10.5220/0007355400780087
In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2019)*, pages 78-87
ISBN: 978-989-758-353-7

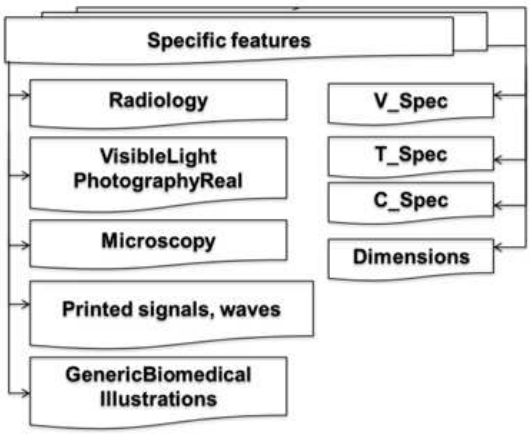

Figure 1: Specific Medical Features (Ayadi et al., 2013).

medical expert using imaging modalities and medical terminology. There are 25 features that are classified into 9 categories as illustrated in Figure 1.

- ”Radiology”: it represents the set of diagnostic and therapeutic modalities using radiation. It denotes ”Ultrasound”, ”Computerized Tomography”, ”Magnetic Resonance”, ”X-Ray”, ”2D Radiography”, ”PET”, ”Angiography” and ”Combined modalities” in one image. These modalities which ensure the provision of medical imagery, are chosen as values for the radiology feature.
- ”Visible light photography”: denotes the set of modalities that use visible light including ”Endoscopy”, ”Skin”, ”Dermatology” and ”Other organs”.
- ”Printed signals and waves”: combines ”electromyography”, ”electroencephalography” and ”electrocardiography”.
- ”Microscopy”: includes ”fluorescence microscopy”, ”transmission microscopy”, ”electron microscopy” and ”light microscopy”.
- ”Generic biomedical illustrations”: denotes, as ”modality tables and forms”, ”programs listing”, ”statistical figures”, ”graphs”, ”charts”, ”screen shots”, ”flowcharts”, ”system overviews”, ”gene sequences”, ”chromatography”, ”Gel”, ”chemical structure”, ”mathematics”, ”formulae”, ”nonclinical photos”, and ”hand-drawn sketches”.
- ”Dimensionality”: using only modality features to determine the best retrieval model is not suffcient. A medical textual query can be expressed without any image modality. However, in a medical query, the user can give information about the searched object dimension such as: ”micro”, ”gross” and ”gross-micro”.
- ”V-spec”: V-spec feature includes a feature related with to the searched image color. An example of V-spec is ”colored”.
- ”T-spec”: includes ”pathology” and ”finding” terms.
- ”C-spec”: includes ”Histology”, which means a study related to microscopic anatomy, so it interesting to applied both image content and its text description for queries containing this term.

3 RELATED WORK

In the litterature, several studies (Qiu et al., 2017) and (Bai et al., 2018) used CNN based model for Information Retrieval (IR) and medical image retrieval. This section briefly summarizes some of these approaches.

3.1 CNN for IR

In recent literature, the CNN is increasingly used in many disciplines such as IR (Tzelepi and Tefas, 2018), text classification (Kim, 2014), sentiment analysis (dos Santos and Gatti, 2014), etc. Thus, it is applied to several types of data such as text and images. For textual data, the CNN has shown the ability to: (1) automatically extract representations from input data and (2) effectively integrate the input sentences in vector spaces that keep the syntactic and semantic aspects of sentences.

Authors in (Huang et al., 2013) proposed a new semantic model based on CNN to enhance the web search performance by extracting semantic structures from queries or documents. In this model, the first layer converts vector of terms to vector of trigrams letters. The neuronal activities of the last layer form a projected vector representation to a semantic space. Finally, the CNN computes similarities of output vectors to evaluate the relevance scores of documents.

In (Shen et al., 2014), a CNN-based model was proposed. It transforms queries and documents to a set of n-gram words. So, the n-gram is projected in low-level feature vectors. Then, a max-pooling operation is applied to select neurons with highest activation values from word features. Finally, a non-linear transformation is performed to extract high level semantic information from sequence of input words. The parameters of the proposed model are learned using click through data. In (Severyn and Moschitti, 2015), a CNN architecture for re-ranking question-answer pairs was presented. Additional features have been integrated in this architecture to offer better performance. This CNN model was expanded and ana-

lyzed in (Rao et al., 2017) and delivered reproducible results with several implementations.

