International Journal of Computer Assisted Radiology and Surgery
https://doi.org/10.1007/s11548-020-02283-z

ORIGINAL ARTICLE

A bilinear convolutional neural network for lung nodules classification on CT images

Rekka Mastouri[1] · Nawres Khlifa[1] · Henda Neji[2,3] · Saoussen Hantous-Zannad[2,3]

Received: 12 June 2020 / Accepted: 21 October 2020

Abstract

Purpose Lung cancer is the most frequent cancer worldwide and is the leading cause of cancer-related deaths. Its early detection and treatment at the stage of a lung nodule improve the prognosis. In this study was proposed a new classification approach named bilinear convolutional neural network (BCNN) for the classification of lung nodules on CT images.

Methods Convolutional neural network (CNN) is considered as the leading model in deep learning and is highly recommended for the design of computer-aided diagnosis systems thanks to its promising results on medical image analysis. The proposed BCNN scheme consists of two-stream CNNs (VGG16 and VGG19) as feature extractors followed by a support vector machine (SVM) classifier for false positive reduction. Series of experiments are performed by introducing the bilinear vector features extracted from three BCNN combinations into various types of SVMs that we adopted instead of the original softmax to determine the most suitable classifier for our study.

Results The method performance was evaluated on 3186 images from the public LUNA16 database. We found that the BCNN [VGG16, VGG19] combination with and without SVM surpassed the [VGG16]2 and [VGG19]2 architectures, achieved an accuracy rate of 91.99% against 91.84% and 90.58%, respectively, and an area under the curve (AUC) rate of 95.9% against 94.8% and 94%, respectively.

Conclusion The proposed method improved the outcomes of conventional CNN-based architectures and showed promising and satisfying results, compared to other works, with an affordable complexity. We believe that the proposed BCNN can be used as an assessment tool for radiologists to make a precise analysis of lung nodules and an early diagnosis of lung cancers.

Keywords Bilinear CNN · SVM · Lung nodules · Classification

Introduction

Lung cancer is the first cause of cancer-related death worldwide. The mortality rate of lung cancer surpasses that of prostate, colon and breast cancers. The American Cancer Society estimates that lung cancer is expected to cause 135,720 deaths in 2020, accounting for about 25% of all cancer deaths [1, 2]. Early screening and localization of lung cancer in situ at its nodular stage are very important and beneficial to improve the patient's treatment effect. In fact, early diagnosis of lung cancer increases the five-year survival rate from 18.6 to 56% [2, 3].

Thanks to its high spatial resolution, computed tomography (CT) is the reference imaging modality used to detect, characterize lung nodules (dimensions, density, shape, contours) and follow their growth [3]. Suspicious nodules are sometimes difficult to identify because of their small size, human subjectivity and technical defaults of some CTs [3]. Therefore, computer-aided diagnosis (CAD) systems are becoming necessary to overcome these deficiencies.

Recently, deep learning (DL) has showed a big success and has become a choice method in medical images analysis due to its promising results. CNN is the prevalent deep learning model, widely applied in various image analysis fields. This model is able to automatically extract relevant features from

✉ Rekka Mastouri
rekka.mastouri@gmail.com

1 Higher Institute of Medical Technologies of Tunis, Research Laboratory of Biophysics and Medical Technologies, University of Tunis el Manar, 1006 Tunis, Tunisia

2 Faculty of Medicine of Tunis, University of Tunis el Manar, 1007 Tunis, Tunisia

3 Medical Imaging Department, Abderrahmen Mami Hospital, 2035 Ariana, Tunisia

Published online: 02 November 2020

images and to solve various learning problems, such as object detection, pattern recognition and classification [3].

In this paper, we proposed a trending CNN model called bilinear CNN (BCNN) to classify pulmonary nodules in CT images. BCNN model relies on the pooling and multiplication of the output features of two independent CNNs by an outer product (Kronecker's matrix product) to form a vector descriptor [4]. The bilinear pooling process ensures the easy computation of gradient and end-to-end learning of both feature extractors using image labels only [4]. Moreover, this type of model is highly recommended for fine-grained classification problems since it can reliably interrelate local features in a translational invariant manner [4]. Pre-trained CNN models for deep feature extraction were particularly implemented, followed by a linear support vector machine (SVM) for classification. Different CNNs (VGG16 and VGG19) were explored via the fine-tuning process to build our BCNN model. In addition, different types of SVMs (linear, polynomial and radial basis function (RBF)) were tested to determine the most appropriate type for our classification process. In the experiments, we used 888 CT scans containing 1186 nodules from the public database LUNA16.

This paper is organized as follows: In "Related works" section, we summarized some relevant works from the literature. In "Materials and methods" section, we described in detail the proposed system. Experimental results and discussion are presented in "Results and discussion" section. Finally, the conclusion and perspectives are summarized in "Conclusion" section.

Related works

CNN was potentially introduced to medical image analysis field thanks to its ability to self-determine relevant features through its convolutional and pooling layers. Several architectures were progressively implemented in CAD systems for lung nodules classification on CT scans, achieving great performance and satisfactory results. Monkam et al. [5] proposed various CNN architectures with various numbers of convolutional layers for micro-nodules (diameter <3 mm) and non-nodules classification on CT scans. They trained the proposed models on different images sizes of 16×16, 32×32 and 64×64 extracted from 512×512 CT images from LIDC–IDRI database using the fivefold cross-validation method. The experiments showed that the CNN with two convolutional layers, in the case of 32×32 patches size, presented a great performance and promising results. As well, in another work, Monkam et al. [6] improved their experimental results by testing various classifiers coupled to CNN instead of softmax. The extreme learning machine (ELM) presented promising results compared to other models, achieving a sensitivity rate of 96.57% and an accuracy rate of 97.35%. Shen et al. [7] developed a multi-crop CNN architecture allowing to crop regions presenting relevant information from convolutional feature maps. The proposed model made it possible to integrate the nodule features into a hierarchical network without a segmentation procedure, thus simplifying classification. Onishi et al. [8] proposed a novel lung cancer lesions classification method in CT images using a deep convolutional neural network (DCNN) and generative adversarial networks (GANs). First, authors used Wasserstein generative adversarial networks (WGANs) for new nodule images generation to increase the performance of AlexNet-based DCNN model. Then, the DCNN was trained on the generated images to classify candidate lesions as malignant and benign using the fine-tuning process. The algorithm was evaluated on private database containing CT images of 60 cases (benign: 27 and malignant: 33), and it achieved an accuracy rate of 66.7% for benign nodules and 93.9% for malignant nodules. Wu et al. [9] proposed to merge residual learning and transfer learning techniques to construct a new CNN model for the classification of pulmonary nodules. ResNet-50 architecture was used for training and testing of the selected datasets based on the tenfold cross-validation method. The model was evaluated on 7685 images of nodules and 7310 images of non-nodules collected from LIDC–IDRI database and achieved accuracy, sensitivity and FP rates of 98.23%, 97.70% and 1.65%, respectively. Kaya et al. [10] combined handcrafted features with deep features extracted from AlexNet architecture and trained various classifiers (SVM, KNN, AdaBoost and random forest) to distinguish between lung nodules and non-nodules. Kido et al. [11] took advantages of the CNN and R-CNN to develop a high-performance system for lung nodules detection. An object detector R-CNN followed by a SVM classifier was implemented for candidate nodules selection. AlexNet architecture was trained on detected nodules and used as a feature extractor for 1304 nodules, obtained after data augmentation. Then, the authors fed the extracted features into a multi-class SVM classifier for the recognition of malignant and benign nodules. The proposed scheme achieved an accuracy rate of 95.2% increased with data augmentation to 99.4%.

