IET Image Processing

Special Section: Multidisciplinary advancement of imaging technologies: from medical diagnostics and genomics to cognitive machine vision, and artificial intelligence

Deep learning-based automated detection of human knee joint's synovial fluid from magnetic resonance images with transfer learning

ISSN 1751-9659
Received on 3rd December 2019
Revised 22nd April 2020
Accepted on 12th May 2020
E-First on 15th July 2020
doi: 10.1049/iet-ipr.2019.1646
www.ietdl.org

Imran Iqbal[1], Ghazala Shahzad[2], Nida Rafiq[3], Ghulam Mustafa[4], Jinwen Ma[1] ✉

[1]*Department of Information and Computational Sciences, School of Mathematical Sciences and LMAM, Peking University, Beijing 100871, People's Republic of China*

[2]*Department of Radiology, Liaquat University of Medical and Health Sciences, Jamshoro 76090, Pakistan*

[3]*Dow Institute of Radiology, Dow University of Health Sciences, Karachi 74200, Pakistan*

[4]*Department of Biomedical Engineering, College of Engineering, Peking University, Beijing 100871, People's Republic of China*

✉ *E-mail: jwma@math.pku.edu.cn*

Abstract: As an analytic tool in medicine, particularly in radiology, deep learning is gaining much attention and opening a new way for disease diagnosis. Nonetheless, it is rather challenging to acquire large-scale detailed labelled datasets in the field of medical imaging. In fact, transfer learning provides a possible way to resolve this issue to a certain extent such that the parameter learning of a neural network starts with its pre-trained weights learned from a large-scale dataset of certain similar task, and fine-tunes on a small comprehensively annotated dataset for the particular target task. The main aim of this study is to apply the deep learning model to detect the synovial fluid of human knee joint from magnetic resonance images. A specialized convolutional neural network architecture is proposed for automated detection of human knee joint's synovial fluid. Two independent datasets are used in the training, development, and evaluation of the proposed model. It is demonstrated by the experimental results that the proposed model obtains high sensitivity, specificity, precision, and accuracy to the detection of human knee joint's synovial fluid. As a result, this proposed approach provides a novel and feasible way for automating and expediting the synovial fluid analysis.

1 Introduction

Object detection has been an important and challenging task in the field of computer vision for the last two decades. It is a process of finding an instance of object in an input image from one of the given classes. If there exists such an instance of the object in the image, the task of object detection is to determine its location or bounding box and category [1]. There are various applications of object detection such as self-driving vehicles, robot vision, and human–machine interaction. As the object detection tasks become more complicated, it is rather difficult for conventional machine learning algorithms to extract effective features from training images. Nonetheless, in recent years, deep learning has developed smart techniques for extracting effective features from images automatically [2]. Krizhevsky *et al.* [3] developed the powerful deep neural network architecture, AlexNet, which attained the excellent performance on ImageNet challenge for classification task [4]. After that, several deep neural network architectures for object detection have been established to achieve much better results, such as Inception [5]. Actually, significant advancement has been made on object detection mainly due to the improvement of architecture of convolutional neural networks (CNNs), large-scale comprehensively labelled datasets, and innovation of processing devices. Thus, deep learning can provide an effective way for detection of human body organs and their tissues and the objective of this study is to apply it to the detection of synovial fluid in human knee joints from magnetic resonance (MR) images.

In anatomy, knee joint is the most complex joint in human body, which appears like a simple hinge joint in which the bones are so articulated as to allow extensive movement in one plane. A knee joint consists of three parts: patella, tibia, and femur. All of these are line with the cartilage that yields less friction surface to achieve different types of body movements. It is encircled by a capsule producing synovial fluid, which serves as a joint lubricating fluid, so that the knee joint can move in an efficient way [6]. Synovial fluid is a viscoelastic liquid emerged in the synovial joints' cavities. This fluid serves as lubricating agent to lessen friction during movement between the cartilages of synovial joints [7]. It also provides nutrition to the articular cartilage to the bones of the joint. This fluid moves into the articular cartilage when the joint is in rest and keep a thin layer on the articular cartilage surface when the joint is active.

Human body makes an additional quantity of synovial fluid in knee joint due to diseases or injury sometimes; blood, pus, or crystalline substance may be added into the synovial fluid, aggravating the joint and affecting knee to swell. Surgeons may suggest knee MR image scan if they suspect abnormalities such as an accumulation of fluid, baker's cyst, with an unknown cause within knee joint. So, it has various significant clinical implications to detect and analyse the synovial fluid of knee joints of a patient. Actually, this helps diagnose the reason of joint inflammation. In joint diseases like arthritis, the synovium of the joint is the main place where inflammation appears.

In order to do so, the MR imaging system with powerful magnet, radio frequency pulses, and a computer can be utilised to reconstruct a comprehensive picture of organs and soft tissues within human bodies. In fact, it can be used for body examination together with anomalies of the head and backbone injuries and abnormalities of the joints such as the hip and knee, heart problems such as congenital heart disease. Consequently, the MR images are valuable to prevent diseases, help diagnoses, prescribe drugs, and evaluate how potent earlier medications have been [8]. Due to the wide application and excellent evaluation performance of MR imaging, X-ray and computed tomography have been abandoned to examine most of knee and hip joints' diseases. Organs and tissues can be observed clearly via MR imaging, such as fibrocartilage and hyaline. Cartilage and synovial-based diseases can also be detected with the help of it [9]. MR imaging further supports the prompt diagnosis and treatment of abnormal feature of femoral head in

osteonecrosis as a result of its high contrast as well as good resolution features [10]. Hence, it is an effective way to use the MR image for detecting the synovial fluid. Indeed, surgeons often analyse and diagnose the cartilage and synovial-based diseases with the help of MR images. However, to the best of our knowledge, there is no automated and effective synovial fluid detection system available in medical centres. In fact, an effective and efficient synovial fluid system can save time, decrease subjectivity, and inaccuracy due to the tiredness of medical specialists. This type of automatic system can greatly lessen the workload of radiologists.

