International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V69I9P223
Volume 69 Issue 9, 193-202, September, 2021

Hybrid Ontology-Deep Learning Integrated CBMIR System for CT Lung Diseases

Nora yahia Ibrahim[1], Amira samy Talaat [2], Hany M. Harb[3]

[1,2]*Computers and systems Department, Electronics Research Institute, Egypt*
[3]*Computers and Systems Engineering Department, Faculty of Engineering, Misr University for Science and technology, Egypt*

[1]nora@eri.sci.eg

Abstract *— Recently, radiologists and general practitioners diagnose abnormal CT (Computed Tomography) lung images based on the CBMIR (Content-Based Medical Image Retrieval) system. This system idea is to extract the image feature and retrieve the relevant CT lung images that match a query image. The semantic gap between low-level visual content images and conceptual high-level semantic knowledge is a challenge in the CBMIR system. Most of the conventional CBMIR system works on the details of the images and doesn't exploit the metadata, for example, tags, image description, etc. This paper tackles this problem by using the metadata that describes the contents of images. This paper utilizes the high-level knowledge representation structure called ontology that represents the semantic annotations extracted from metadata to improve the classification accuracy of the CBMIR. Two stages are proposed to retrieve medical images: the classification and the retrieval stages. In the classification stage, the images are trained using a Deep Convolutional Neural Network (DCNN), and then the trained model is used to classify the CT lung images. However, the images are retrieved from the predicted class according to a query image in the retrieval stage. In addition, it retrieves similar images from the whole database without incorporating class prediction. The proposed method is evaluated in retrieving common CT Imaging Signs (CISs), which are vital for diagnosing lung diseases. The proposed method achieved an accuracy of 95.2 for the classification task and a mean average precision (mAP) of 0.742 for the retrieval task.*

Keywords *— ontology, deep learning, semantic, image retrieval, lung diseases.*

I. INTRODUCTION

Many hospitals and clinics are generating enormous medical images with various medical imaging types such as CT, Magnetic Resonance Imaging (MRI), X-ray, and many more, which cause an extensive collection of images. A CT scan for a patient consists of few hundred slices (images). However, handling such a vast image set manually is a tedious task and labor-consuming concerning time consumption and efficient interpretation or even tricky for radiologists because radiologist's decision is based on (CISs) in CT lung diseases. These signs can be classified into different forms without correlating them to a specific disease because the same signs can present in different diseases [1]. CISs are divided into nine categories which are essential for the diagnosis of lung diseases, including ground-glass opacity (GGO), pleural indentation (PI), calcification, lobulation, speculation, air bronchogram (AB), cavity & vacuoles (C&V), bronchial mucus plugs (BMP), and obstructive pneumonia (OP)[2].

CBMIR system plays a vital role in this challenge by retrieving CT images that match a query image and helps the physician diagnose the abnormal lung tissues in CT images accurately.

The idea of CBMIR is to use a suitable distance to discover the best matches corresponding to feature space with a given query image [3]. In the CBIR system, the image is represented by a set of visual representations such as shape, texture, color [4], etc. Although these representations are powerful for describing the image in an automated process, they are often not precise enough in capturing the visual appearance of images. Therefore, the CBMIR system will not map human concepts into image features (i.e., form a semantic gap between the low-level visual representations of image and high-level human-understandable concepts). In order to minimize this gap, the semantic analysis must be integrated into the CBMIR system to improve its performance. Although the metadata-based search is inadequate when dealing with visual content, it describes the image content more accurately. However, most conventional CBMIR doesn't exploit the advantage of the metadata. This paper tackles this problem by integrating two techniques to merge the advantages of semantic and visual content.

The semantic web, such as ontology, presents new insights into image retrieval because it attempts to extract semantic annotations from metadata to describe the contents of images in a unified description base.

Ontologies are used in image retrieval systems to acquire semantic knowledge (i.e., mapping image features to concepts); and provides relevant and more appropriate results to a user's query rather than traditional keyword searching.

This paper exploits the semantic annotations extracted from metadata to construct an ontology that describes the ROI of images. The constructed ontology is integrated with DCNN to propose a new technique that accurately retrieves the CISs lung diseases. The proposed algorithm consists of two stages: stage one trains the VGG-16 model based on merging the features extracted from the described metadata and the features extracted from the CISs, the class of query image is predicted using the trained VGG-16. The second stage retrieves the images from the predicted class according to a query image. In addition, the system compares the visual features extracted from the query image and the database images to retrieve similar images without incorporating class prediction.

This paper is divided as follows. The related work is presented in section 2. Section 3 discusses an explanation of the proposed CBMIR model. Section 4 includes experimental results. Finally, the conclusion is presented in section 5.

II. RELATED WORK

According to the previous research, some techniques used the CBIR framework based on ontology in general-purpose applications and specific domains, such as medical sciences. On the other hand, other techniques apply different machine learning techniques to medical images. Therefore, this section concentrates on the conventional techniques which are applied to medical images.

A. The approaches applied ontologies to the biomedical domain

Daniela S. and Raicu [5] proposed a computer-assisted architecture focused on the automated discovery of image-semantic mappings. These mappings are based on Probabilistic models between the low-level image features (shape, size, gray-level intensity, and texture) and the semantic lung nodule characteristics (lobulation, malignancy, margin, sphericity, speculation, subtlety, and texture) that are associated with the visual appearance of the nodules. The acquisition of the mappings is utilized to develop a visual ontology for lung interpretation to annotate and recover the input image and provide a context-sensitive retrieval system of the pulmonary nodules.

Johannes and Georg Langs [6] presented a strategy that visual remapping features extracted from medical images data depend on weak labels that can be found in corresponding radiology reports defining descriptions of local image contents to capture the relevant information. The extraction and identification of relevant terms in radiology reports are not in the scope of this paper.

