The Veterinary Journal 237 (2018) 43–48

Contents lists available at ScienceDirect

The Veterinary Journal

journal homepage: www.elsevier.com/locate/tvjl

Original Article

Prediction of radiographic abnormalities by the use of bag-of-features and convolutional neural networks

Y. Yoon, T. Hwang, H. Lee*

Institute of Animal Medicine, College of Veterinary Medicine, Gyeongsang National University, Jinju 30488, South Korea

ARTICLE INFO

Article history:
Accepted 25 May 2018

Keywords:
Computer aided detection
Convolutional neural networks
Dog
Machine learning
Thoracic radiography

ABSTRACT

This study evaluated the feasibility of bag-of-features (BOF) and convolutional neural networks (CNN) for computer-aided detection in distinguishing normal from abnormal radiographic findings. Computed thoracic radiographs of dogs were collected. For the purposes of this study, radiographic findings were used to distinguish between normal and abnormal in the following areas: (1) normal cardiac silhouette vs. cardiomegaly, (2) normal lung vs. abnormal lung patterns, (3) normal mediastinal position vs. mediastinal shift, (4) normal pleural space vs. pleural effusion, and (5) normal pleural space vs. pneumothorax. Images for training and testing the models consisted of ventrodorsal and lateral projection images of the same scale. The number of images used for each finding are as follow: 3142 for cardiomegaly (1571 normal and 1571 abnormal from 1143 dogs), 2086 for lung pattern (1043 normal and 1043 abnormal from 1247 dogs), 892 for mediastinal shift (446 normal and 446 abnormal from 387 dogs), 940 for pleural effusion (470 normal and 470 abnormal from 284 dogs), and 78 for pneumothorax (39 normal and 39 abnormal from 61 dogs). All data samples were divided so that 60% would be used for training the algorithms and 40% for testing the two models. The performance of the classifiers was evaluated by calculating the accuracy, sensitivity and specificity.

The accuracy of both models ranged from 79.6% to 96.9% in the testing set. CNN showed higher accuracy (CNN; 92.9–96.9% and BOF; 79.6–96.9%) and sensitivity (CNN; 92.1–100% and BOF; 74.1–94.8%) than BOF. In conclusion, both BOF and CNN have potential to be useful for improving work efficiency by double reading.

Introduction

Computers are used in diagnostic imaging for image acquisition, management, storage, and reporting. Computer algorithms have also been developed and used in human clinical practice for what is commonly called computer-aided detection (CAD). CAD is a technology combining elements of machine learning and computer vision with medical imaging. The primary goal of CAD is to support the detection of disease and provide a consistent and reproducible second opinion for clinicians. This strategy of double reading can potentially improve workflow efficiency by reducing the rate of false negatives due to observational oversights (Castellino, 2005; Jiang et al., 2015; AlZubaidi et al., 2017). Previous CAD studies focused on the differentiation of diseases (Kadah et al., 1996; Sujana et al., 1996; Chen et al., 1998; Chen et al., 2013). These CAD algorithms usually detected the disease based on a specific radiological finding.

Most studies of CAD include the following steps: cropping the region of interest (ROI) from the medical image, extracting features from the ROI, selecting features, training algorithms using the features, and making predictions using the training algorithms (Baxt, 1991; Chen et al., 1998; Vlahou et al., 2003; Polat and Güneş, 2007; Akay, 2009; Malon and Cosatto, 2013; Roth et al., 2014; Wang et al., 2014). The role of the supervisor who labels the class of training data is important in this process because supervisor's subjective standards can influence the ROI, the method for feature extraction and selection, and ultimately the performance of the CAD.

Machine learning is the study that constructs algorithms which can learn from and make predictions on data. Artificial neural networks, a set of well established machine learning algorithms, are models that solve problems and represent information in a way analogous to that of a human brain (Baxt, 1991; Sujana et al., 1996). A basic operating unit in artificial neural networks is a neuron-like node. Each connection between nodes can transmit numbers from

* Corresponding author.
E-mail address: lhc@gnu.ac.kr (H. Lee).

https://doi.org/10.1016/j.tvjl.2018.05.009

one to another, and the networks solve the problem by changing the binding strength of nodes. Several machine learning algorithms that require large memory usage have received attention due to the increase in computing power now available (Csurka et al., 2004; Bay et al., 2008; Yang et al., 2009; Ciresan et al., 2011; Krizhevsky et al., 2012; Jiang et al., 2015; AlZubaidi et al., 2017). Bag-of-features (BOF) (Csurka et al., 2004) and convolutional neural networks (CNN) (Ciresan et al., 2011; Krizhevsky et al., 2012) both represent such algorithms, using the image itself as the input rather than the numerical values used in conventional algorithms (Kadah et al., 1996; Sujana et al., 1996; Chen et al., 1998; Fei-Fei and Perona, 2005; McEvoy and Amigo, 2013). BOF extracts features from images, such as the margin, shape, alignment, and so forth, and classifies them by their occurrence (Csurka et al., 2004; Fei-Fei and Perona, 2005). Several local patches of image are extracted as numerical vectors and defined as visual words by clustering similar patches, which is called feature extraction. Based on the occurrence of visual words, a histogram is generated. A classifier is then trained by matching the histogram and predicts categories of general untrained images. CNN is similar to ordinary neural networks in that they are made up of layers that have learnable weights and biases, but differs in that it has a convolutional layer (CONV) encoding the features of the input image (Ciresan et al., 2011; Krizhevsky et al., 2012). CONV applies various filters to local patches of the image, and layers of other functions tile the outputs to obtain a better representation from the image. Repeated tiling and various combinations of layers reduces the amount of non-critical information thereby improving the ability to generalize. Local connectivity and weight-sharing between layers then allows for training of the networks. BOF and CNN can both extract and learn the features of images and thereby classify the images (Csurka et al., 2004; Yang et al., 2009; Ciresan et al., 2011; Krizhevsky et al., 2012). This suggests that both BOF and CNN can be used for radiologic interpretation.