The objective of this research work is to detect the synovial fluid of human knee joint automatically (rather than human) with high accuracy from MR images. First, Digital Imaging and Communications in Medicine (DICOM) files of human knee joint are acquired from Shanghai Key Laboratory of Orthopaedic Implants [11] for training the models, and from PC Hospital Liaoning [12] for development and testing the models. As a matter of fact, these datasets are annotated by the radiologists with the help of PixelAnnotationTool program [13]. A specialised CNN architecture is then proposed to solve this problem effectively and efficiently. In fact, this network is initialised with the pre-trained weights obtained on the Microsoft common objects in context (COCO) dataset [14] via transfer learning. These weights are further updated through the fine-tuning on the specific training set. In this way, it can learn new features which are more precise to the synovial fluid detection. Furthermore, the hyperparameters are tuned to improve the performance of the proposed model. Specifically, the hyperparameters are selected according to the lowest average loss on the development or validation set. Finally, the proposed model is evaluated on the test set.

For comparison, four state-of-the-art CNN architectures are used: Inception-v2, ResNet-50, Inception-ResNet-v2, and ResNet-101, being implemented in the same way as above, to make the synovial fluid detection. Specifically, the proposed model is compared with these four baseline models in terms of sensitivity, specificity, precision, accuracy, and evaluation time. The MR image annotations for the two datasets of human knee joint are performed by two co-authors (G. S. and N. R.), both have more than 10 years of experience in radiology field.

The remainder of the paper is organised as follows. Section 2 concisely explains recent medical applications using deep learning. Section 3 briefly describes the detail of the proposed architecture and baseline models which are used in this research work. Section 4 presents the experimental results of the proposed CNN architecture as well as comparison with the baseline models. Finally, Section 5 concludes the paper.

2 Deep learning to healthcare

Recently, deep learning is totally reshaping healthcare as an industry. In fact, fast-paced developments of deep neural networks have made progresses on medical imaging in an incredible way. CNN-based healthcare aided systems can now offer medical professional precise disease diagnosis and even assist doctors to make better medical decisions. Although certain conventional machine learning algorithms have been adopted in some healthcare aided systems, they need human experts to design the feature extractors which convert raw data into appropriate feature representations. On the contrary, deep learning is a form of feature learning and representation in which a machine receives raw data and creates their own feature representations needed to detect the patterns [15], which is more beneficial to the healthcare aided systems than as conventional machine learning. In the following paragraphs of this section, the major related works of deep learning to healthcare are reviewed.

The classification of heartbeats via electrocardiogram is a time-consuming and challenging task because there always exists certain noise which corrupts the actual signals. In fact, the heartbeats can be classified into five different classes namely fusion, non-ectopic, ventricular ectopic, supraventricular ectopic, and unknown beats. Acharya *et al.* [16] developed six layers' network architecture to classify arrhythmia and normal heart rhythm. Javadi and Mirroshandel [17] further developed a deep model of 26 convolutional layers based on Visual Geometry Group (VGG) network [18] for assessment of human sperm. They prepared 1540 sperm image collection, MHSMA dataset, from 235 subjects with male factor infertility. Their experimental results showed good accuracy for morphological deformities detection. In another study, Hernandez *et al.* [19] investigated two to eight densely connected layers' networks to predict the force capabilities using the force feasible set of the human upper-limbs. They created 17 musculoskeletal models based on anthropometric data of right upper-limb. Their method could be used smoothly with real-time feedback to evaluate human gesture and hand force control for demanding tasks.

On the other hand, the deep network models trained on medical images could reach the dermatologist-level result on identifying skin malignancy. In fact, the recent deep learning methods excelled the ordinary dermatologist in a contrast of predictions and the judgements by a team of dermatologists on a photographic images set. In such a work [20], the Inception network [21] was used for the classification of skin lesions using a dataset of more than hundred thousand clinical images. The doctors and machine learning experts trained the model via retinal photographs to detect diabetic retinopathy and diabetic macular oedema. In that research work, the achieved results were very close to the performance of eye specialist [22]. That group also showed that CNNs can extract earlier unrecognised association between retinal image patterns and age, sex, smoking status, and systolic blood pressure [23]. Both of these retinal fundus photographs researches used Inception network architecture [21] to train the model. Chang *et al.* [24] introduced the SCIAN dataset which have about 2000 images for classification of human sperm heads for semen analysis. In that work, they used four classification techniques namely K-nearest neighbours [25], naive-Bayes [26], decision trees [27], and support vector machine [28] for classification of human sperm heads into five different categories. Another research [29] related to human sperm classification used VGG-based network. They achieved 72 and 62% classification accuracy in 3-out-of-3 and at least 2-out-of-3 expert agreement datasets, respectively. That approach shows the potential of deep learning technologies to surpass specialists in terms of reliability and accuracy.

All of the above works in the field of healthcare utilise the deep learning approaches to make a remarkable success in medical analysis and diagnosis. They certainly open new ways for automatic medical diagnosis. However, these deep learning models such as Inception network are not so beneficial to extract effective features of synovial fluid of human knee joint from MR images due to the complexity and specificity of this problem. In order to solve this particular synovial fluid detection problem, a specialised deep CNN architecture is designed and presented in the next section.