3.2 CNN for Medical Image Retrieval

The CNN models have been recently used for medical TBIR systems. In (Rios and Kavuluru, 2015), an approach based on bag of words was proposed. It used CNN to index biomedical articles by building binary text classifiers. In this model, the input is matrix of real numbers which represent the medical terms of the input document. Then, a succession of processing layers is done to classify the document. Another method for medical text classification that can be used for retrieval tasks was presented in (Hughes et al., 2017). In fact, it uses a bag of words training on a CNN to represent the semantics of an input sentence; especially it uses the Word2vec algorithm to represent the input medical sentences. Also, it keeps the stop-words during the training of the CNN model which is constituted by several convolutional layers, max-pooling and fully-connected layers. In (Soldaini et al., 2017), authors proposed a CNN to reduce noise in clinical notes to be used for medical literature retrieval. They used GloVe vectors (Pennington et al., 2014) to represent terms of input queries.

Despite the large number of works using CNN, there is a lack of studies using external semantic resources such as specific features to represent queries and documents. Therefore, we propose a new medical image re-ranking model based on CNN and SMF using Word2vec to improve retrieval accuracy.

4 A NEW CNN MODEL FOR TEXT-BASED MEDICAL IMAGE RETRIEVAL: CSMF MODEL

In this section, we explore the use of CNN for medical image retrieval. Our model, called CSMF, aims to re-rank medical images based on their textual description. The input of the CSMF model is a set of queries and documents indexed to a set of SMF (Ayadi et al., 2017b) as detailed in section II. The output of our model is a set of relevant documents to a given query. Our model is composed of several layers: (1) the input layer, which is a vector representing the query/document, (2) the convolutional layer, (3) the pooling layer and (4) the Fully Connected Layer (FCL) representing the output layer of the CSMF model. The output contains the scores of the similarity between query and documents.

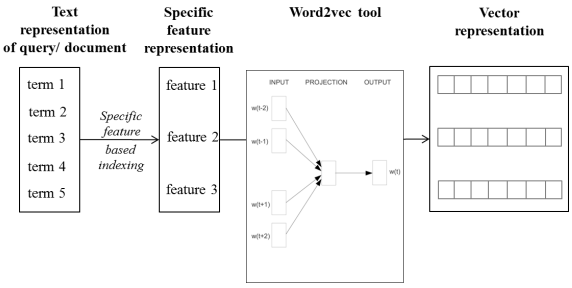

Figure 2: Transformation from text representation to vector representation.

4.1 Vector Representation of Queries and Documents

In this section, we detail the Word2vec method (Mikolov et al., 2013) for presenting queries/documents as vectors.

The input layer of the CSMF model is a query/document presented by features: $[f_1, ..., f_n]$, where each feature f_i is presented by a vector $V_i \in \mathbb{R}^d$ using the Word2vec tool. The obtained set of vectors are then concatenated to a matrix $S \in \mathbb{R}^{n \times d}$, where n is the number of the query (or document) features and d is the number of all features (in our case d=25 as mentioned in section II). Each vector V_i contains features representation using the Word2vec tool. For each input query/document, the matrix S is built. Each row i of S represents a feature f_i at the corresponding feature position i in the query/document.

Figure 2 shows an example of transforming a text representation to a vector representation according to the CSMF model. The queries/documents are represented as a set of SMF in order to extract semantic and specific features from the text representation. Finally, the Word2vec tool is used to transform each feature to a vector.

To capture semantic features in a given query/document and reach high level semantic information, the neural network applies a series of transformations to the input matrix S using convolution, non-linearity and pooling operations.

4.2 Convolutional Layer

In this layer, a set of filters $F \in \mathbb{R}^d$ are applied to the query/document vectors representation to produce different feature maps. Each feature map includes a level of semantic features extracted by the CNN. Each component of the feature map $c_k \in \mathbb{R}$ is computed by the following Equation:

$$c_k = \sum_{1}^{d} V_i F_i \quad (1)$$

Value 1	Value 2	Value 3	Value 4	…	Value 25
Feature 1 *x-ray*	**Feature 2** *mri*	**Feature 3** *ct*	**Feature 4** *micro*		**Feature 25** *skin*

Figure 3: Example of filter containing 25 values corresponding to a feature.

where V_i is the vector representing the query/document feature, F_i is a filter applied to the vector V_i, and d is the number of features. In our work, each filter contains 25 values where each value corresponds to a feature semantic degree as shown in Figure 3.