In recent years, bilinear CNN model was highly proposed in the literature for fine-grained classification problems. BCNN multiplies the convolutional layer outputs of two different CNNs by an outer product, which gives the quadratic number of characteristic maps. Therefore, BCNN takes advantages of two networks and combines them in one which outperforms most of the existing algorithm of classification. Liu et al. [12] proposed to combine compact BCNN with histogram loss while adding two-part networks to reinforce their effectiveness in person re-identification. Meanwhile, Ustinova et al. [13] proposed multi-region BCNN for person re-identification. Chen et al. [14] implemented two pre-trained CNN architectures (inception-v3 and inception-

Table 1 Overview of the existing classification methods

References	Database	Model	Performance metrics
Lung nodule classification			
Monkam et al. [5]	LIDC–IDRI	3 CNNs with various patch sizes	Sensitivity = 83.82% Accuracy = 88.28% F1-score = 83.45% AUC = 0.87
Monkam et al. [6]	LIDC–IDRI	Ensemble 3D CNNs + ELM	Sensitivity = 96.57% Accuracy = 97.35%
Shen et al. [7]	LIDC–IDRI	Multi-crop CNN	Sensitivity = 77% Accuracy = 87.14% AUC = 0.93
Onishi et al. [8]	Fujita Health University Hospital	WGAN + AlexNet	Accuracy (benign) = 66.7% Accuracy (malignant) = 93.9%
Wu et al. [9]	LIDC–IDRI	Deep residual ResNet-50	Sensitivity = 97.7% Accuracy = 98.23% FPR = 1.65%.
Kaya et al. [10]	LIDC–IDRI	AlexNet + cascaded classifier	Sensitivity = 67.37% Specificity = 95.46%. Accuracy = 84.70%
Kido et al. [11]	LIDC–IDRI	CNN + SVM	Accuracy = 95.2%
Fine-grained classification			
Liu et al. [12]	CUHK03 detected CUHK03 labeled	Multi-part compact Bilinear CNN	Recall@K($r = 5$) = 90.28% Recall@K($r = 5$) = 94.86%
Ustinova et al. [13]	Market-1501 CUHK01 CUHK03 detected CUHK03 labeled	Multi-region bilinear CNN	Recall@K($r = 5$) = 85.01% Recall@K($r = 5$) = 78.08% Recall@K($r = 5$) = 89.15% Recall@K($r = 5$) = 93.37%
Chen et al. [14]	UECFOOD-100 UECFOOD-256 ETHFOOD-101	BCNN [Inception-v3, Inception-v4]	$\text{Accuracy}_{\text{TOP-5}} = 99.28\%$ $\text{Accuracy}_{\text{TOP-5}} = 95.49\%$ $\text{Accuracy}_{\text{TOP-5}} = 98.76\%$
Wang et al. [15]	H&E-stained colorectal cancer histopathological image dataset	BCNN + SVM	Accuracy = 92.6% AUC = 98.5%

v4) for food identification as fine-grained recognition tasks using the bilinear technique. They performed three combinations using the two CNNs and compared their performance using three different datasets. In the medical imaging field, Wang et al. [15] developed a new BCNN-based model able to decompose histopathological images into two components and extract more effective feature maps of the two decomposed images. Classification methods, databases and performance metrics of previous works are summarized in Table 1.

Materials and methods

In this work, we propose a BCNN-based lung nodule classification model. CNN-based systems require a large amount of data for their training, and it is not easy to recover an extensive database of medical images. That is why we used a public database (LUNA16) to train two pre-trained CNN architectures (VGG16 and VGG19). Precise fine-tuning of pre-trained model parameters is recommended using our targeted database for training and testing. The proposed work is shown in the block diagram in Fig. 1.

Dataset and preprocessing

In this work, we relied on the benchmark datasets available from the LUNA16 challenge [16]. The challenge includes 888 CT scans in MetaImage format (mhd/raw) filtered out from the public Lung Image Database Consortium and Image Database Resource Initiative (LIDC–IDRI) database [17]. These scans are stored at 512×512 pixels resolution with a slice thickness under 2.5 mm. The LUNA16 database includes all nodules with diameter greater than 3 mm annotated by at least three out of four radiologists, resulting in 1186 nodules. Non-nodules, micro-nodules (<3 mm) and masses (≥ 3 mm) annotated by less than three radiologists are treated as non-nodules. There are 549,870 non-nodules.