3 Models and methods

Inception [5] is a well-known deep network architecture. When it needs to design a layer for an architecture, kernel size should be carefully selected such as 5-by-5 or 3-by-3, or 1-by-1. That network uses different kernel sizes and concatenate the results, which makes the architecture of that network more powerful for classification and regression tasks. As a matter of fact, the Residual Network (ResNet) [30] is a deep neural network architecture which can handle more sophisticated learning tasks. Usually, it faces two problems with a learning task. First, the accuracy can saturate, while the degradation may arise, which is triggered by the overfitting. Second, there may appear the situation of vanishing or exploding gradient, that is, the gradient becomes vanishingly small or explodingly large to prevent the weights from modifying their values. Fortunately, ResNet can deal with these problems by using skip connections, which uses the residual mapping to preserve the inputs. Xception network [31] uses the modified form of depthwise separable convolutions. It is actually inspired by Inception-v3 in which first apply pointwise convolution then use a depthwise convolution. Therefore, that neural network architecture has better ImageNet validation accuracy with less number of parameters in comparison with Inception-v3. Another renowned deep neural

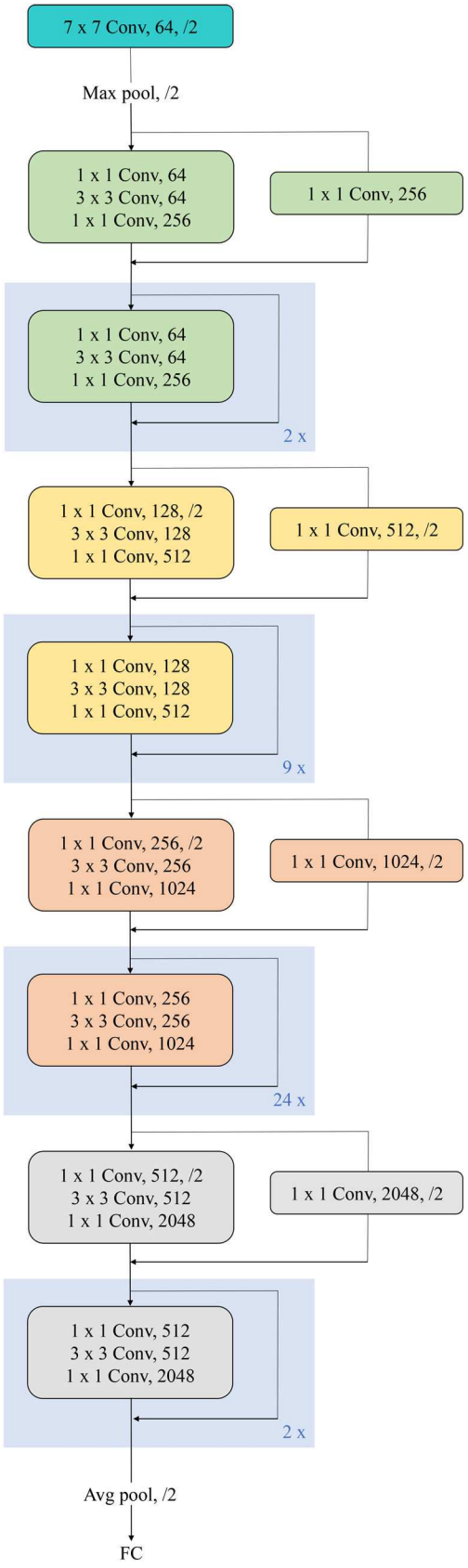

Fig. 1 *Layout of the proposed CNN architecture. Note that all convolutional layers are followed by ReLU [33] and batch normalisation [34] (not included in the diagram). 'FC' signifies fully connected layer*

network architecture is Inception-ResNet-v2 [32]. That network is a combination and modification of Inception and ResNet networks. As it uses skip connections and several kernel sizes, that network attains better accuracy in a smaller number of epochs in comparison with Inception-v4.

Being inspired by these advanced deep learning architectures, a specialised CNN architecture is proposed to solve the problem of automatic detection of synovial fluid in human knee joints, which is a difficult task even for experienced radiologist. To tackle this problem, the proposed architecture is composed of four main parts. In the first part, there are 9 convolutional layers for detecting simple features such as features of femur and tibia. In the second part, there are 30 convolutional layers for detecting complex patterns such as patella and bursa patterns. In the third part, there are further 75 convolutional layers often looking for more complex features such as features of synovial membrane. In the last part, there are finally 9 convolutional layers for learning features which are quite precise to describe the synovial fluid. There are total 125 convolutional layers in this architecture. After every three convolutional layers, there is a residual mapping to preserve the inputs and make the architecture more effective for this task. All convolutional layers are followed by rectified linear units (ReLU) [33] and batch normalisation [34]. Fig. 1 illustrates this architecture's layout schematically in detail. 'x' with prefix 2, 9 and 24 signify that that block repeat prefix times.

In various medical fields, such as medical imaging and oncology, it is hard to create a large-scale comprehensively labelled dataset due to the data confidentiality and expensive labelling. However, transfer learning [35] is a machine learning technique such that a network trained on one problem is re-purposed on another similar problem. It is an enhancement of learning in a new problem through the knowledge transfer from a related problem that has already been learned. The pre-trained neural network model can be fine-tuned on a medium or small medical dataset so that it can converge quicker and learn the problem-specific features. With this tactic, the weights of the network can be effectively and efficiently learned in the course of fine-tuning step from the pre-trained weights, allowing the network to learn the features precise to the new problem. In recent years, various research works have shown that fine-tuning is capable for a range of problems in the field of medicine [36]. According to this fact, the weights of the proposed model can be initialised with the learned values obtained on the COCO dataset via transfer learning. Actually, the COCO dataset is an outstanding large-scale image dataset aimed to help the advanced research in object recognition, image captioning, segmentation, keypoint detection, and scene understanding. It was an initiative to collect images that reflect everyday scene and provides contextual information. In natural images, several objects can be found in one image and each must be labelled as a different object and segmented appropriately. This dataset has 91 classes, and about 2.5 million labelled object instances in more than 300 thousand images.

TensorFlow object detection application programming interface (API) [37] is a framework for using TensorFlow [38] deep learning library that allows to train the model and also evaluate it with ease. The COCO dataset is used to train this API. For comparison, four pre-trained well-known models are used from this API: Inception-v2, ResNet-50, Inception-ResNet-v2, and ResNet-101, which are considered as the baseline model-1, baseline model-2, baseline model-3, and baseline model-4, respectively. Actually, these four baseline models use the Mask RCNN [39], Atrous Convolutions [40], and COCO dataset for training. The pre-trained models are fine-tuned using TensorFlow deep learning platform with the specific training set.