Matthew S. Simpson [7] introduced a technique that maps lesions in CT images to the terminal in the RadLex ontology by using natural language and image processing to create a new visual ontology of biomedical finding imaging entities, which are contained in biomedical articles and pairing these entities with textual descriptions taken from their associated captions. The labels of these entities are used to train a classifier for providing concepts to ROI that does not contain associated text. The retrieval process of these types of images is not in the scope of this paper.

Camille Kurtz [8] proposed a semantic image retrieval system using an ontology that describes the content of medical images of the liver. The retrieval method is based on the similarity between images annotations using a dissimilarity measure.

Yang Chen and Xiaofeng Ren [19] merged a visual object recognition algorithm with a medical ontology for mining web images and retrieving many diseases automatically with minimum manual labeling work.

B. The approaches applied ontologies to the general domain

Jalila Filali and Hajer Zghal [10] combined the ontology and Hierarchical Max-pooling techniques to classify 22 different animal classes from ImageNet. In order to improve image classification accuracy, they constructed an ontology that represents the classes' domain. The outputs of the hypernym-hyponym ontological relationships are utilized to discriminate between the classes.

Jalila Filali and Hajer Baazaoui [11] developed an image retrieval system based on integrating low-level visual features and semantic features. The low-level features are extracted using BOVW (Bag of visual words) based on the SIFT technique. However, the concepts and relationships derived from the image annotations are used to build ontologies that represent the semantic features. Finally, these ontologies are used in the retrieval process.

Umar Manzoor and Mohammed A. Balubaid [12] presented an ontology-based image retrieval system that enables users to enter images or keywords as text. They used a hybrid strategy and utilized classification approaches based on shape, color, and texture.

C. The approaches applied learning techniques to the biomedical domain

R. Biswas, S. Roy [13] presented the CBMIR system to extract and detect lung nodules in CT images. The gray level co-occurrence matrix (GLCM) and shape features are used for feature extraction by applying Fast Discrete Curvelet Transform (FDCT). The authors used the selection feature approach to select the limited features using the Enhanced Moth Flame Optimization algorithm. Finally, they applied a deep learning algorithm to detect the various types of lung nodules.

M. Kashif, Gulistan Raj [14] presented a group of descriptors such as local ternary pattern, local phase quantization, and discrete wavelet transform to extract the visual features of images. After the features are extracted, a feature selection strategy is applied to select the limited features. The selected features are used to apply the different distance metric similarities such as Euclidean, city block, spearman, and cosine distances between the database images and query image of CT lung image. However, the semantic features are extracted from images using the support vector machine (SVM) classifier. Then, both semantic and visual similarities are combined in a created weighted balanced graph applying the shortest path technique to get the final similarity.

Ling Ma, Xiabi Liu [15] proposed a fused context-sensitive similarity (FCSS) method of CBMIR based on the SVM technique to classify the extracted semantic features. On the other hand, the visual features from CT lung images are extracted using the textures retrieval method. The cosine similarity matches the extracted visual features of database images with the query image features. Using the shortest path algorithm, they constructed a weighted graph whose nodes represent the images and edges to measure their pairwise similarity (semantic and visual similarities). However, the shortest path

computation consumes a lot of the running time because they did not cluster the datasets.

R. Biswas, S. Roy, A. Biswas [16] presented a Triplet-CBMIR framework that extracted low-level features such as shape and texture and semantic features with relevant feedback techniques using deep learning techniques. Then, the authors remove the irrelevant features by applying the mutual information-based Neighbourhood Entropy algorithm for feature selection.

III. THE ONTOLOGY-CNN BASED INTEGRATED METHODOLOGY

This section describes the proposed ontology and CNN-based integrated CBMIR framework, which is composed of two stages: the classification and the retrieval stages. In the classification stage, the proposed CNN model is trained using visual and semantic representations. VGG-16 model is used to extract the visual features of medical images. On the other hand, the semantic annotations are extracted from metadata to build an ontology that is converted to a features vector using the Term Frequency-Inverse Document Frequency (TF-IDF) word embedding technique [17]. All these features are concatenated to train the proposed VGG-16 model. In the retrieval stage, the features are extracted from the query image, and the class is predicted using the trained VGG-16. According to the predicted class, the proposed system retrieves all the similar images which belong to the query image. The similarity distance between the features of the query image and images in the whole database is also evaluated to retrieve the matched images to the query image.

A. The Classification stage

In this stage, the proposed algorithm is trained on the CISs dataset to classify these signs into different diseases. The images and annotations metadata will be passed through three modules for the training process to perform a specific task, as shown in Fig.1.