In the current work, we propose to use six filters initialized statistically as detailed below. In addition, the filters applied to the queries are initialized differently compared to the ones applied to the documents because the latter's size is greater than the queries' size. The query filter and the document filter are henceforth called (QF) and (DF), respectively.

4.2.1 Co-occurrence Filter (CoF)

(QF) The idea consists of calculating the co-occurrences of query features with all the terminology features.

$$CoF(QF) = \frac{\sum_0^d FR(FQi)}{\sum_0^d FR(FFj)} \quad (2)$$

Where $FR(FQ)$ is the frequency of query features and $FR(FF)$ is the frequency of features.

(DF) The document filter calculates the occurrence of the set of document features in the query. The more the document contains query features, the more it is relevant.

$$CoF(DF) = \sum_0^n FR(FD \in Q) \quad (3)$$

Where n is the number of the set of document features and $FR(FD \in Q)$ is the occurrence of document features in the query.

4.2.2 Length Filter (LF)

(QF) For each query, we compute the documents' size containing query features (SD). As normalization, we divide each obtained value by the highest sum of sizes (Max (SD)).

$$LF(QF) = \frac{\sum_0^n SD}{Max(SD)} \quad (4)$$

Where n is the number of the documents containing all query features.

(DF)For documents, we calculate the occurrence of the set of document features in the corresponding query (FD) and then we divide this value by the document length (LD). Indeed, if the document and the query share several features and the document has a small size, this document becomes more relevant.

$$LF(DF) = \frac{FD}{LD} \quad (5)$$

4.2.3 Rank Filter (RF)

(QF) We calculate documents' ranks (RD) containing query features. As normalization, we divide each obtained value by the highest rank.

$$RF(QF) = \frac{\sum_0^n RD}{Max(\sum_0^n RD)} \quad (6)$$

Where n is the number of documents containing all query features.

(DF) If the organization of features in a document is the same as in the query, the document should be organized.

$$RF(DF) = FR(FQ) \times Fact_{org} \quad (7)$$

Where $FR(FQ)$ is the frequency of query features in the document and $Fact_{org}$ is the organization factor of query in the document: $Fact_{org}$ equals 1 if the query preserves its organization in the document and 0.5 if not.

4.2.4 Proximity Filter (PF)

(QF) If a document contains query features, we compute the distances between its features (DD). Then, we divide each value by the biggest distance. In our case, the distance between two features is the number of features between them.

$$PF(QF) = \frac{\sum_0^n DD}{Max(\sum_0^n DD)} \quad (8)$$

Where n is the number of documents which contain query features.

(DF) The more the document's features existing in the query are closer, the more it is relevant.

$$PF(DF) = \frac{1}{|FD \in Q|} \quad (9)$$

Where $FD \in Q$ is the set of document features existing in the query.

4.2.5 PMI Filter (PMIF)

(QF/ DF) The PMI (Pointwise Mutual Information) (Church and Hanks, 1990) is a proposed metric to find features with a close meaning. Indeed, the PMI of features x and y is defined using the occurrences of x ($FR(x)$) and y ($FR(y)$), the co-occurrences $FR(x,y)$

within a vector of features, and N the collection size for QF and the document size for DF.

$$PMIF(QF) = log\frac{N \times FR(x,y)}{FR(x) \times FR(y)} \quad (10)$$

This equation calculates the semantically closest features of the collection to the features x and y.

4.2.6 Feature Difference Filter (FDF)

(QF) For each query, we compute the number of its different features comparing to the document ($DiffD$). Then, we divide this number by the maximum value.

$$FDF(QF) = \frac{\frac{1}{DiffD}}{Max(\frac{1}{DiffD})} \quad (11)$$

(DF) The more the number of the set of document features not belonging to the query is small, the more the document is relevant.

$$FDF(DF) = \frac{1}{|FD \notin FQ|} \quad (12)$$

Where FD is the set of document features and FQ is the query features.

4.2.7 Application of Filters

Given that the input of the SemRank model is a matrix $S \in \mathbb{R}^{n \times d}$, the convolutional filters are also matrices $F \in \mathbb{R}^{d}$. It should be noted that these filters have the same dimensionality d as the input matrix. Moreover, these filters scan the vectors representation producing a vector $C \in \mathbb{R}^{n}$ at the output. Each component c_i of C is the result of computing the product between a vector V and the filter F, which is summed to produce a single value.

$$c_i = \sum_{k=1}^{d} V_k F_k \quad (13)$$

As an example, Fig. 4 shows a matrix representation of the query "ct x-ray micro", as well as the six filters.