Fig. 2 shows the complete pipeline of the proposed approach for detecting the synovial fluid of human knee joint from MR images. First, the MR images are pre-processed and annotated by expert panel to form the training, development, and test sets. In pre-processing step, the images are resized to 450-by-450 pixels. The ground truth of the MR image is a binary image with the mask pinpointing the synovial fluid. Then, the specialised CNN model is fine-tuned with the training set. The hyperparameters are selected according to the best performance of the model, i.e. the lowest average loss, on the development set. Finally, the obtained model is chosen for evaluation on the test set.

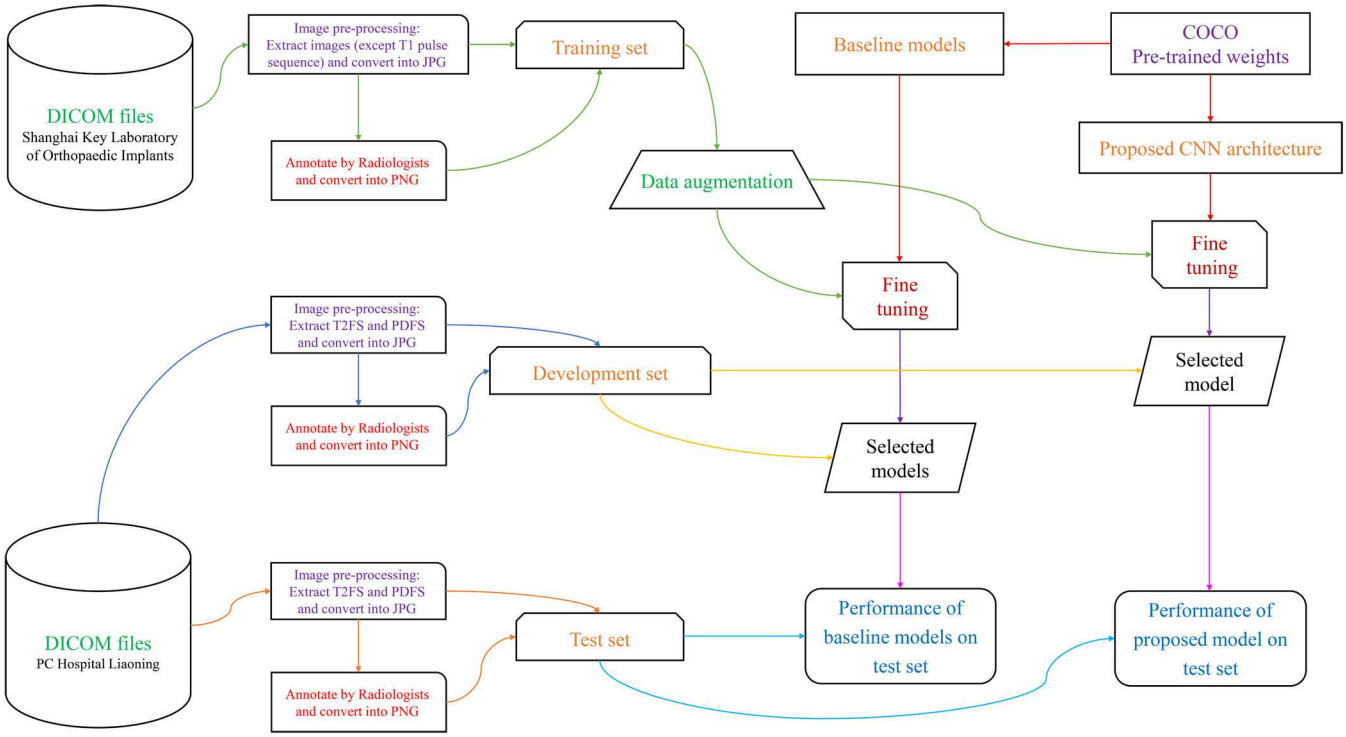

Fig. 2 *Automated deep learning-based human knee joint's synovial fluid detection pipeline. For training set: extract all pulse sequences except T1 from 49 subjects. For development and test sets: extract coronal proton-density weighting with fat suppression (PDFS), sagittal PDFS, and transverse T2 weighting with fat suppression (T2FS) from 15 subjects for development and 25 subjects for evaluation*

4 Experimental results

In this section, several experiments of the proposed CNN architecture and baseline models are performed to detect the synovial fluid of human knee joint from MR images. All of the MR image annotations of the knee joint are performed by two co-authors, radiologists (G. S. and N. R.), who have more than 10 years of radiology experiences. Throughout the experiments NVIDIA GeForce GTX 1080 card with 8 GB GDDR5X memory is used.

4.1 Training setting

DICOM files from Shanghai Key Laboratory of Orthopaedic Implants of 49 subjects (mean age 41.7 years; 28 female subjects) from 2015 to 2017 are acquired for training and 1433 images are extracted in which 890 images (296 transverse, 296 coronal, and 298 sagittal planes) have synovial fluid (T2 and PD pulse sequences) and 543 images (213 transverse, 151 coronal, and 179 sagittal planes) do not have synovial fluid. To make the model more robust to numerous shapes of synovial fluid and to achieve better generalisability, this dataset is augmented with the following operations: horizontal flip, crop image, crop to aspect ratio, pixel value scale, rotation, adjust brightness, adjust contrast, adjust hue, adjust saturation, distort colour and black patches. There are total 17 196 images including 1433 original images and 15 763 augmented images for training. All of the subjects are de-identified in pre-processing step.