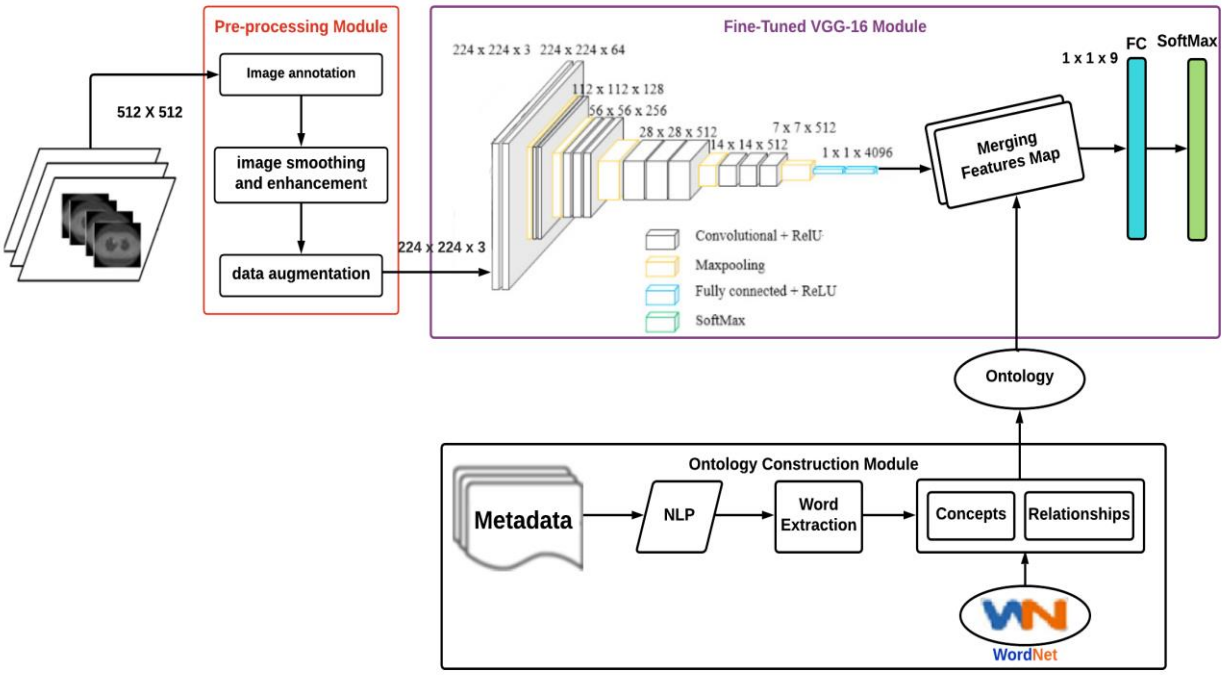


Fig 1: The proposed training process of classification stage for CT lung diseases

a) Image pre-processing module: Before the feature extraction process, the data is prepared and cleaned for training the model. The pre-processing image module is composed of three steps, as described in the following subsections.

1) Annotation of training images: In deep learning, image annotation is necessary to clarify the images with labels. In order to train the VGG-16 model with the dataset, it is necessary to annotate the dataset in VGG-16 format. Our dataset images are transformed from DICOM to JPG format. We resize JPG images from 512*512 to 224*224 and generate a label file for each image.

2) Image smoothing and enhancement: Two main steps are applied to the dataset to increase the numbers of images for specific classes and balance the images: image smoothing and image enhancement techniques. Image smoothing removes unwanted noise based on the Median filter without blurring the image. Contrast adjustment enhances the contrast of images to become saturated at the law and high intensity of the input images.

3) Augmentation of training images: Another common pre-processing technique generates different training images to artificially increase their amounts and expose the neural network to a broad range of different input images. In the proposed method, we apply the following augmentations.

- Shifting: with a shift range of 0.2 for width and height.

- Zooming: with a zoom range of 0.2, then the range will be [0.8, 1.2], or between 80% zoom in and 120% zoom out.
- Shearing: with a shear range of 0.2.
- Flipping: horizontally.
- Fill mode: points outside of the image boundaries are filled by default with "nearest".

b) Ontology construction module: ontology is a formal conceptually knowledge representation in a certain domain and its relationships. Ontologies can be created in two ways, generic and domain-dependent. CYC [18] and WordNet [19] [20] are examples of generic ontologies. Domain Ontology is specific for specific domains. However, the generic ontology defines concepts that are generic across many fields [12]. This paper utilizes the domain-dependent visual concept ontology strategy to provide a vocabulary for the visual description of domain classes based on image processing techniques. Some of the visual concept ontology is described in [21], [5]. In our medical domain, an ontology is created to provide controlled terminology for describing and interpreting the contents of CT medical imaging, which requires knowledge of the ROI.

As clarified in Fig. 1, the input metadata that we collected from [2] in the text file form is used to extract the semantic concepts and relationships to build a visual ontology that describes images' ROI. Three steps are used to achieve this goal: Three steps are used to achieve this goal: the first step, the TF-IDF word embedding technique, is utilized to extract the words from textual information and perform morphological and semantic analysis. TF means the number of times that the word occurs in the document, While IDF means maximizing the importance of words that occur rarely and minimizing the importance of words that occur very frequently in the document collection. The second step, the filtering of words, is performed by removing words that are not defined by the WordNet dictionary. The third step is to capture the meanings, contexts, and semantic and taxonomic relationships of words to get the lemmatized words (morphologic of word and derivation of the word to a common base form) in a vector form that is ready to build our ontology. The constructed ontology contains a vector of each concept, and it is a label name. This vector is a set of keywords and their weights. These weight vectors are merged with the training images features as discussed in the next section in detail.

c) Fined-tuned VGG-16 classifier module: This module is used to train the model for classifying the CT lung images, as shown in figure1. VGG-16 is a pre-trained deep learning model that contains 41 layers that have been trained on ImageNet datasets. VGG-16 has been trained on over a million images and is capable of classifying images into1000 different object sectors. This paper used the pre-trained VGG-16 network and fine-tuned its weights using the pre-processed images from annotation and augmentation processes to represent low-level visual information. In parallel, the features vectors from textual information are extracted using the word embedding technique as discussed in section 3.1.2 (ontology construction module). These features are merged to the extracted features of database images that are extracted from the fully connected (FC) layer (layer 39). The network was fine-tuned for the classification process by replacing the last layer of VGG-16 (softmax layer) with three layers: a merging layer (layer 40), a new fully-connected layer, and a softmax layer (final layer of our fine-tuned VGG-16), as shown in Fig. 2. The new merging layer takes the output of layer number 39 (input images features) and the output of the ontology construction module. The output of the merging layer is fed to the new fully connected layer for classification. The final softmax layer of a new fine-tuned VGG-16 takes the output of the new fully connected layer as an input with the number of our classes.