In our study, we reframed the image according to the (x, y, z) annotations to create the learning datasets. We noted

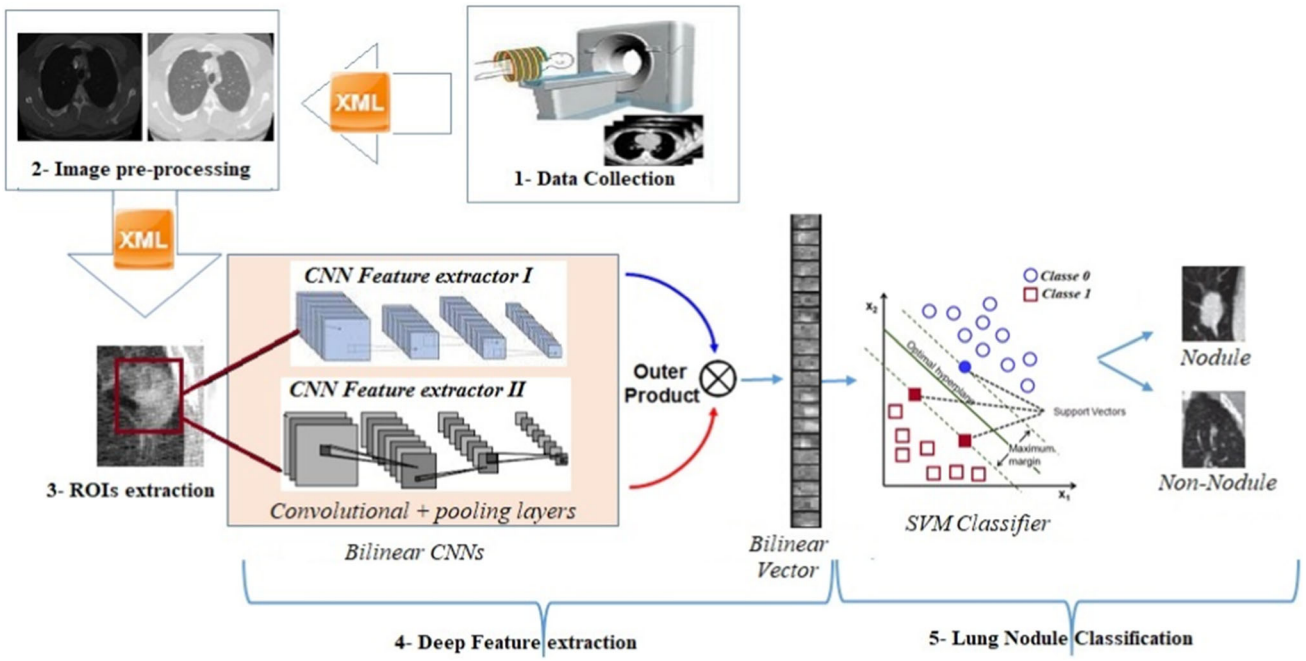

Fig. 1 Workflow of the proposed bilinear CNN model for lung nodule classification

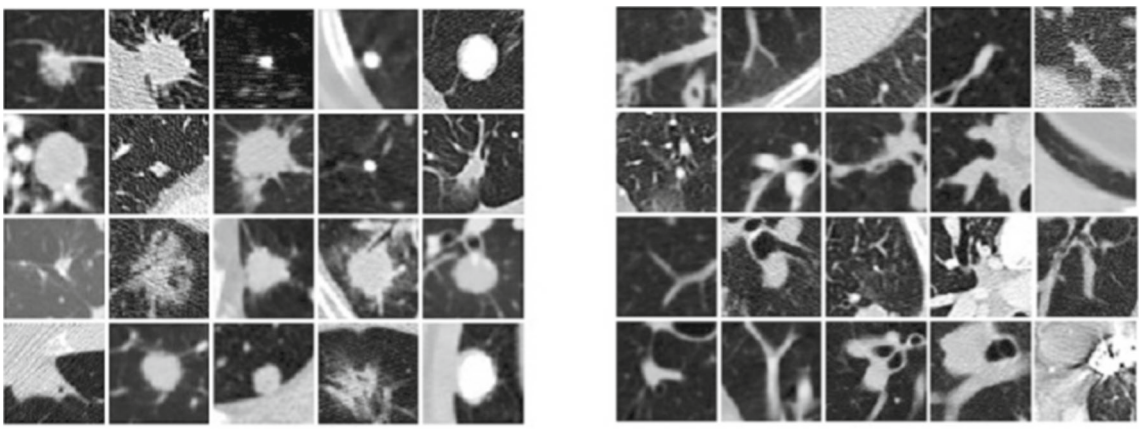

Fig. 2 Example of extracted datasets from LUNA16 database: **a** nodule and **b** non-nodule ROIs

a concern issue related to class imbalance, so we manually and randomly sub-sampled the majority class. Then, we rescaled the intensity of images that were defined at the Hounsfield scale, to standardize the image data before being used as input for training the model. Finally, we generated 50×50 grayscale images including 1186 nodules and 2000 non-nodules for the training and validation of the proposed BCNN model. Figure 2 shows some examples of generated images.

The training step of CNN requires a high number of datasets. For this reason, we used data augmentation techniques prior to training the model. We chose then to apply uniform transformations to keep the shape of nodules given its importance in the diagnosis and detection process. We applied the following augmentations to each patch: rotation 45°, zoom range of [0.5, 0.9], flipping horizontally and flipping vertically. We obtained after data augmentation 5930 patches for the nodule class, and we reserved 80% of the dataset to the training and 20% to the validation of the model.

Bilinear CNN description

Bilinear model was originally developed by Lin et al. [18] for fine-grained classification and person identification tasks. The bilinear CNN is made of two-branch CNNs as feature extractors whose output vectors are bilinearly pooled through

an outer product function. Thus, the BCNN model produces a large amount of information compared to conventional CNN models. According to the methodology proposed in [18], a BCNN model for image classification is shown in Fig. 1 and it can be formally represented as the following quadruple:

$$\mathcal{B}_{\text{CNN}} = \left(f_1, f_2, \mathcal{P}_f, \mathcal{C}\right) \tag{1}$$

where f_1 and f_2 are featured extractors of CNN1 and CNN2, respectively, $\mathcal{P}_f$ is a pooling function, and $\mathcal{C}$ is a classification function. The feature extractor produces a mapping function: $F : \mathcal{I} \times \mathcal{L} \rightarrow \mathbb{R}^{\mathcal{K}\times\mathcal{D}}$ that considers an image $I \in \mathcal{I}$ and a location $l \in \mathcal{L}$ and gives in its output a feature $\mathbb{R}$ of size $\mathcal{K} \times \mathcal{D}$. Generally, $\mathcal{L}$ refers to position and scale. To obtain bilinear features, we combine the output of f_1 and f_2 at each location l by computing the outer matrix product as follows:

$$\mathfrak{b}(l, \mathcal{I}, f_1, f_2) = f_1(l, \mathcal{I})^T f_2(l, \mathcal{I}) \tag{2}$$

The feature $\mathcal{K} f_1$ and f_2 must have the same dimension to be compatible.