In this study, the goal was to discriminate normal vs. abnormal radiographic findings using BOF and CNN and to evaluate their performance.

Materials and methods

A computer with Microsoft Windows 10 (64 bit), an Intel Core i7 4.9 gigahertz central processing unit, 32 gigabyte random-access memory, and a NVIDIA Quadro M4000 graphics card was used for this study. The algorithms for image acquisition and machine learning were coded in MATLAB (R2016b, MathWorks). The toolboxes used in MATLAB were computer vision system, curve fitting, data acquisition, global optimization, image acquisition, image processing, neural network, optimization, parallel computing, and statistics and machine learning.

Computed thoracic radiographs with normal and abnormal radiographic findings from dogs were collected retrospectively from a database at the Veterinary Medical Teaching Hospital of Gyeongsang National University from 2012 to 2016. A direct digitizer Regius190 (Konica Minolta) was used to image radiographs (50–70 kVp, 300 mA, and 6.0 mAs). Three veterinary radiologists (one veterinary radiology Ph.D. and two veterinary radiology masters) evaluate the radiographs individually, and the radiographs were excluded if there was a disagreement in individual interpretations. Images with inappropriate posture or severe rotation were excluded. Based on subjective interpretations of the radiographs (Table 1), each image was classified as being normal or abnormal with regard to five specific findings: (1) normal cardiac silhouette vs. cardiomegaly, (2) normal lung vs. abnormal lung patterns, (3) normal mediastinal position vs. mediastinal shift, (4) pleural space without pleural effusion vs. pleural effusion, and (5) pleural space without pneumothorax vs. pneumothorax (Fig. 1). The images were included if the distinction between normal and abnormal for findings could be made even if other findings were identified. The number of normal and abnormal images was the same for each finding; this was done to avoid problems arising from imbalanced data, such as a false trained model that highly favors the overrepresented class (Seiffert et al., 2008). All of the selected data samples were divided so that 60% would be used for training the algorithms and 40% for testing the two models. Follow-up images of the same patient were included to increase the number of training data images, and no image from patients in the training set was included in the testing set for each finding. Table 2 presents in detail the images that met the above criteria.

Lateral (LAT) and ventrodorsal (VD) images were used to discriminate cardiomegaly, abnormal lung patterns, pleural effusion, and pneumothorax. VD images were used to classify the mediastinal shift. Each image of different size was resized and cropped to include the entire thorax because using images of the same size as the image used for the actual reading required a prolonged training time or made computation impossible in the system specification of this study. We therefore used a smaller sized image that was the largest possible size for computation through trial and error to optimize performance in a pilot study. The final sizes of the images used in BOF models (five binary classification models) were as follow: VD Image 120×160 pixels, and VD-LAT combined Image 320×160 pixels. The final size of the VD and VD-LAT combined images used in CNN models (five binary classification models) was 100×50 pixels.

The configuration of the BOF for each finding (five binary classification models) was the same. A detailed configuration and process of the BOF model is shown in Fig. 2. Speed-up robust features (SURF), a local feature detector and descriptor, was used to extract features. Local patches of image made by grid were extracted as features. The features were defined as visual words by *k*-mean clustering. Based on the occurrence of visual words, a histogram is generated. Binary support vector machine, learning algorithms that analyze data used for classification, was trained by matching the histogram and predict categories of general untrained images (Fei-Fei and Perona, 2005; Escalera et al., 2010).

The configuration of the CNN for each finding (five binary classification models) was the same. CNN model consisted of a CONV to output multi-channel data by applying various filters, rectified linear unit layers (ReLU) to determine the degree of activation of the input value, channel-wise local response normalization layers (NORM) to put the input values on the same scale, max pooling layers (MaxP) to reduce the number of parameters, fully connected layers (FullC) to multiply the input by a weight matrix and then adds a bias vector, a drop out layer (DropO) that randomly set the input elements to zero with a given probability, and a softmax layer (SoftM) that

Table 1
Criteria for radiographic findings.

	Criteria
Normal cardiac silhouette	Vertebral heart score 10.5 or less on lateral view without chamber dilation
Cardiomegaly	Vertebral heart score 11.5 or higher on lateral view
	Right atrium and/or right ventricle enlargement on ventrodorsal and lateral views
	Left atrium and/or left ventricle enlargement on ventrodorsal and lateral views
Normal lung	Lucent lung lobe with little interstitial lung tissue and vessels
Abnormal lung pattern	Alveolar or interstitial or bronchial patterns on ventrodorsal and lateral views
Normal mediastinal position	Located in the middle of the thoracic cavity
Mediastinal shift	Left-sided or right-sided shift on ventrodorsal view
Pleural space without pleural effusion[a]	No fissure lines and retraction of the lung lobes away from the thoracic wall
Pleural effusion	Prominent fissure lines with retraction of the lung lobes away from the thoracic wall and a fluid-filled pleural space on ventrodorsal and lateral views
Pleural space without pneumothorax[b]	No retraction of lung lobes away from the thoracic wall
Pneumothorax	Retraction of lung lobes away from the thoracic wall and a lack of vascular and bronchial markings beyond the border of the collapsed lobes on ventrodorsal and lateral view

[a] Pleural space to distinguish pleural effusion.
[b] Pleural space to distinguish pneumothorax.