4.3 Activation Function

The convolutional layer is followed by a non-linear activation function α applied to the output of the preceding layer. This function allows transforming the input signal in a neuron to an output signal.

Several activation functions are proposed in the literature such as:

- Sigmoid (Norouzi et al., 2009) which is defined by:

$$\alpha(x) = \frac{1}{1 - e^{-\lambda x}} \quad (14)$$

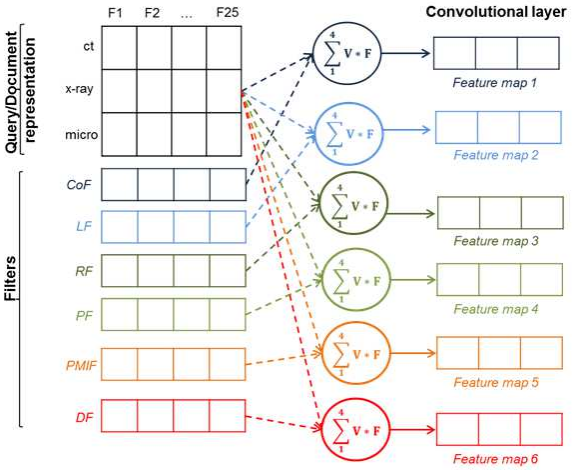

Figure 4: Example of convolutional layer for the query "ct x-ray micro".

where x is the input of a neuron and λ a parameter of the sigmoid function. Its name indicates in practice an S shape. It represents the logistic distribution function.

- Hyperbolic tangent (tanh) (Nguyen and Widrow, 1990) is an hyperbolic function defined by:

$$tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \quad (15)$$

where x is the input of a neuron.

- Rectified Linear Unit (ReLU) (Jarrett et al., 2009) which is defined by:

$$\alpha(x) = max(0,x) \quad (16)$$

where x is the input of a neuron.

The ReLU function ensures that neural values transmitted to the next layer are always positive. In fact, authors in (Nair and Hinton, 2010) showed that: the ReLU function is efficient, simple and allows to reduce complexity and calculation time. Hence, we use it as an activation function in our model.

4.4 Pooling Layer

The pooling layer aims to aggregate information, reduce representation and extract global features from local ones of convolutional layer. In the literature, two functions have been applied:

- Average: consists of computing the average of each feature map of the convolutional layer and storing it in the pooling layer. However, this method suffers from a major drawback: all elements of the input are considered even if many have low weights (Zeiler and Fergus, 2013).

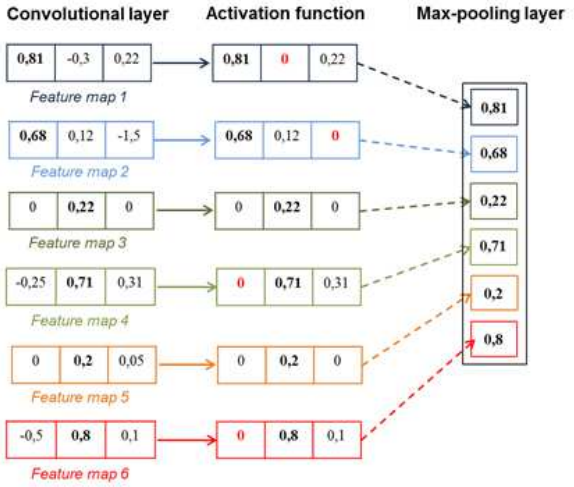

Figure 5: Example of a max-pooling layer.

- Max: consists of selecting the maximum value of each feature map of the convolutional layer. Thus, Max method only considers neurons with high values of activation which can lead to poor generalization of input data (Zeiler and Fergus, 2013).

While max pooling does not suffer from this drawback, we chose to use it as illustrated in Figure 5.

4.5 Fully Connected Layer

A Fully Connected Layer (FCL) is, then applied to the resulting vector, to obtain a final vector representation of the query/document. As our objective is only to interconnect all neurons together, we propose to initialize the weight vector to 1.