The global image descriptor $\phi(\mathcal{I})$ is obtained when bilinear features are aggregated by the pooling function $\mathcal{P}_f$ at all positions of the image. The sum pooling is then used, and the descriptor is formally defined as:

$$\phi(\mathcal{I}) = \sum_{l\in\mathcal{L}} \mathfrak{b}(l, \mathcal{I}, f_1, f_2) \tag{3}$$

The feature descriptor $\phi(\mathcal{I})$ is an orderly representation since pooling disregards the feature locations. Let's consider $\mathcal{K} \times N$ and $\mathcal{K} \times M$ the dimensions of the extracted feature by f_1 and f_2, respectively; the final bilinear feature $\phi(\mathcal{I})$ is then a general-purpose image descriptor of size $M \times N$ that will be used with a classification function $\mathcal{C}$. The bilinear descriptor is reshaped to 1D bilinear vector $\mathcal{V}(\mathcal{I})$ and undergoes an element-wise signed square root operation as:

$$\mathcal{Y}(\mathcal{I}) = sign(\mathcal{V}(\mathcal{I}))\sqrt{\mathcal{V}(\mathcal{I})} \tag{4}$$

$\mathcal{Y}(\mathcal{I})$ followed by ℓ_2 normalization to improve model performance in practice.

$$\mathcal{Z}(\mathcal{I}) = \frac{\mathcal{Y}(\mathcal{I})}{\mathcal{Y}(\mathcal{I})^2}. \tag{5}$$

The proposed model

In this section, we describe our proposed BCNN model used for the classification of pulmonary nodules on CT images. As shown in Fig. 2, we used two pre-trained CNN architectures for feature extraction. It is worth noting that the right choice of the feature extractor is critical and deemed to be the main part in a BCNN architecture because of its influence on the determination of model performance. In our BCNN model, we chose the most popular VGG16 [19] and VGG19 [19] networks as feature extractors. We designed three combinations using the two CNN architectures: (1) The bilinear model consists of two VGG16 streams denoted by $[\text{VGG16}]^2$, (2) the bilinear model consists of two VGG19 streams denoted by $[\text{VGG19}]^2$, and (3) the bilinear model consists of VGG16 and VGG19 indicated by [VGG16, VGG19].

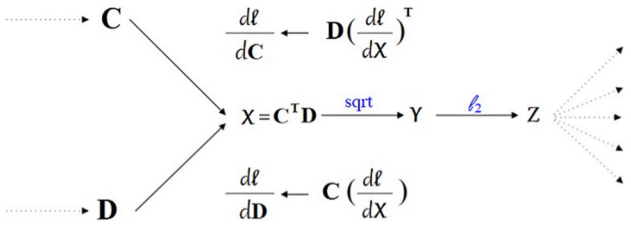

Fig. 3 End-to-end gradient computation in BCNN model [18]

Since the bilinear form performs a much more straightforward gradient computation at the pooling layer, we built our BCNN model with an end-to-end manner and we trained it by back-propagating the gradients of classification loss (e.g., cross-entropy loss function). Here, we considered two matrices C and D of size $\mathcal{K} \times N$ and $\mathcal{K} \times M$, respectively, to represent the output maps of the two feature extractors, and $X = C^T D$ of $M \times N$ dimension is the pooled bilinear feature obtained by formula (2). We denoted by $\frac{\partial \ell}{\partial X}$, the gradient of the loss function ℓ with respect to representation X. As the classification gradient and normalization layer are straightforward, we efficiently obtained the backward propagation of bilinear pooling layer by the following chain rule of gradients:

$$\frac{\partial \ell}{\partial C} = D\left(\frac{\partial \ell}{\partial X}\right)^T, \quad \frac{\partial \ell}{\partial D} = C\left(\frac{\partial \ell}{\partial X}\right) \tag{6}$$

The global scheme describing the end-to-end gradient computation in the BCNN model is illustrated in Fig. 3.

Afterward, we introduced the extracted bilinear vector features in various SVM types that are adopted for final prediction instead of the original softmax classifier. SVM is a supervised classifier originally developed for a binary classification. Its objective is to find an optimal hyperplane that separates sets of negative and positive nodules with a maximum margin depending on the extracted bilinear features. For further study, we implemented SVM with different kernel functions:

$$\text{Linear kernel}: K(x, x_i) = x^t \times x_i \tag{7}$$

$$\text{Polynomial kernel}: K(x, x_i) = \left(x^t \times x_i + b\right)^d; \ b > 0 \tag{8}$$

Radial basis function kernel (RBF)

$$: K(x, x_i) = \exp\left(\frac{|x - x_i|^2}{2\sigma^2}\right);\ \sigma \neq 0 \tag{9}$$

where is degree polynomials, $\in \mathbb{R}$, and is the point belonged to the hyperplane satisfying $(w \cdot x) + b = 0$, where is normal to the hyperplane.

Moreover, we modified the parameters of SVMs and compared their performances to get a better classification rate. Finally, we compared our best results to various classifiers [AdaBoost, k-nearest neighbor (KNN) and random forest (RF)] results and other state of the arts.

Results and discussion

In this section, we evaluate the achievements of our BCNN-based system and we describe the details of the experiment and analyze the results. Finally, we compare our experimental results with previous works to highlight the model performances.

BCNN-softmax parameter settings

All experiments were performed using NVIDIA GeForce GTX 1080 Ti with 11 GB of graphics memory and via Keras libraries. In this work, we took into account the prominence of the cost of calculation and the training time of BCNN models. In medical applications, the most important parameter is the error rate. Thus, we attributed optimal parameters to best reduce mis-classification rates while focusing on training time.

Moreover, we fine-tuned the networks and we trained in our models the last 4 layers of the VGG16 [19] architecture and the last 5 layers of the VGG19 [19] architecture. Figure 4 and Table 2 show the training/validation errors/accuracy in each epoch of the three proposed BCNN-softmax models and the used hyper-parameters for the training of models, respectively.

BCNN-SVM parameter settings

As described in "The proposed model" section, we introduced the extracted bilinear features in various SVM types that are adopted for final prediction instead of the original softmax classifier. To improve the computational efficiency of the system and obtain a better classification rate, we compared the multi-SVM classifier (linear, RBF and polynomial) with different parameters (C, γ and d). Thus, we tested the performance of each SVM type in each BCNN model with a different combination of parameters. To clarify the influence of these parameters on nodules classification efficiency, we presented in Fig. 5 the results of different performed tests.