As for the development and test sets, DICOM files are acquired from PC Hospital Liaoning. There are 40 subjects (mean age 45.5 years; 17 female subjects) in this dataset. The first 15 subjects are considered for development set and the remaining 25 subjects for test set from this dataset. Splitting one dataset or combination of the datasets randomly into training and testing can be considered a restriction. Test set should be left out so that the generalisation measures, such as sensitivity, and specificity will not be biased and will show the performance of real model. In this study, independent datasets are used for training and testing to prove the robustness of the proposed approach. Thus, there is no overlap (same instance, multiple views) exists between the training and test sets because of the different datasets but the instances might be similar within the training set due to the data augmentation. All of the knee joint MR images are annotated by two co-authors, proficient medical specialists. These annotations are used to create images of binary mask as the same resolution as the MR images. All the pixels within annotated area are labelled as synovial fluid, while rest of them are labelled as none. Three models are fine-tuned, one for each plane (coronal, sagittal, transverse), therefore, there are total 15 fine-tuned models; 12 for the baseline and three for the proposed architecture.

Fig. 3 illustrates some typical MR images of the training, development, and test sets in three planes. Table 1 shows the training time as well as the hyperparameters which are used to fine-tune the models for detecting the synovial fluid from the MR images of the knee joint. The proposed CNN architecture has 30% less training time in comparison with the baseline model-3 due to the simplicity of this network. Fig. 4*a* shows the classification loss. In this subfigure, the classification loss can be seen in three planes of the proposed and baseline models. Fig. 4*b* shows the localisation loss during the training of the proposed and baseline models. Fig. 4*c* shows the total loss of training. The total loss is the sum of region proposal network (RPN) loss and box classifier loss. Fig. 4*d* shows the training time for the knee joint of the proposed and baseline models. According to this figure, it can conclude that the proposed model takes about 7 min per epoch for each plane while the baseline model takes about 9.1 min per epoch for each plane for training of the knee joint. The average training MR image size of the knee joint is ∼450-by-450 by three colour channels.

4.2 Hyperparameter configuration

The hyperparameters of the proposed and baseline models are tuned with the development set. There are 360 knee MR images in development set, in which 120 are transverse, 120 are sagittal, and 120 are coronal planes. This number is similar to the standard of the ImageNet computer vision challenge [4], which has 50–100 images per object category for development set. The average development MR image size of the knee joint is ∼512-by-512 by three colour channels. Several hyperparameters are tuned to improve the performance of the proposed CNN architecture and baseline models. Adam [41] and Momentum [42] optimisers are used with learning rates from 0.00001 to 0.0008, but it has been found that the Momentum algorithm with momentum term 0.9

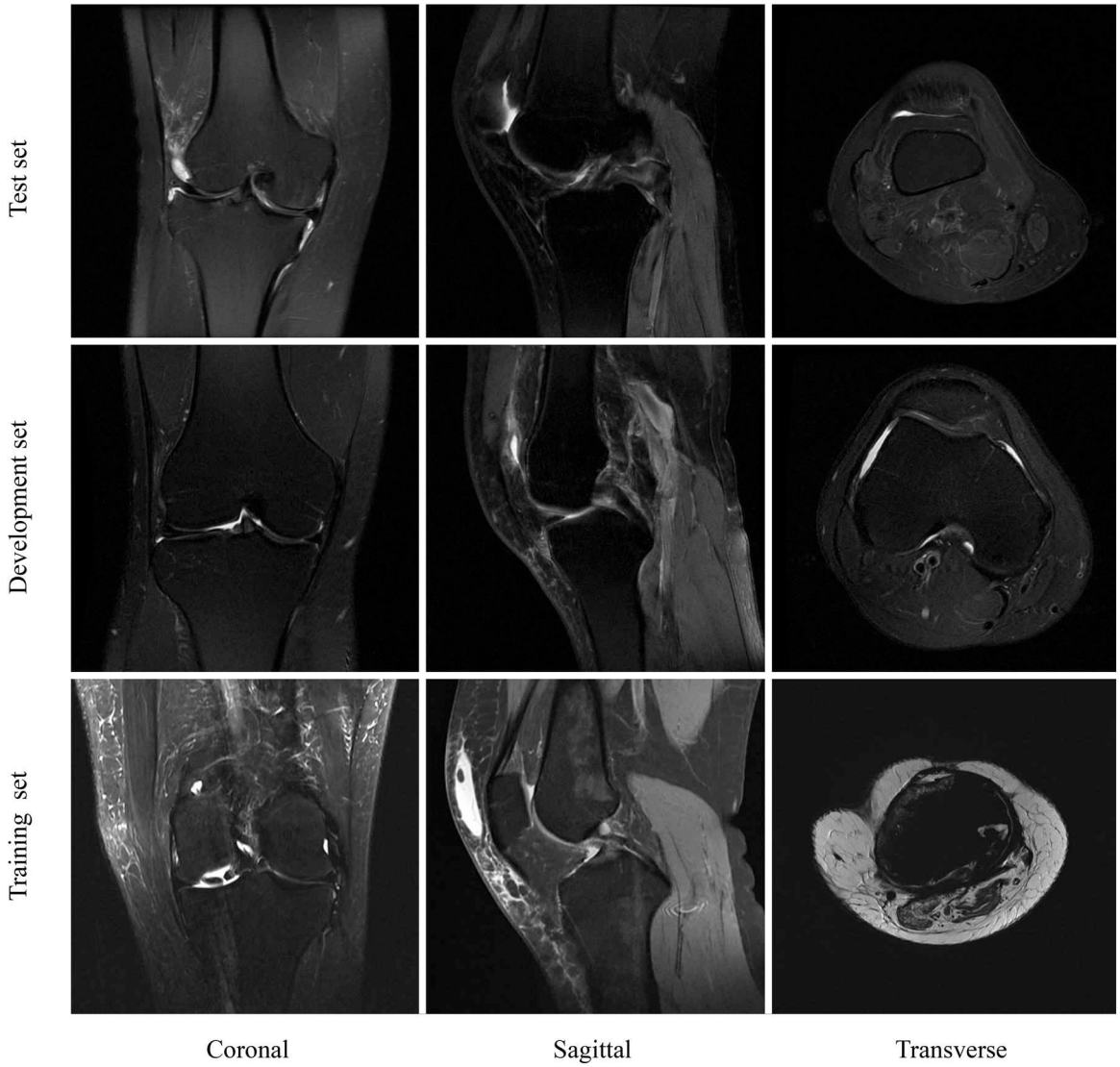

Fig. 3 *Typical training, development, and test MR images of human knee joint of coronal (left), sagittal (middle), and transverse (right) planes*