4.6 The Query/Document Matching Function

We compute the relevance score between queries and documents by calculating the cosine similarity between query vector representation $\overrightarrow{Q}$ and document vector representation $\overrightarrow{D}$. This relevance score is defined as follows:

$$RSV(Q,D) = S_{CSMF}(D) = cosine(\overrightarrow{Q},\overrightarrow{D}) = \frac{\overrightarrow{Q}\,\overrightarrow{D}}{\left\|\overrightarrow{Q}\right\|\left\|\overrightarrow{D}\right\|} \quad (17)$$

Finally, we combine the CSMF scores (S_{CSMF}) with Baseline model scores ($S_{Baseline}$) using a linear combination:

$$S_{combination}(d_i) = \alpha \times S_{Baseline}(d_i) + (1-\alpha) \times S_{CSMF}(d_i) \quad (18)$$

where α is a parameter ($\alpha \in [0..1]$) and d_i is a document retrieved by the Baseline model.

As a baseline we propose to use the well known probablistic model BM25 model (Robertson et al, 1994).

5 EXPERIMENTS

In this section, we first describe the datasets and the evaluation metrics. Then, we present the baseline approach which is BM25. Finally, we discuss the experimental results by presenting a comparative study with BM25, DLM and Bo1PRF models.

5.1 Datasets and Evaluation Metrics

To evaluate the proposed CSMF model, we conducted experiments using medical ImageCLEF datasets from 2009 to 2012 (Dimitrovski et al., 2009), (Benavent et al., 2010), (Kalpathy-Cramer et al., 2011) and (Müller et al., 2012)). Each image in the collection has a textual description presented in semi structured format including an identifier, an URL, a caption, a title, etc. These ImageCLEF collections are presented in Table 1. We note that each query is composed of a text representation and few sample images. In our work, we use only textual representations of the queries. We note that ImageCLEF 2011 and 2012 datasets contain a greater image diversity and also include charts, graphs and other, similar, non-clinical images (Ayadi et al., 2013).

We note that the size of the collection of ImageCLEF 2011 and 2012 has been significantly increased. Indeed, these datasets contain a greater image diversity and also include charts, graphs and other, similar, non-clinical images (Ayadi et al., 2013).

In our experiments, we propose to use two metrics in the evaluation process: the Precision at k documents (P@K) and the Mean average precision (MAP).

5.2 CSMF_BM25 Model Results

We propose to combine the scores obtained by the CSMF model with those obtained by the BM25 model to improve medical image retrieval accuracy. So, we conduct a set of experiments. Consequently, we obtain a new model called CSMF_BM25 model. In fact, α = 0 means that only the CSMF score is used and α = 1 means that only the BM25 score is used.

Figure 6 shows that the combination of scores obtained by the baseline model and those obtained by the CSMF model improves the results compared to the baseline. According to MAP measures, there are improvements of: 7% in the ImageCLEF 2009 when

Table 1: Statistics of ImageCLEF datasets.

	2009	2010	2011	2012
Total number of images	74902	77500	231000	306528
Number of queries	25	16	30	22

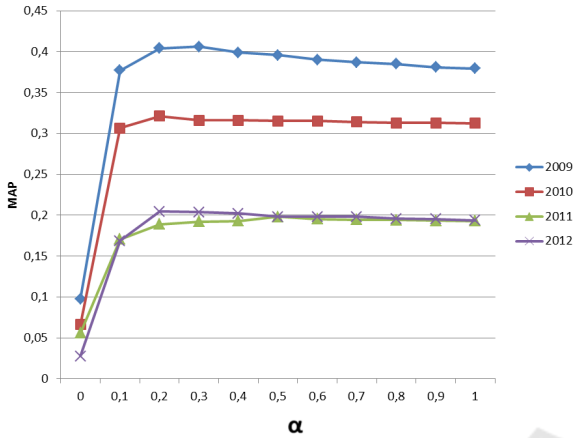

Figure 6: MAP according to α of CSMF_BM25 model in ImageCLEF datasets.

α = 0.3, 2% in the ImageCLEF 2010 when α= 0.2, 2% in the ImageCLEF 2011 when α = 0.5 and 5% in the ImageCLEF 2012 when α = 0.3 compared to the baseline.

We notice that best results are obtained when α ∈ [0.1..0.5]. Therefore, we chose to set α = 0.3 in the remaining experiments.

To compare the CSMF_BM25 model with the BM25 one, we determine the improvement rate and we conducted a statistic significance test. The significance value $p \in [0..1]$ estimates the probability that the difference between two methods is due to randomness. The difference is considered statistically significant if $p < 0.05$ (Hull, 1993). In this paper, the results are followed by the * when $p < 0.05$. According to Table 2, we note that the improvement obtained by the CSMF_BM25 model is statistically significant compared to the BM25 model for 2009 and 2012 ImageCLEF collections ($p < 0.05$).