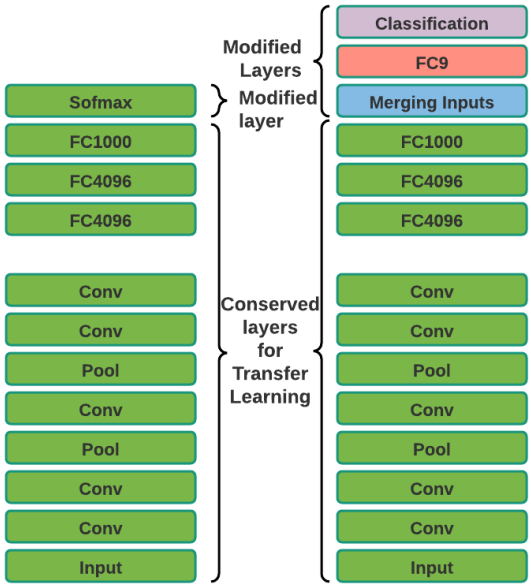


Fig. 2 The proposed VGG-16 model

B. Retrieval stage

For the query image, the features are extracted by passing query image from trained VGG-16; the visual information is extracted at layer 39. Then the class is predicted to which the query image belongs using the trained VGG-16. Finally, the similar images in that predicted class are retrieved by matching the query features and the images features of the predicted class using similarity distance, as shown in Fig. 3. In addition, it retrieves similar images from the whole database without incorporating class prediction by comparing visual information extracted from a query image and database images.

In our context, we apply the cosine similarity distance Eq. (1) to compute the similarity between the query image features and database images features then the relevant images are retrieved, which are matched with the query image. In addition, the classification results (predicted class) are used in the retrieval stage by applying the

similarity metrics between the query image features and images features in that predicted class. This strategy limits the search area in the database by decreasing the number of computations and eliminating irrelevant images from retrieval results. For ranking the results, the images with the shortest distance or high match compared to others are shown as leaders in the retrieval results.

$$D(A,B) = 1 - \frac{\sum_{j=1}^{m} X_j Y_j}{\sqrt{\sum_{j=1}^{m} X_j^2}\ \sqrt{\sum_{j=1}^{m} Y_j^2}} \quad (1)$$

Where m represents the number of samples in the training set while the query and database images features are represented by Xj and Yj, respectively.

A. *Dataset*

For evaluating the proposed ontology and CNN-based image retrieval method, LISS datasets of CISs [2] are used. Fig. 4 shows a sample of nine categories of CISs datasets. This dataset contains the CT imaging in DICOM format with an image resolution of 512*512 pixels. In addition, 511 rectangular ROIs are manually marked and annotated by experienced radiologists.

B. *Implementation details*

The input dataset is split into 80:20, 80% for training, 10% for testing, and the same ratio to validate the model to evaluate the proposed method. As shown in Fig.1, in the first step of the training model, the DICOM images are converted to JPG format and annotated in VGG-16 format. Then all the images are resized from 512*512 to 224*224.

IV. EXPERIMENTAL RESULTS

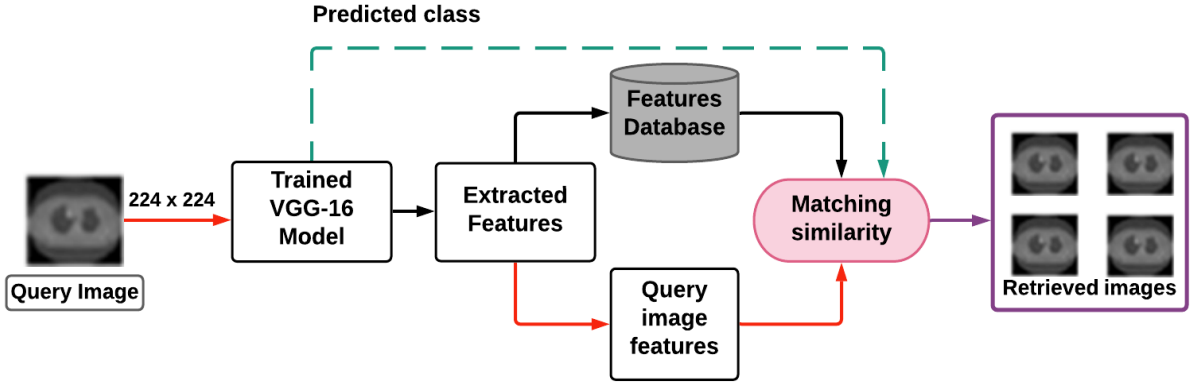
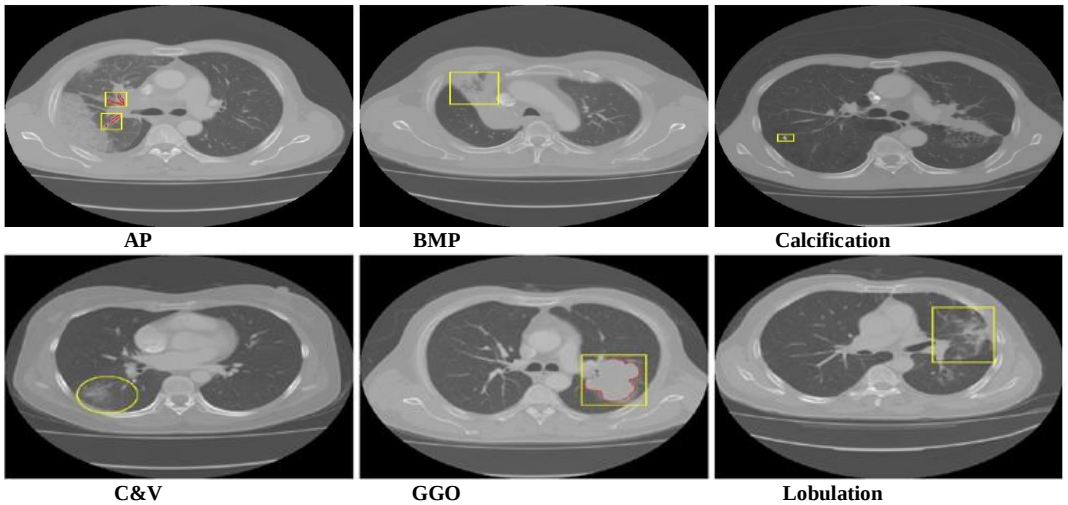
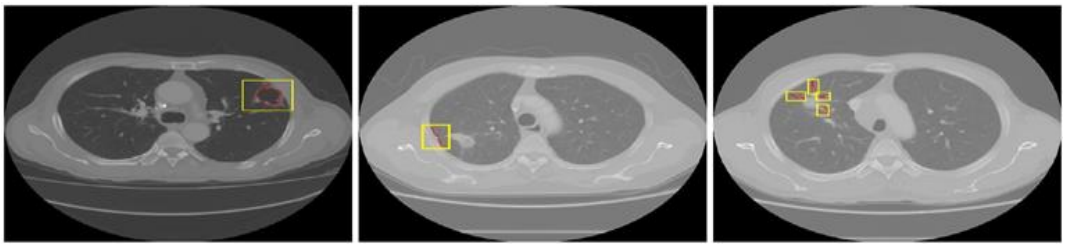