Table 1 Training time and hyperparameter configuration of the proposed and baseline models

CNN architecture	Learning algorithm	Learning rate	Mini-batch size	Epochs	Dropout rate	Data augmentation	No. of subjects	Total training time (hours)
baseline model-1	momentum	2×10^{-4}	1	~400	0.3	horizontal flip, crop image, crop to aspect ratio, pixel value scale, rotation, adjust brightness, adjust contrast, adjust hue, adjust saturation, distort colour, and black patches	49	**~37.50**
baseline model-2		3×10^{-4}						~81.57
baseline model-3								~182.16
baseline model-4								~127.53
proposed model								~140.15

performs the best. The mini-batch size is set to 1. Various epochs from 100 to 500 are used for fine-tuning, however, it hardly gets any major improvement after 400 epochs. Due to limited number of training MR images, dropout [43] is used during training to avoid overfitting with a number of dropout rate between 0.2 and 0.5, and it makes significant improvement on the performance with 0.3 dropout rate.

Table 1 shows the near optimal hyperparameters, which are selected according to the best performance of the models on the development set. The model with the lowest average loss on the development set is chosen for evaluation on the test set. The average loss is the average of four losses. Table 2 shows the RPN loss and box classifier loss on the development set for the proposed and baseline models. This hyperparameter configuration is optimised for the best performance of the models on the development set.

4.3 Evaluation and comparison

In this section, the proposed and baseline models are evaluated and compared on the test set for detecting the synovial fluid of human knee joint in MR images. The percentage of synovial fluid images for training, development, and testing in coronal, sagittal, transverse planes are given in Fig. 5. In this figure, it can be seen

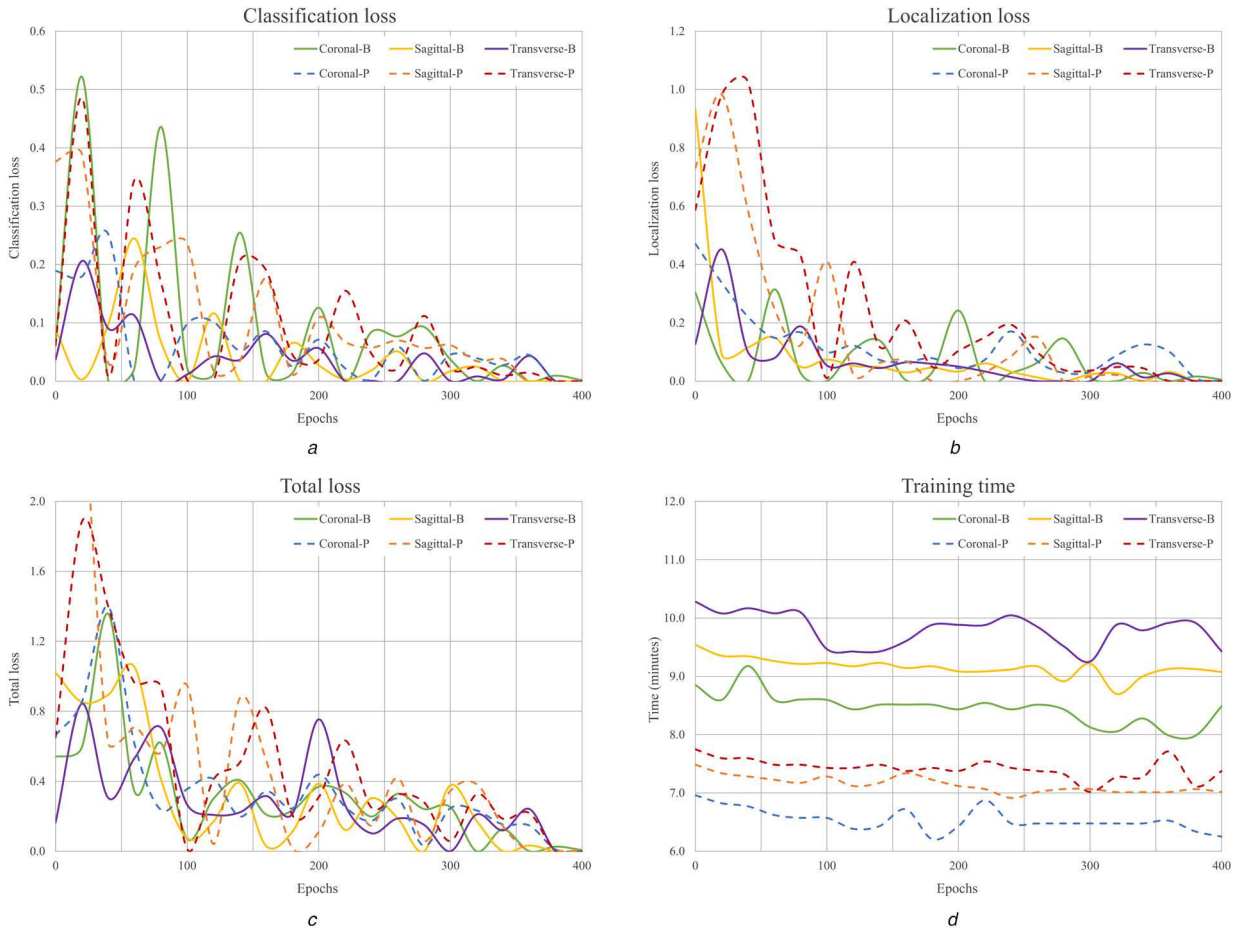

Fig. 4 *Comparison of proposed and baseline models in coronal, sagittal, and transverse planes*