5.3 Comparison between CSMF_BM25 and Some Literature Models

In this section, we propose to compare our proposed model with DLM and Bo1PRF models according to P@5, P@10 and MAP measures. The DLM (Dirichlet Language Model) (Yu et al., 2005) is a statistic model that allows modeling the arrangement of words in a language, capturing the distribution of words and measuring the probability of observing a sequence of words. The purpose of the Bo1 PRF (Bo1 pseudo relevance feedback) (Lioma and Ounis, 2008) is to consider the relevance jugement of users on the documents obtained initially.

Table 3 summarizes the comparison of the CSMF_BM25 model with the DLM and the Bo1PRF models. The best result across all models and for each metric is presented in bold. Our model outperforms other models significantly and reached between 9% and 24% on the 2009 dataset. For 2010 dataset, the CSMF_BM25 model improves the retrieval performance compared to DLM and Bo1PRF models. This could be explained by the fact that 2009 and 2010 datasets contain images proposed by clinicians and physicians answering the information needed.

For the 2009 and 2010 datasets, the combination of BM25 and CSMF improves the results. For the 2011 and 2012 datasets, the results are reduced compared to the baseline.

First, we observe that the CSMF_BM25 model outperforms the BM25 model with a substantial margin from 1% to 7% in MAP for the 2009, 2010 and 2012 datasets. Our model also outperforms DLM model with a statistically significant margin from 1% to 39% for different datasets. Further, compared to PRF model, the CSMF_BM25 model shows a significant improvement of 9% and 4% MAP respectively for the 2009 and 2010 datasets. For the 2011 and the 2012 datasets, however, no significant gain is observed. This can be explained that these datasets contain a diversity of images types (tables, shapes, graphs ...). Moreover, the Bo1 PRF model is based on the relevance feedback technique that improves retrieval results.

The accuracy gain is presented in Table 4.

First, we observe that the CSMF_BM25 model outperforms the BM25 model with a substantial margin from 1% to 7% in MAP for the 2009, 2010 and 2012 datasets. Our model also outperforms DLM model with a statistically significant margin from 1% to 39% for different datasets. Further, compared to PRF model, the CSMF_BM25 model shows a significant improvement of 9% and 4% MAP respectively for the 2009 and 2010 datasets. For the 2011 and the 2012 datasets, however, no significant gain is observed. This can be explained that these datasets contain a diversity of images types (tables, shapes, graphs ...). Moreover, the Bo1 PRF model is based on the rele-

Table 2: Comparison between CSMF_BM25 and BM25 according to MAP values.

ImageCLEF datasets				
	2009	**2010**	**2011**	**2012**
BM25	0.379	0.312	0.193	0.193
CSMF	0.097	0.066	0.055	0.027
CSMF_BM25 (α=0.3)	0.405 (+7%*)	0.316 (+1%)	0.190 (-)	0.203 (+5%*)

Table 3: Comparative results CSMF_BM25 with some literature models.

		DLM	**Bo1PRF**	**CSMF_BM25 (α=0.3)**
2009	**P@5**	0.592	0.608	**0.688**
	P@10	0.524	0.568	**0.664**
	MAP	0.327	0.371	**0.405**
2010	**P@5**	**0.436**	0.361	0.413
	P@10	0.375	0.330	**0.460**
	MAP	0.313	0.305	**0.316**
2011	**P@5**	0.240	0.386	**0.406**
	P@10	0.223	0.326	**0.330**
	MAP	0.138	**0.211**	0.192
2012	**P@5**	0.281	**0.554**	0.436
	P@10	0.240	**0.409**	0.336
	MAP	0.146	**0.361**	0.203

Table 4: Accuracy gain of the CSMF_BM25 compared to other models.

	2009	**2010**	**2011**	**2012**
CSMF_BM25/ BM25	+7% (*)	+1%	-	+5% (*)
CSMF_BM25/ DLM	+24% (*)	+1%	+38% (*)	+39% (*)
CSMF_BM25/ Bo1PRF	+9%	+4%	-	-

vance feedback technique that improves retrieval results.