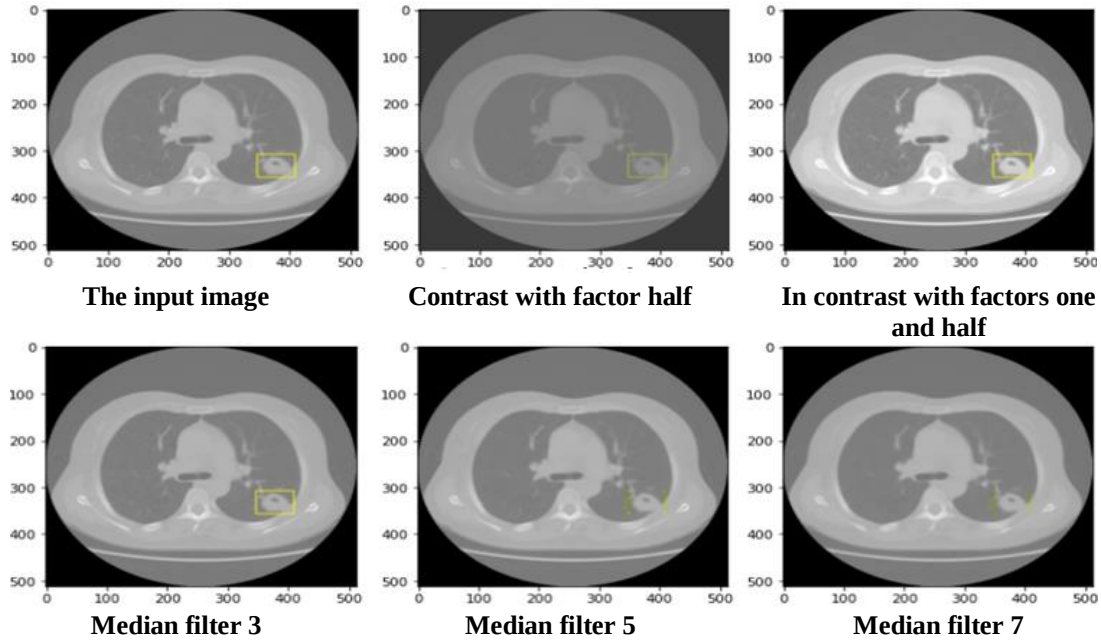
Fig. 3: The retrieval stage for CT lung diseases





OP **PD** **Spiculation**

Fig. 4: The annotated CISs dataset samples



The input image **Contrast with factor half** **In contrast with factors one and half**

Median filter 3 **Median filter 5** **Median filter 7**

Fig 5: The sample of the original annotated image, and after applying the smoothing and contrast adjustment techniques

Finally, to increase the number of input images and balance our CISs classes, we applied the median filtering smoothing and contrast adjustment techniques for the annotated images as clarified in Fig. 5. In the final step in the pre-processing image module, the annotated enhanced images with their labels are augmented based on the proposed augmentation. As discussed in the ontology construction module, the input to this module is metadata that describes the sign in each image.

In order to describe the visual appearance of the C&V sign category, as shown in Fig. 6, a set of vocabularies such as a cavity, modifier, modality, shape, and margin are used.

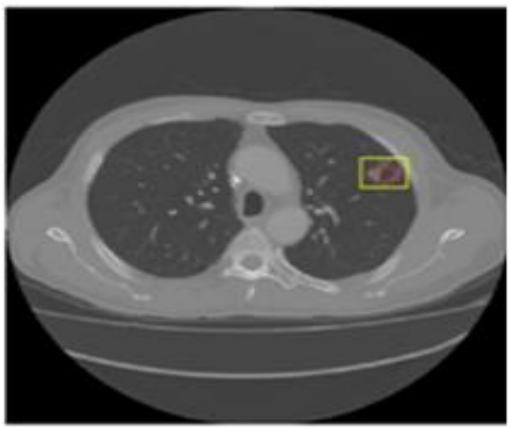


Fig 6: The description of the C&V sign image

These vocabularies are collected in a text file to describe the sign of CT lung image that is a part of the constructed ontology. Each semantic concept is dominated by value as a list form. For example, the cavity is a gas-filled composition, the modifier is a malignant aggressiveness, the modality is a low attenuation, the shape is round, and the margin is regular. This textual information is converted to a feature vector using word embedding after filtering the words by removing words that are not defined

by the WordNet dictionary. Fig. 7 shows the constructed ontology that contains 29 "is-a" taxonomic relations and 30 concept classes. The last layer of VGG-16 is modified to fit the target domain, as presented in section 3.1.3. The merged features from the visual and semantic are used as input for the new FC layer. The output of this new FC layer is nine according to the number of classes.

After applying the constructed ontology in VGG-16 and testing the trained CNN model, the training and validation classification accuracy is improved from 0.7 to 0.952 after 70 epochs. Fig. 8 clarifies the optimal validation classification accuracy of the proposed model.