(a) Classification loss, ***(b)*** Localisation loss, ***(c)*** Total loss, ***(d)*** Training time for the baseline model-3 and proposed model for each plane. 'B' indicates for the baseline model-3 and 'P' indicates for the proposed model

Table 2 Development losses of the proposed and baseline models in coronal, sagittal, and transverse planes

CNN architecture	Plane	No. of subjects	RPN loss		Box classifier loss	
			Localisation loss	Objectness loss	Classification loss	Localisation loss
baseline model-1	coronal	15	0.1024	0.1841	0.0801	0.0391
	sagittal		0.1108	0.2254	0.1012	0.0512
	transverse		0.0269	0.1722	0.0465	0.0212
	overall		0.0800	0.1939	0.0759	0.0372
baseline model-2	coronal		0.1547	0.2470	0.0942	0.0462
	sagittal		0.1834	0.3978	0.1057	0.0530
	transverse		0.0493	0.1884	0.0477	0.0215
	overall		0.1291	0.2773	0.0825	0.0402
baseline model-3	coronal		0.1084	0.2309	0.0764	0.0341
	sagittal		0.1446	0.3247	0.1108	0.0636
	transverse		0.0346	0.1576	0.0541	0.0196
	overall		0.0959	0.2377	0.0804	0.0391
baseline model-4	coronal		0.1344	0.2056	0.1068	0.0483
	sagittal		0.1940	0.3794	0.1287	0.0635
	transverse		0.0288	0.1151	0.0831	0.0471
	overall		0.1191	0.2334	0.1062	0.0529
proposed model	coronal		0.1239	0.1870	0.1026	0.0463
	sagittal		0.1801	0.3906	0.0979	0.0642
	transverse		0.0383	0.1212	0.0632	0.0266
	overall		0.1141	0.2329	0.0879	0.0457

that the similar number of synovial fluid images are available in three planes for training, development as well as for testing. Some typical evaluation results of detecting synovial fluid are shown in Fig. 6 in three planes. The detection time of synovial fluid could be much faster if multiple graphics processing unit or tensor processing unit are used.

For evaluation, there are 582 knee joint MR images in the test set, in which 200 are transverse, 194 are sagittal, and 188 are coronal planes. The average test MR image size of the knee joint is

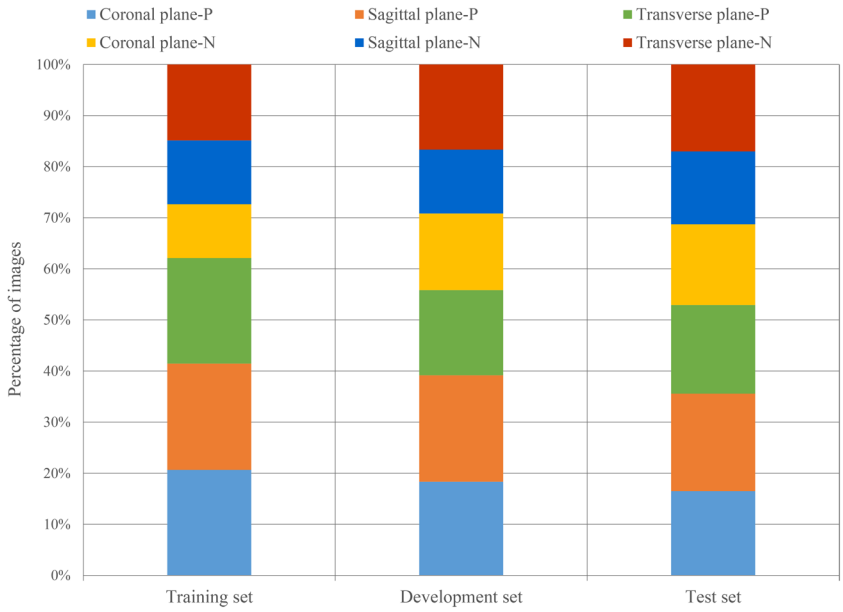

Fig. 5 *Percentages of synovial fluid images in coronal, sagittal, and transverse planes for training, development, and test sets. 'P' denotes for positive example and 'N' denotes for negative example of synovial fluid*