Using the optimal parameters and among the different SVMs that were tested, we have significantly found that a linear SVM with $C_{\mathrm{SVM}} = 0.1$ gave the best accuracy rate for the three BCNN combinations. In fact, the system achieved accuracies of 91.84%, 90.58% and 91.99% for features extracted from $[\mathrm{VGG16}]^2$, $[\mathrm{VGG19}]^2$ and [VGG16, VGG19] models, respectively, as best results of the linear SVM with $C_{\mathrm{SVM}} = 0.1$.

In the next section, we will consider only the linear SVM classifier to interpret and discuss the performance results of the proposed models.

Experiments

In our approach, we executed three BCNN combinations using two of the most popular pre-trained networks (VGG16 and VGG19). First, we trained each combination using a softmax classifier. Then, we fed the bilinear vector features of each combination into a linear SVM classifier. In addition, we implemented various other classifiers and we compared their performances. Table 3 shows the classification performance of each bilinear model with the softmax and the linear SVM classifiers.

To further demonstrate the effectiveness of the proposed approach, we displayed in Fig. 6 the receiver operating characteristic (ROC) curves and the AUC for each BCNN-SVM model. The nearest curve of the upper left corner of the figure corresponds to the best model. As shown here and when we took AUC into account, [VGG16, VGG19] outperforms $[\mathrm{VGG16}]^2$ and $[\mathrm{VGG19}]^2$ architectures by 1% and 2%, respectively, since it covers more surface than other models. Thus, considering the three combinations, [VGG16, VGG19]-SVM achieves promising results with an accuracy rate of 91.99% and an AUC of 95.9%, which exceed those of the other two models.

Furthermore, an objective comparison between our results and the existing works was made and is denoted in Table 4. We noticed that the three BCNN structures performed better than those in the literature. Our experimental studies substantiated the efficiency and reliability of the proposed BCNN-SVM method in regard to accuracy and AUC metrics.

Discussion

In this study, we proposed three BCNN structures associated with a linear SVM for the classification of lung nodules on CT scans. We used two of the most known pre-trained networks (VGG16 and VGG19) as feature extractors to build our three bilinear proposed architectures. First, we fine-tuned the

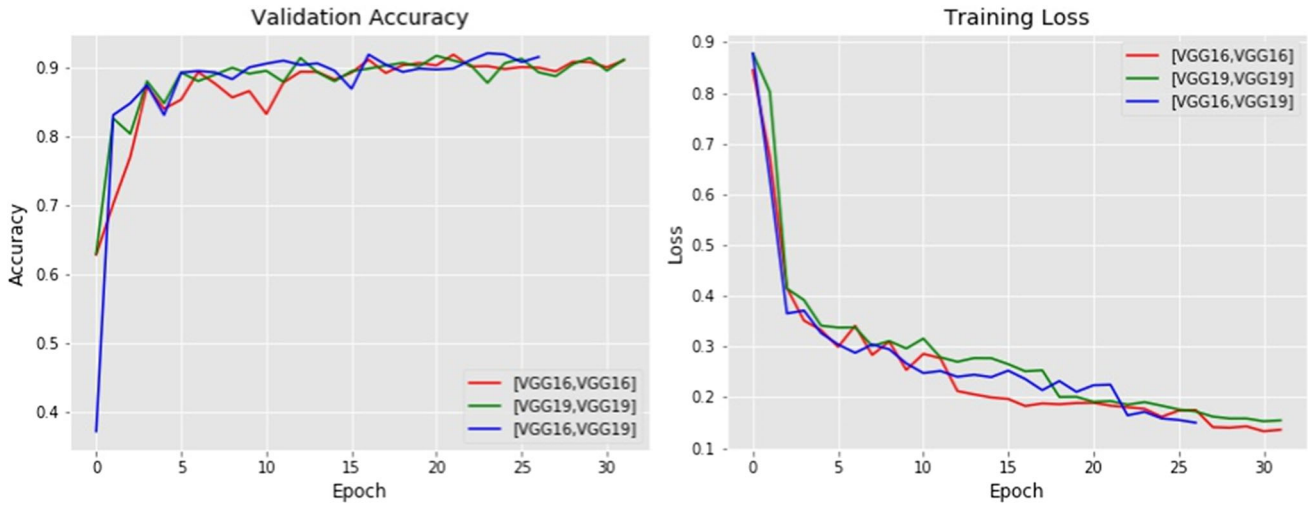

Fig. 4 Validation accuracy and training loss of the three BCNN-softmax models

Table 2 Setup of hyper-parameters

Model	Initializer	Optimizer	Batch size	Learning rate	Steps/epoch	Epoch	Trained layer
[VGG16]^2	Random	SGD	16	0.1	800	32	− 4
[VGG19]^2	Random	SGD	16	0.1	800	32	− 5
[VGG16, VGG19]	He_uniform	SGD	16	0.1	800	27	[− 4, − 5]