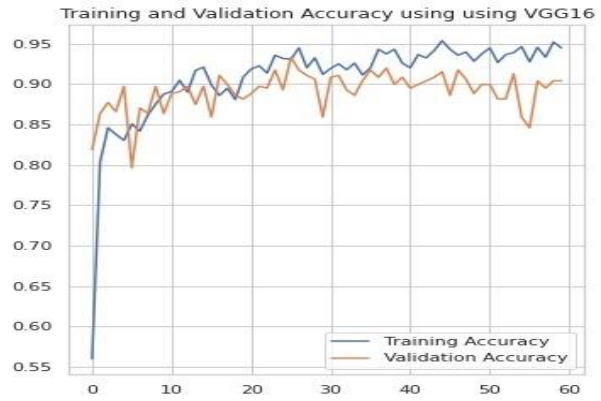


Fig. 8: Optimal training and validation classification accuracy of the proposed system

C. Classification Performance Metrics

The classification performance metrics: Average Precision (AP), Average Recall (AR), accuracy, and F1 measure as explained in Eq.(2), Eq.(3), Eq.(4), and Eq.(5) respectively are used to evaluate the proposed classification stage.

$$AP = \frac{1}{K}\sum_{i=1}^{K}\frac{TP_i}{TP_i + FP_i} \quad (2)$$

$$AR = \frac{1}{K}\sum_{i=1}^{K}\frac{TP_i}{TP_i + TN_i} \quad (3)$$

$$Accuarcy = \frac{1}{K}\sum_{i=1}^{K}\frac{TP_i + TN_i}{TP_i + TN_i + FP_i} \quad (4)$$

$$F1 = \frac{AP * AR}{AP + AR} \quad (5)$$

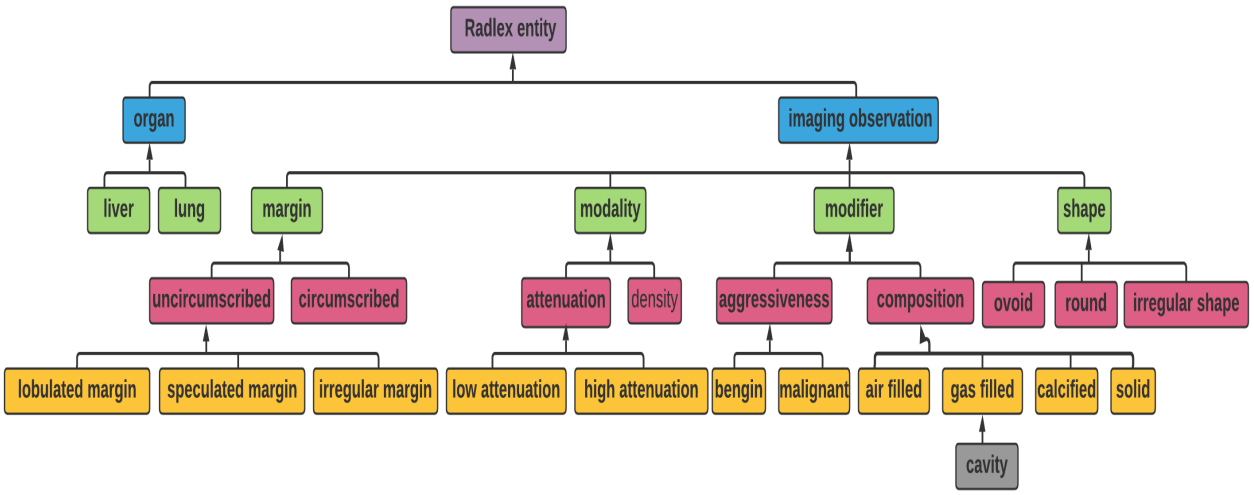


Fig. 7: The constructed visual ontology from CISs for the proposal

Where TP (True Positive) indicates the number of images from class X and correctly classified, FP (False Positive) represents a number of images not from class X but misclassified as class X, TN (True Negative) and point to the number of images that are correctly classified as not belonging to class X, FN (False Negative) shows the images that are from class X but are misclassified and 'K' represents the total number of categories that equals 9 classes in this case.

The precision, recall, and f1-score of the proposed classification stage on nine categories are presented in Table 1. It can be noticed from Table 1 that the average value of the precision, recall, and F1 for all 9 categories is greater than 0.85. Also, our classification method achieved the max precision in GGO and speculation categories. In addition, in some categories, the accuracy and recall are high, and in others, it is low due to the imbalance in the dataset.

Table 1: The precision, recall, and f1-score for signs of CT lung images

CISs	Precision	Recall	F1-score
GGO	**1.00**	0.94	0.97
Spiculation	**1.00**	1.00	1.00
Lobulation	0.96	0.96	0.96
C&V	0.85	0.77	0.80
OP	0.93	0.97	0.95
Calcification	0.68	0.81	0.74
AB	0.98	0.65	0.76
BMP	0.76	0.79	0.77
PD	0.76	0.95	0.84
Average	**0.87**	0.86	0.86

On the other hand, the confusion matrix is computed, as shown in Fig. 8. It concludes that the trained proposed system predicts the input image classes accurately.

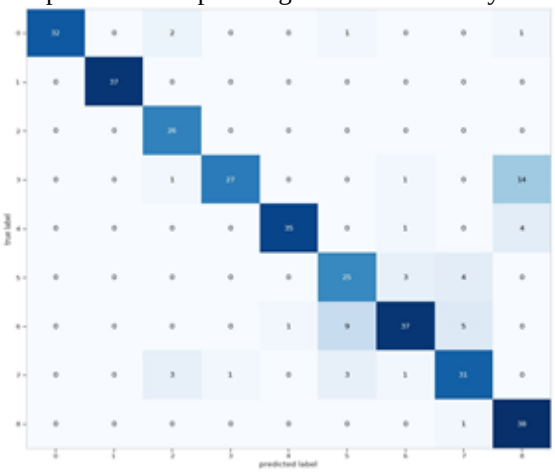


Fig. 8: The confusion matrix of the predicted CISs

D. Retrieval Performance Metrics

For evaluating the performance of the proposed system for the retrieval stage, the retrieval performance metrics: precision at position n (p@n), average precision (AP), and mAP are explained in Eq. (6), Eq. (7), and Eq. (8) respectively.