To evaluate how well our proposed approach performs compared to the state of the art approaches (Hersh et al., 2009), (Popescu et al., 2010), (Kalpathy-Cramer et al., 2011) and (Müller et al., 2012), we further compared our approach with those of the four teams that achieved the best MAP using textual runs for the medical image retrieval tasks from 2009 to 2012 which are:

- LIRIS (France)
- SINAI (Spain)
- YORK (Canada)
- ISSR (Egypt)
- XRCE (France)
- AUEB (Greece)
- OHSU (USA)
- LABERINTI (Spain)
- UNED (Spain)
- IPL (Greece)
- MRIM (France)
- BIOINGENIUM (Colombia)
- BUAA AUDR (China)
- DEMIR (Turkey)

Table 5 lists the MAP, and P@10 values of our model and those of the state of the art approaches. These evaluation measures are the most commonly used measures for ranking participant runs in the ImageCLEFmed competition from 2009 to 2012. The results of our approach are comparable to the state of the art approaches. We first observe that the CSMF BM25 model gives the best result in terms of P@10 for the 2009 dataset. For the same dataset, our model does not outperform the highest values of MAP obtained by existing ImageCLEFmed approaches. However, it was the second best approach with a MAP of 0.405.

Table 5: Comparative results with the official submissions of the clef medical image retrieval track.

ImageCLEF 2009			**ImageCLEF 2010**		
Group	**MAP**	**P@10**	**Group**	**MAP**	**P@10**
LIRIS	0.430	0.660	XRCE	0.338	0.506
CSMF_BM25	**0.405**	**0.664**	AUEB	0.323	0.648
SINAI	0.380	0.620	**CSMF_BM25**	**0.316**	**0.460**
YORK	0.370	0.600	OHSU	0.302	0.431
ISSR	0.350	0.560	SINAI	0.276	0.425
ImageCLEF 2011			**ImageCLEF 2012**		
Group	**MAP**	**P@10**	**Group**	**MAP**	**P@10**
LABERINTI	0.217	0.346	BIOINGENIUM	0.218	0.340
UNED	0.215	0.353	BUAA AUDR	0.208	0.309
IPL	0.215	0.403	**CSMF_BM25**	**0.203**	**0.336**
MRIM	0.200	0.303	IPL	0.200	0.295
CSMF_BM25	**0.192**	**0.330**	DEMIR	0.190	0.331

In ImageCLEF 2011, no outperformance is shown.

We conclude that integrating SMF in a CNN improves results comparing to the baseline and other models. This could be limited to the SMF that are purely medical.

6 CONCLUSION AND FUTURE WORK

We proposed in this paper a novel CNN model for re-ranking medical images based on Specific Medical image Features (SMF) called CSMF. In this model, queries and documents are represented as a set of SMF. The Word2vec method is used to construct vector representations for each query/document. The resulting vectors are then integrated into a CNN process. The output is a query vector and a document vector used to calculate new relevance scores for documents given a query. A linear combination of obtained scores with baseline scores is then used.

We carried out experiments using the Medical ImageCLEF collections from 2009 to 2012. The results showed that the combination of CSMF scores and baseline scores improves the retrieval accuracy. In addition, we compared our model with other state of the art models and we noticed a significant improvement in the most of metrics' values.

In future work, we plan to use CSMF model as a ranking model by applying the deep learning technique on the CNN for updating the filter values of this model. Furthermore, we plan to integrate visual features in the CSMF model and combine them with textual features to improve retrieval accuracy.

REFERENCES

Ayadi, H., Khemakhem, M. T., Huang, J. X., Daoud, M., and Jemaa, M. B. (2017a). Learning to re-rank medical images using a bayesian network-based thesaurus. In *European Conference on Information Retrieval*, pages 160–172. Springer.

Ayadi, H., Torjmen, M., Daoud, M., Ben Jemaa, M., and Xiangji Huang, J. (2013). Correlating medical-dependent query features with image retrieval models using association rules. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 299–308. ACM.

Ayadi, H., Torjmen-Khemakhem, M., Daoud, M., Huang, J. X., and Ben Jemaa, M. (2017b). Mining correlations between medically dependent features and image retrieval models for query classification. *Journal of the Association for Information Science and Technology*, 68(5):1323–1334.

Ayadi, H., Torjmen-Khemakhem, M., Daoud, M., Huang, J. X., and Ben Jemaa, M. (2018). Mf-re-rank: A modality feature-based re-ranking model for medical image retrieval. *Journal of the Association for Information Science and Technology*, 69(9):1095–1108.

Bai, C., Huang, L., Pan, X., Zheng, J., and Chen, S. (2018). Optimization of deep convolutional neural network for large scale image retrieval. *Neurocomputing*, 303:60–67.