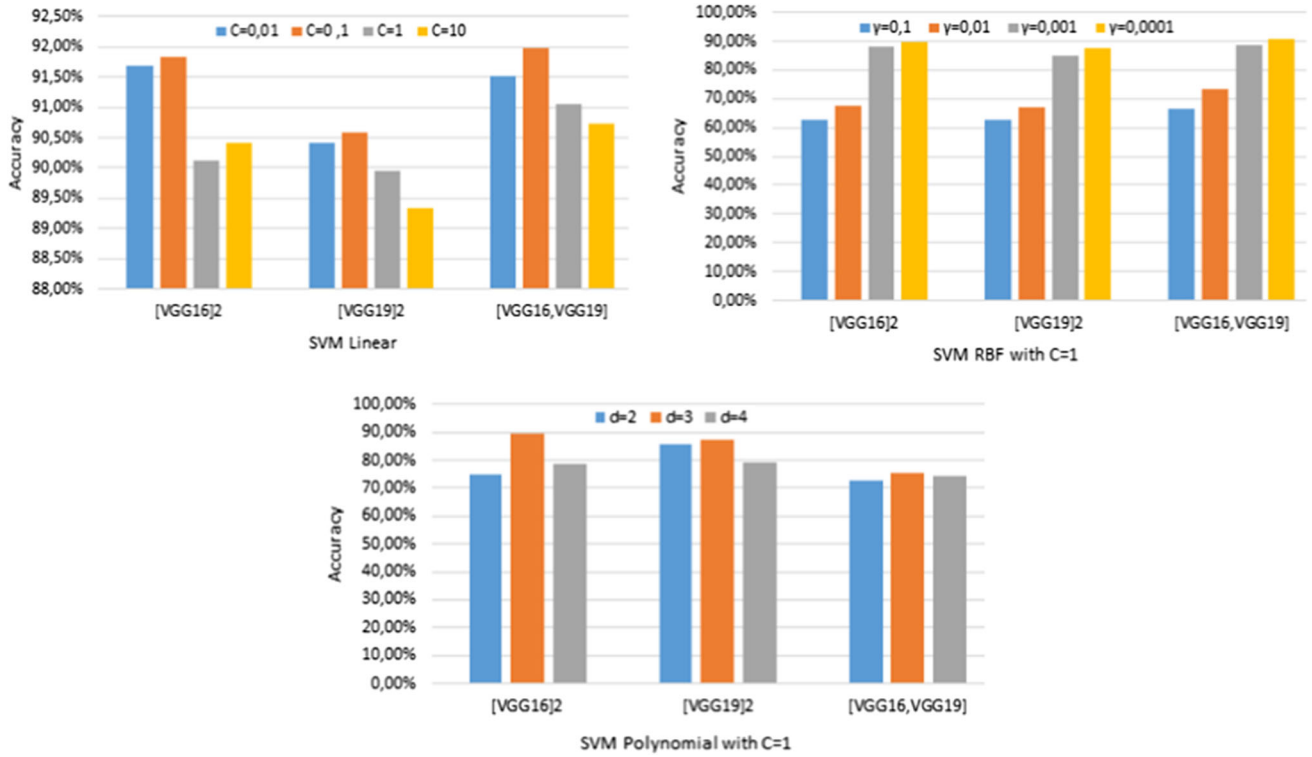

Fig. 5 Various SVM tests with different parameters

Table 3 Performance metrics of each BCNN architecture with softmax and SVM classifiers

Bilinear CNN architecture		Accuracy (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	FPR (%)
$[VGG16]^2$	Softmax	91.46	91.94	90.63	93.35	9.28
	SVM	91.84	93.28	89.36	93.52	8.14
$[VGG19]^2$	Softmax	91	91.73	89.82	92.97	10.18
	SVM	90.58	92.29	87.66	92.52	12.29
[VGG16, VGG19]	Softmax	91.62	91.81	91.44	93.50	8.63
	SVM	91.99	91.85	92.27	93.76	7.72

Table 4 Classification results of BCNN models compared to previous works

References	Database	Methods	Accuracy (%)	AUC
Monkam et al. [5]	LIDC–IDRI (2635 nodules)	Three CNNs with various patch sizes	88.28	0.87
Shen et al. [7]	LIDC–IDRI (2618 nodules)	Multi-crop CNN	87.14	0.93
Kaya et al. [10]	LIDC–IDRI (1402 nodules)	AlexNet + cascaded classifier	84.70	–
Shi et al. [20]	LIDC–IDRI (1400 images)	VGG16 with SVM	91.5	–
Zhao et al. [21]	LIDC–IDRI (743 images)	Agile CNN (LeNet + AlexNet)	82.23	0.877
Zhao et al. [22]	LIDC–IDRI (743 images)	Transfer learning CNNs	85	0.94
Shen et al. [23]	LIDC (4252 images)	Hierarchical semantic CNN	84.2	0.856
Our proposed method	LIDC–IDRI (3186 images)	$[VGG16]^2$ + SVM	91.84	0.948
		$[VGG19]^2$ + SVM	90.58	0.94
		[VGG16,VGG19] + SVM	91.99	0.959

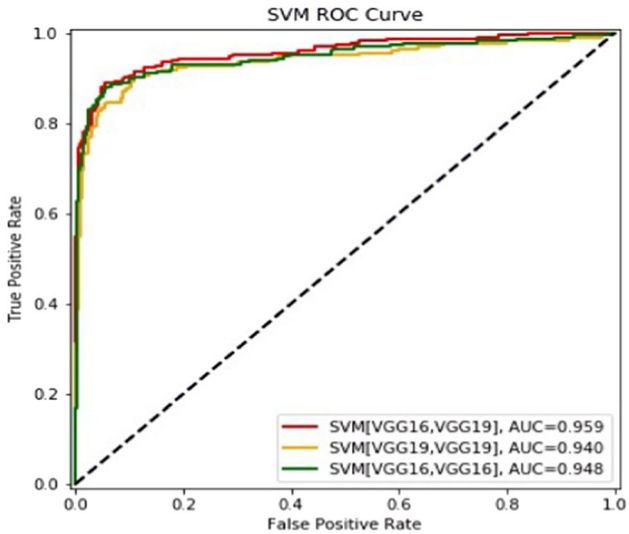

Fig. 6 ROC curves and AUC of the proposed BCNN-SVM structures

CNNs and we compared the performance of each BCNN with a softmax classifier. We found that [VGG16, VGG19] structure exceeded the $[VGG16]^2$ and $[VGG19]^2$ architectures and achieved an accuracy rate of 91.62% against 91.46% and 91%, respectively.

Originally, CNN-based image classification is provided by the fully connected layers with a softmax function for its ease of use. In fact, the softmax itself is not a classifier but just a typical logistic regressor that generates a vector representing the probability distributions ranging between 0 and 1 of a list of logits scores. So, it is suitable as a cost function for multi-class categories detection. However, the SVM performs a different interpretation of the scores: Thanks to its loss function, which results in a hyperplane of maximum margin, it considers them as class scores and induces the correct class to have a higher score than those of the other classes. In addition, it is considered as very efficient for a binary classification.