$$p(n) = \frac{1}{m}\sum_{n=1}^{m} rel(n) \quad (6)$$

Where m is the number of the retrieved images and rel(n) indicates relevancy, rel(n) is 1 for the relevant item at position n, otherwise 0.

$$AP = \frac{\sum_{n=1}^{m} p(n) * rel(n)}{mr} \quad (7)$$

Where mr is the number of relevant images.
The mAP for a set of queries "q" is computed as follows.

$$mAP = \frac{\sum_{n=1}^{q} AP(n)}{Q} \quad (8)$$

E. Comparison

To evaluate the proposed method, we compared the performance of the proposed method with other different techniques such as FCSS [17] and Kashif [16], based on the LISS dataset. As a result, our system achieved an average classification accuracy of 95.2 and outperformed Kashif [16] by approximately 13.5% and FCSS [17] by 9%, as shown in Table 2.

Table2: Comparison of the proposed compared to the conventional techniques in terms of the classification accuracy

Methods	FCSS	Kashif	Proposed
Classification Accuracy	86.27	81.75	95.2

On the other hand, the proposed method is compared to the conventional techniques for the retrieval stage in terms of the AP and mAP. The AP and mAP results of Top 10, 20, 50, and 100 retrieved images on nine categories based on visual similarity are presented in Table 3. It can be noticed from Table 3 that the proposed method performance for the most number of categories in terms of mAP for the whole retrieval process outperforms that of FCSS [15] by approximately 9% and Kashif [14] by 2% at Top 10 retrieved images.

Finally, our system has achieved the mAP of 0.742 using the class prediction that outperforms Kashif [14] by approximately 4% and FCSS [15] by 11% after applying the visual and semantic similarities as shown in Table4.

Also, it can be observed that if a query image is classified correctly, the irrelevant retrieved images do not appear in the retrieval results in contrast to FCSS [15] and Kashif [14].

Table 3: Comparison of retrieval results for the proposed and conventional techniques on the LISS dataset based on visual similarity

Category	AP@10			AP@20			AP@50			AP@100		
	FCSS	Kashif	Proposed	FCSS	Kashif	Proposed	FCSS	Kashif	Proposed	FCSS	Kashif	Proposed
GGO	0.75	0.5454	0.76	0.65	0.495	0.69	0.32	0.4	0.51	0.17	0.32	0.48
Spiculation	0.3	0.5241	0.74	0.225	0.436	0.65	0.16	0.34	0.48	0.12	0.25	0.4
Lobulati	0.71	0.557	0.7	0.65	0.482	0.64	0.3	0.39	0.42	0.155	0.34	0.4

on		9		2								
C&V	0.4	0.7401	0.58	0.575	0.696	0.6	0.47	0.62	0.48	0.411	0.55	0.39
OP	0.45	0.5986	0.63	0.3	0.432	0.54	0.16	0.34	0.35	0.084	0.28	0.29
Calcification	0.4	0.626	0.45	0.25	0.558	0.37	0.22	0.46	0.19	0.135	0.40	0.11
AB	0.45	0.4922	0.63	0.25	0.391	0.51	0.14	0.29	0.38	0.08	0.22	0.28
BMB	0.74	0.7472	0.54	0.722	0.697	0.53	0.55	0.60	0.41	0.31	0.53	0.37
PD	0.6	0.5899	0.56	0.52	0.526	0.42	0.46	0.44	0.29	0.341	0.38	0.19
mAP	0.533	0.6024	0.6211	0.461	0.523	0.55	0.309	0.431	0.39	0.201	0.363	0.3233

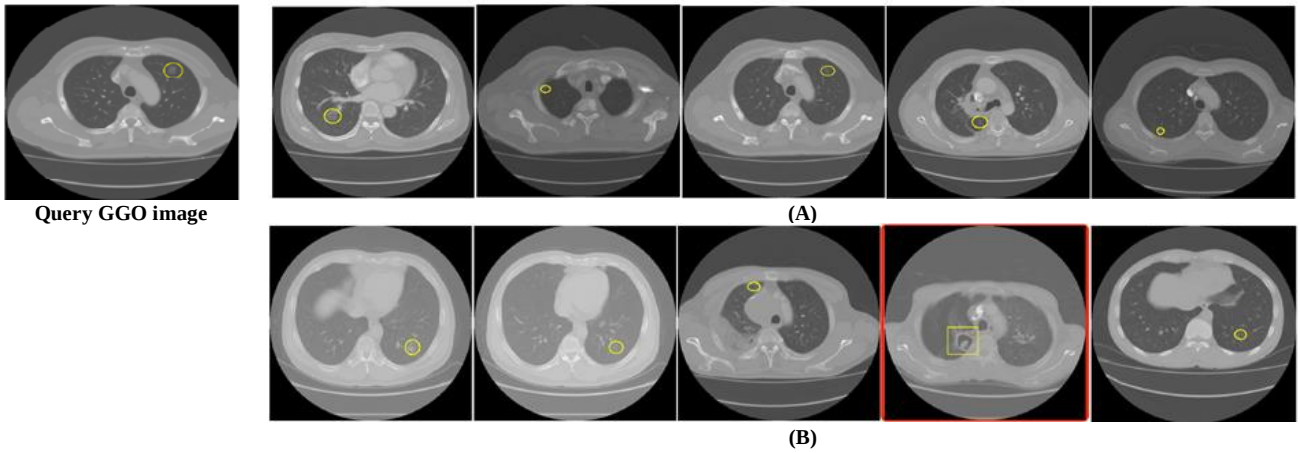


Fig.9: Retrieval results for GGO class (A) retrieved images using class prediction (B) retrieved images without using class prediction

Table4: The mAP values from the proposed and compared methods.