Benavent, J., Benavent, X., de Ves, E., Granados, R., and García-Serrano, A. (2010). Experiences at imageclef 2010 using cbir and tbir mixing information approaches. In *CLEF (Notebook Papers/LABs/Workshops)*.

Church, K. W. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational linguistics*, 16(1):22–29.

Dimitrovski, I., Kocev, D., Loskovska, S., and Džeroski, S. (2009). Imageclef 2009 medical image annotation task: Pcts for hierarchical multi-label classification. In *Workshop of the Cross-Language Evaluation Forum for European Languages*, pages 231–238. Springer.

dos Santos, C. and Gatti, M. (2014). Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 69–78.

Hersh, W., Müller, H., and Kalpathy-Cramer, J. (2009). The imageclefmed medical image retrieval task test collection. *Journal of Digital Imaging*, 22(6):648.

Huang, P.-S., He, X., Gao, J., Deng, L., Acero, A., and Heck, L. (2013). Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 2333–2338. ACM.

Hughes, M., Li, I., Kotoulas, S., and Suzumura, T. (2017). Medical text classification using convolutional neural networks. *Stud Health Technol Inform*, 235:246–50.

Hull, D. (1993). Using statistical testing in the evaluation of retrieval experiments. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 329–338. ACM.

Jarrett, K., Kavukcuoglu, K., LeCun, Y., et al. (2009). What is the best multi-stage architecture for object recognition? In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2146–2153. IEEE.

Kalpathy-Cramer, J., Müller, H., Bedrick, S., Eggel, I., de Herrera, A. G. S., and Tsikrika, T. (2011). Overview of the clef 2011 medical image classification and retrieval tasks. In *CLEF (notebook papers/labs/workshop)*, pages 97–112.

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *arXiv preprint arXiv:1408.Conference on Empirical Methods in Natural Language Processing*.

Lioma, C. and Ounis, I. (2008). A syntactically-based query reformulation technique for information retrieval. *Information processing & management*, 44(1):143–162.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *ICLR Workshop*.

Müller, H., de Herrera, A. G. S., Kalpathy-Cramer, J., Demner-Fushman, D., Antani, S. K., and Eggel, I. (2012). Overview of the imageclef 2012 medical image retrieval and classification tasks. In *CLEF (online working notes/labs/workshop)*, pages 1–16.

Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814.

Nguyen, D. and Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *Neural Networks, 1990., 1990 IJCNN International Joint Conference on*, pages 21–26. IEEE.

Norouzi, M., Ranjbar, M., and Mori, G. (2009). Stacks of convolutional restricted boltzmann machines for shift-invariant feature learning. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2735–2742. IEEE.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Popescu, A., Tsikrika, T., and Kludas, J. (2010). Overview of the wikipedia retrieval task at imageclef 2010. In *CLEF (notebook papers/LABs/workshops)*.

Qiu, C., Cai, Y., Gao, X., and Cui, Y. (2017). Medical image retrieval based on the deep convolution network and hash coding. In *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017 10th International Congress on*, pages 1–6. IEEE.

Rao, J., He, H., and Lin, J. (2017). Experiments with convolutional neural network models for answer selection. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1217–1220. ACM.

Rios, A. and Kavuluru, R. (2015). Convolutional neural networks for biomedical text classification: application in indexing biomedical articles. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 258–267. ACM.

Robertson, S. E., Walker, S. (1994). Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. Springer-Verlag New York, Inc., pp. 232–241.

Severyn, A. and Moschitti, A. (2015). Learning to rank short text pairs with convolutional deep neural networks. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, pages 373–382. ACM.

Shen, Y., He, X., Gao, J., Deng, L., and Mesnil, G. (2014). A latent semantic model with convolutional-pooling structure for information retrieval. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 101–110. ACM.

Soldaini, L., Yates, A., and Goharian, N. (2017). Denoising clinical notes for medical literature retrieval with convolutional neural model. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 2307–2310. ACM.

Tzelepi, M. and Tefas, A. (2018). Deep convolutional image retrieval: A general framework. *Signal Processing: Image Communication*, 63:30–43.

Yu, G., Li, X., Bao, Y., and Wang, D. (2005). Evaluating document-to-document relevance based on document language model: modeling, implementation and performance evaluation. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 593–603. Springer.

Zeiler, M. D. and Fergus, R. (2013). Stochastic pooling for regularization of deep convolutional neural networks.