In order to decrease the false positive rate and improve our experiment results, we decided to use the SVM classifier instead of the softmax. Series of experiments were carried out while varying several parameters to determine the most suitable classifier. Hence, we found that the linear SVM with $C = 0.1$ is the most appropriate for the classification of pulmonary nodules. As shown in Table 3, the accuracy of the system was enhanced from 91.62–91.99% and from 91.46–91.84% for [VGG16, VGG19] and $[VGG16]^2$ architectures, respectively, when we used the linear SVM classifier. However, it was decreased from 91 to 90.58% for the $[VGG19]^2$ structure when we used the SVM. We can, therefore, report that, in some cases, bilinear CNN may be more efficient for nodule classification than when coupled with an SVM. Additionally, the sensitivity was enhanced from 91.94–93.28%, from 91.73–92.29% and from 91.81–91.85% for $[VGG16]^2$, $[VGG19]^2$ and [VGG16, VGG19] architectures, respec-

Table 5 Classification performances of BCNN models with various classifiers

BCNN Model	Classifier	Sensitivity (%)	Specificity (%)	F1-score (%)	Accuracy (%)	FP/scan
$[VGG16]^2$	RF	91.89	88.70	92.69	90.74	0.113
	KNN	91.81	87.18	92.15	90.11	0.128
	AdaBoost	92	86.50	92	89.95	0.135
	SVM	**93.28**	**89.36**	**93.52**	**91.84**	**0.081**
$[VGG19]^2$	RF	90.27	87.17	91.49	89.16	0.128
	KNN	90.55	84.68	90.77	88.38	0.153
	AdaBoost	91.48	85.29	91.36	89.17	0.147
	SVM	**92.29**	**87.66**	**92.52**	**90.58**	**0.123**
[VGG16, VGG19]	RF	91.28	89.73	92.74	90.74	0.102
	KNN	91.11	87.61	92.40	90.42	0.124
	AdaBoost	91.60	87.50	92.17	90.11	0.125
	SVM	**91.85**	**92.27**	**93.76**	**91.99**	**0.077**

tively, when we used the linear SVM classifier. Moreover, the FP of the models was decreased from 9.28–8.14% and from 8.63–7.72% for $[VGG16]^2$ and [VGG16, VGG19] architectures, respectively, with the SVM classifier. However, it was increased from 10.18 to 12.29% for the SVM-$[VGG19]^2$ structure. In addition, the ROC curve demonstrated the outperformance of the [VGG16, VGG19] combination compared to $[VGG16]^2$ and $[VGG19]^2$ architectures achieving an AUC rate of 95.9% against 94.8% and 94%, respectively.

We noticed that the performance of BCNN [VGG16, VGG19] with and without SVM excelled over the other two BCNN structures. The outcomes indicated that a BCNN model with two different architectures could provide greater performance. This may be due to a low correlation between the outputs of feature extractors with different structures. Thus, this can be effective in increasing the discriminating ability, leading to more consistent classification accuracy. Moreover, we compared our results to those of conventional CNN models as shown in Table 4. Since the bilinear CNN generates wealthy representations compared to conventional CNN models, our three BCNN structures presented promising results and outperformed existing state-of-the-art models.

We evaluated the performance of various other classifiers including RF, KNN and AdaBoost coupled to BCNN models using the testing datasets. The obtained results are summarized in Table 5.

As shown in Table 5, the best performances were obtained with the SVM classifier. As shown in Table 5, the proposed SVM-$[VGG16]^2$ model, surpassed the RF classifier by 1.39%, 0.66%, 0.83% and 1.1%, in terms of sensitivity, specificity, F1-score and accuracy, respectively, and the FP rate decreased by 3.16% which presented a significant improvement with ($p<0.01$) in a one-tailed z-test. The model also surpassed the KNN and AdaBoost classifiers by 1.47% and 1.28%, respectively, in terms of sensitivity, by 2.18% and 2.86% in terms of specificity, by 1.37% and 1.52% in terms of F1-score and by 1.73% and 1.89% in terms of accuracy, respectively. This also represented a significant improvement with ($p<0.05$) in a one-tailed z-test for accuracy and specificity, but there is no significant improvement with ($p<0.05$) for the sensitivity and F1-score. FP rate decreased by 4.66% and 5.36% for KNN and AdaBoost classifiers, respectively, and this also presented a significant improvement with ($p<0.01$) in a one-tailed z-test. Again, the proposed SVM-$[VGG19]^2$ model outperformed the RF, KNN and AdaBoost classifiers in terms of sensitivity by 2.02%, 1.74% and 0.48%, respectively, which presented a significant improvement with ($p<0.05$) over the RF and KNN classifiers, but there is no significant improvement over the AdaBoost classifier ($p<0.05$) in a one-tailed z-test. It also outperformed them in terms of specificity by 0.49%, 2.98% and 2.37%, respectively, F1-score by 1.03%, 1.75% and 1.16%, respectively, and accuracy by 1.42%, 2.2% and 1.41%, respectively. FP rate decreased by 0.5%, 3% and 2.4% over RF, KNN and AdaBoost classifiers, respectively. This is represented as significant improvement with ($p<0.01$) in a one-tailed z-test. Similarly, the proposed SVM-[VGG16, VGG19] model outperformed the RF classifier by 0.57%, 2.54%, 1.02% and 1.25%, in terms of sensitivity, specificity, F1-score and accuracy, respectively, and the FP rate decreased by 2.48% which presented a significant improvement with ($p<0.01$) in a one-tailed z-test. Also, the proposed model surpassed the KNN and AdaBoost classifiers by 0.74% and 0.25%, respectively, in terms of sensitivity (no significant improvement ($p<0.05$) in a one-tailed z-test), by 4.66% and 4.77% in terms of specificity (significant improvement ($p<0.01$) in a one-tailed z-test), by 1.36% and 1.59% in terms of F1-score (no significant improvement ($p<0.05$) in a one-tailed z-test) and by 1.57% and 1.88% in terms of accuracy, respectively, which presented a significant improvement

over the AdaBoost classifier ($p<0.05$). FP rate decreased by 4.68% and 4.78% for KNN and AdaBoost classifiers, respectively, and this also presented a significant improvement with ($p<0.01$) in a one-tailed z-test.

Conclusion

In this paper, we proposed a lung nodule classification system using three BCNN models, followed by a linear SVM classifier. BCNN models are built using two of the best-known pre-trained CNNs, which are VGG16 and VGG19. The fine-tuning process has been performed for the proposed model's adjustments. Aiming to improve classification results, we trained the SVM classifier instead of the original softmax. We investigated various SVM types to find the most appropriate one for our classification problem. Experimental results indicated that the BCNN architecture combining VGG16 and VGG19 networks, followed by the linear SVM, outperformed the other two architectures in terms of accuracy and AUC. The performance results were very promising and surpassed those of existing works.