Methods	FCSS	Kashif	Proposed
mAP	0.6333	0.7027	0.742

Fig. 9 shows the GGO query image and the retrieved images similar to the query image on the top 5 results after applying classification results and without incorporating class prediction. Red frames surround the retrieved images that have a different diagnosis, and the other images have the same diagnosis as the GGO query image.

V. CONCLUSIONS

This paper proposed an ontology-CNN-based framework for the CBMIR system by training a VGG-16 model for the classification task. Three strategies are proposed for retrieving a CT scan; the first strategy uses class prediction of the query image by the trained ontology-CNN model and then searches for relevant images in the predicted class. The second method is without incorporating the information about the class of the query image and therefore searching the whole database for relevant images. However, the final strategy did not fuse the semantic knowledge into the trained model to get the classification predictions. Also, the proposed ontology for signs in CT scans of lung diseases has improved the classification performance. The proposed hybrid method improves the classification accuracy by learning discriminative features and fusing the high-level features directly from images. To the best knowledge of the authors of this paper, this is the first time an ontology-deep learning model has been used in a CT scan. The proposed system achieved the highest mAP of 0.742 for retrieving signs in CT scans of lung diseases with class prediction. Furthermore, the evaluation results showed improved precision and recall for retrieval and classification of CT scans.

REFERENCES

[1] Liu X., Ma L., Song L., Zhao Y., Zhao X., Zhou C.: Recognizing common ct imaging signs of lung diseases through a new feature selection method based on fisher criterion and genetic optimization, IEEE journal of biomedical and health informatics, 19(2) (2014) 635–647.

[2] Han G., Liu X., Han F., Santika I. N. T., Zhao Y., Zhao X., Zhou C.: The liss —a public database of common imaging signs of lung diseases for computer-aided detection and diagnosis research and medical education, IEEE Transactions on Biomedical Engineering, 62(2) (2014) 648–656.

[3] Aigrain, P., Zhang, H., Petkovic, D. Content-based representation and retrieval of visual media: a state-of-the-art review,".Multimedia Tools Appl. 3 (1996) 179–202.
[4] Suresh M B, Dr.B Mohankumar Naik. An Efficient Approach of Content-Based Image Retrieval using Texture, Color and Shape Features of an Image, International Journal of Engineering Trends and Technology (IJETT) – 48(2017) 316-320.
[5] Raicu DS, Varutbangkul E, Furst JD, Armato III, SG: Modeling semantics from image data: opportunities from LIDC. IJBET 2 (2008) 1–22.
[6] M.M. Rahman, S.K. Antani, G.R. Thoma, A medical image retrieval framework in correlation enhanced visual concept feature space, in 22nd IEEE International Symposium on Computer-Based Medical Systems, (2009) 1–4.
[7] Johannes Hofmanninger, Georg Langs, Mapping Visual Features to Semantic Profiles for Retrieval in Medical Imaging, Proceedings of the IEEE, – open access. The cvf.com, (2015) 457-465.
[8] Camille Kurtz, Adrien Depeursinge, Sandy Napel, Christopher F. Beaulieu, Daniel L. Rubin. On combining image-based and ontological semantic dissimilarities for medical image retrieval applications, Medical Image Analysis, 18(7) (2014) 1082-1100.
[9] Yang Chen, Xiaofeng Ren, Guo-Qiang Zhang, Rong X, Ontology-guided organ detection to retrieve web images of disease manifestation: towards the construction of a consumer-based health image library, Med Inform Assoc (2013) 1076–1081.
[10] Filali J, Zghal HB, Martinet J. Ontology and HMAX features-based image classification using merged classifiers. In: Proceedings of the 14th international joint conference on computer vision, imaging and computer graphics theory and applications, VISIGRAPP 5 (2019). SciTePress, (2019) 124–134.VISAPP, Prague, Czech Republic, February 25-27.
[11] Filali, J., Zghal, H. B., and Martinet, J.. Towards Visual Vocabulary and Ontology-based Image Retrieval System. In Proceedings of the 8th International Conference on Agents and Artificial Intelligence (ICAART 2016) – 2 (2016) 560-565.
[12] Umar Manzoor, Mohammed A., Bassam Zafar, Hafsa Umar, M. Shoaib Khan, Semantic Image Retrieval: An Ontology-Based Approach, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, 4(4) (2015).
[13] R. Biswas, S. Roy Content-Based CT Image Sign Retrieval using Fast Discrete Curvelet Transform and Deep Learning, International Journal of Advanced Trends in Computer Science and Engineering 8(3) (2019).
[14] M. Kashif, Gulistan Raj.,f. Shaukat, An Efficient Content-Based Image Retrieval System for the Diagnosis of Lung Diseases, Journal of Digital Imaging (2020) 971–987.
[15] Ma L., Liu X., Gao Y., Zhao Y., Zhao X., Zhou C.: A new method of content-based medical image retrieval and its applications to ct imaging sign retrieval, Journal of biomedical informatics, 66 (2017) 148–158.
[16] Ranjit Biswas, Sudipta Roy, Abhijit Biswas, Triplet Contents based Medical Image Retrieval System for Lung Nodules CT Images Retrieval and Recognition Application, International Journal of Engineering and Advanced Technology (IJEAT)ISSN: 2249 –8958, 8(6) (2019).
[17] Luhn, H. P. A statistical approach to the mechanized encoding and searching of literary information. IBM Journal of Research and Development 1(4), (1957) 309–317.
[18] D. B. Lenat, Cyc: A Large-scale investment in Knowledge Infrastructure, Communications of the ACM, 38(11) (1995) 33-38.
[19] G. A. Miller, Wordnet: a lexical database for English, Communications of the ACM, 38(11) (1995) 39–41.
[20] G. Miller, Nouns in WordNet: a Lexical Inheritance System, International Journal of Lexicography, 3(4) (1994) 245-264.
[21] Maillot N, Thonnat M, Boucher A. Towards ontology-based cognitive vision, Machine Vision and Applications. Springer. 16(1) (2004) 33–40.