Limitations and future works

The bilinear pooling is high dimensional, and it generally contains several thousand features, which often requires massive calculation and storage costs and makes it impractical for further analysis. This is the main drawback of the bilinear CNN implementation. In the future, we plan to overcome this limitation by adopting the idea proposed in [24] allowing to compact the bilinear CNN by polynomial kernel approximation of random projection of the bilinear descriptor or by kernel approximation of Random Maclaurin and Tensor Sketch, as proposed in [25]. Moreover, it has turned out that the BCNN model cannot handle the information inherent in different convolutional layers [26]. Thus, we propose in the future work to overcome this limitation by using the bilinear pooling function in the outputs of cross-convolutional layers of a single CNN. This is inspired by the forward propagation of ResNet [27] architecture, which is characterized by cross-layer connections.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Informed consent This article does not contain patient data.

References

1. Cancer Facts and Figures (2020) Atlanta: American Cancer Society. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2020.html. Accessed 06 June 2020
2. Lung Cancer Fact Sheet (2020). American Lung Association. https://www.lung.org/lung-health-and-diseases/lung-disease-lookup/lung-cancer/learn-about-lung-cancer/lung-cancer-factsheet.html. Accessed 06 June 2020
3. Mastouri R, Khlifa N, Neji H, Hantous-Zannad S (2020) Deep learning-based CAD schemes for the detection and classification of lung nodules from CT images: a survey. J X Ray Sci Technol 28(4):591–617. https://doi.org/10.3233/XST-200660
4. Ben Jabra M, Guetari R, Chetouani A, Tabia H, Khlifa N (2020) Facial expression recognition using the bilinear pooling. In: 15th international joint conference on computer vision, imaging and computer graphics theory and applications (VISIGRAPP 2020), pp 294–301. https://doi.org/10.5220/0008928002940301
5. Monkam P, Qi S, Xu M, Han F, Zhao X, Qian W (2018) CNN models discriminating between pulmonary micro-nodules and non-nodules from CT images. BioMed Eng OnLine 17(1):96. https://doi.org/10.1186/s12938-018-0529-x
6. Monkam P, Qi S, Xu M, Li H, Han F, Teng Y, Qian W (2019) Ensemble learning of multiple-view 3D-CNNs model for micro-nodules identification in CT images. IEEE Access 7:5564–5576
7. Shen W, Zhou M, Yang F, Yu D, Dong D, Yang C, Zang Y, Tian J (2017) Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. Pattern Recogn 61:663–673
8. Onishi Y, Teramoto A, Tsujimoto M, Tsukamoto T, Saito K, Toyama H, Imaizumi K, Fujita H (2019) Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks. Biomed Res Int 6:1–9. https://doi.org/10.1155/2019/6051939
9. Wu P, Sun X, Zhao Z, Wang H, Pan S, Schuller B (2020) Classification of lung nodules based on deep residual networks and migration learning. Comput Intell Neurosci. https://doi.org/10.1155/2020/8975078
10. Kaya A (2018) Cascaded classifiers and stacking methods for classification of pulmonary nodule characteristics. Comput Methods Programs Biomed 166:77–89
11. Kido S, Hirano Y, Hashimoto N (2018) Detection and classification of lung abnormalities by use of convolutional neural network (CNN) and regions with CNN features (R-CNN). Int Workshop Adv Image Technol. https://doi.org/10.1109/iwait.2018.8369798
12. Liu J, Yang Z, Zhang T, Xiong H (2017) Multi-part compact bilinear CNN for person re-identification. In: 2017 IEEE international conference on image processing (ICIP)
13. Ustinova E, Ganin Y, Lempitsky V (2015) Multiregion bilinear convolutional neural networks for person re-identification. arXiv preprint arXiv:1512.05300
14. Chen H, Wang J, Qi Q, Li Y, Sun H (2017) Bilinear CNN models for food recognition. In: 2017 International conference on digital image computing: techniques and applications (DICTA)
15. Wang C, Shi J, Zhang Q, Yin S (2017) Histopathological image classification with bilinear convolutional neural networks. In: 39th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)
16. LUng Nodule Analysis (2016). Grand challenge. https://luna16.grand-challenge.org/Data/. Accessed 26 Feb 2019
17. Wang W, Luo J, Yang X, Lin H (2015) Data analysis of the lung imaging database consortium and image database resource initia-

tive. Acad Radiol 22(4):488–495. https://doi.org/10.1016/j.acra.2014.12.004
18. Lin TY, RoyChowdhury A, Maji S (2015) Bilinear cnn models for fine-grained visual recognition. In: Proceedings of the IEEE international conference on computer vision, pp 1449–1457
19. Simonyan K; Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations (ICLR). arXiv:1409.1556
20. Shi Z, Hao H, Zhao M, Feng Y, He L, Wang Y, Suzuki K (2018) A deep CNN based transfer learning method for false positive reduction. Multimed Tools Appl. https://doi.org/10.1007/s11042-018-6082-6
21. Zhao X, Liu L, Qi S, Teng Y, Li J, Qian W (2018) Agile convolutional neural network for pulmonary nodule classification using CT images. Int J Comput Assist Radiol Surg 13(4):585–595. https://doi.org/10.1007/s11548-017-1696-0
22. Zhao X, Qi S, Zhang B, Ma H, Qian H, Yao Y, Sun J (2019) Deep CNN models for pulmonary nodule classification: model modification, model integration, and transfer learning. J X-Ray Sci Technol. https://doi.org/10.3233/xst-180490
23. Shen S, Han SX, Aberle DR, Bui AA, Hsu W (2019) An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification. Expert Syst Appl 128:84–95. https://doi.org/10.1016/j.eswa.2019.01.048
24. López-Sánchez D, Arrieta AG, Corchado JM (2020) Compact bilinear pooling via kernelized random projection for fine-grained image categorization on low computational power devices. Neurocomputing 398:411–421. https://doi.org/10.1016/j.neucom.2019.05.104
25. Gao Y, Beijbom O, Zhang N, Darrell T (2016) Compact bilinear pooling. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 317–326
26. Sun Q, Wang Q, Zhang J, Li P (2018) Hyperlayer bilinear pooling with application to fine-grained categorization and image retrieval. Neurocomputing 282:174–183
27. Moussa O, Khachnaoui H, Guetari R, Khlifa N (2019) Thyroid nodules classification and diagnosis in ultrasound images using fine-tuning deep convolutional neural network. Int J Imaging Syst Technol 30(1):185–195. https://doi.org/10.1002/ima.22363