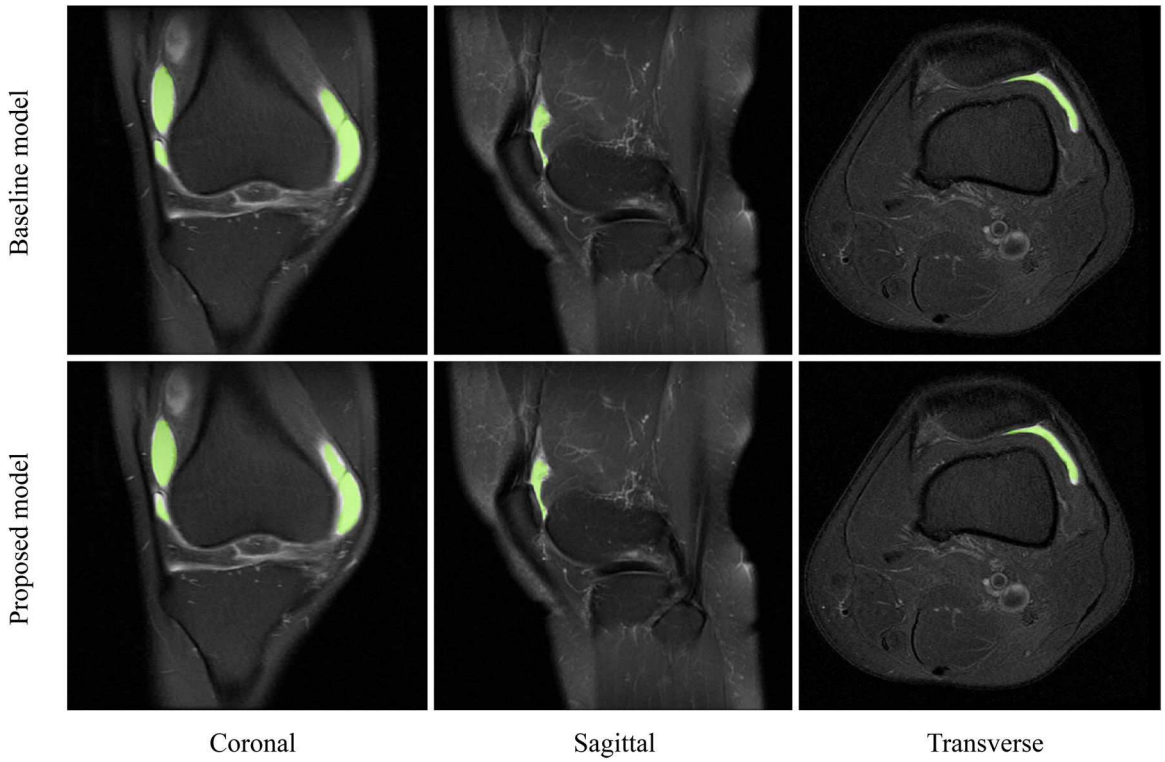

Fig. 6 *Successful detection of human knee joint's synovial fluid instances in coronal (left), sagittal (middle), and transverse (right) planes of the baseline model-3 and proposed model*

roughly 512-by-512 by three colour channels. If the centre of predicted mask is within the ground truth or vice versa, it considered the detection result as a true-positive (TP). Fig. 6 shows the typical detection results of synovial fluids of the knee joint by the proposed and baseline models. Table 3 shows a detailed comparison of the proposed and baseline models. All of the optimal metrics of the baseline models are less than those of the proposed model in coronal sagittal and transverse planes, except for the specificity of the baseline model-3 on sagittal plane. The overall values in Tables 2 and 3 are the corresponding weighted average values of coronal, sagittal, and transverse planes. The proposed model is 1.1 times faster than the baseline model-3 for detecting synovial fluid of the knee joint. The bold values in this table show the best results. The false-positive (FP) and false-negative (FN) indices may be reduced by increasing the number of training examples. The accuracy of the proposed model for detecting the synovial fluid is approximately similar in three planes. Despite the limited number of training examples, it achieves high precision, sensitivity, and specificity. These results support the statement that deep learning models can be helpful to medical practitioners during medical imaging interpretation. In comparison to the medical expert, who requires more than 5 h to trace/contour the synovial fluid from 25 subjects, with the help of deep learning it can be easily done within 7 min.

4.4 Discussion

The objective of this work is to make and investigate a deep learning model for synovial fluid detection on knee MR images. Due to incorrect or late diagnosis, a few cases of wrong treatment

Table 3 Evaluation results of the proposed and baseline models in coronal, sagittal, and transverse planes

CNN architecture	Plane	No. of subjects	Optimising metrics							Satisfying metric
			Sensitivity (recall)	Specificity	Precision	*F*-score (Dice coefficient)			Accuracy	Detection time (seconds per image)
						$F_{0.5}$	F_1	F_2		
baseline model-1	coronal	25	0.4835	0.7583	0.7803	0.6949	0.5970	0.5233	0.5825	~**0.27**
	sagittal		0.5138	0.7476	0.7750	0.7034	0.6179	0.5509	0.6006	
	transverse		0.6404	0.7674	0.7916	0.7559	0.7080	0.6658	0.6938	
	overall		0.5475	0.7578	0.7824	0.7187	0.6421	0.5815	0.6267	
baseline model-2	coronal		0.8075	0.7076	0.8190	0.8167	0.8132	0.8097	0.7696	~0.67
	sagittal		0.8232	0.7053	0.8186	0.8195	0.8208	0.8223	0.7781	
	transverse		0.7640	0.7857	0.8343	0.8192	0.7976	0.7770	0.7730	
	overall		0.7977	0.7336	0.8241	0.8185	0.8103	0.8027	0.7738	
baseline model-3	coronal		0.8779	0.7177	0.8423	0.8492	0.8597	0.8705	0.8189	~0.78
	sagittal		0.6740	**0.8350**	0.8840	0.8321	0.7648	0.7076	0.7302	
	transverse		0.8426	0.7741	0.8426	0.8426	0.8426	0.8426	0.8145	
	overall		0.7978	0.7761	0.8563	0.8412	0.8222	0.8066	0.7878	
baseline model-4	coronal		0.8638	0.7250	0.8479	0.8510	0.8558	0.8606	0.8138	~0.70
	sagittal		0.7956	0.7255	0.8372	0.8285	0.8159	0.8036	0.7703	
	transverse		0.8089	0.8305	0.8780	0.8632	0.8420	0.8218	0.8176	
	overall		0.8222	0.7614	0.8547	0.8477	0.8378	0.8283	0.8006	
proposed model	coronal		**0.8967**	**0.8073**	**0.9009**	**0.9001**	**0.8988**	**0.8975**	**0.8665**	~0.71
	sagittal		**0.9111**	0.8021	**0.8962**	**0.8991**	**0.9036**	**0.9081**	**0.8732**	
	transverse		**0.8708**	**0.8522**	**0.9012**	**0.8950**	**0.8857**	**0.8767**	**0.8635**	
	overall		**0.8925**	**0.8209**	**0.8994**	**0.8980**	**0.8958**	**0.8939**	**0.8677**	

have been filed. Since effect of treatment takes time to appear, sometimes misdiagnosis leads to an increased need for surgical treatment and hospitalisation duration [44]. Developing an automatic synovial fluid detection system could be good to lessen the workload of radiologists or even orthopaedic surgeons and also decreasing the subjectivity, inaccuracies due to the tiredness of medical specialists. This proposed work can also be useful when experienced radiologists are not readily available and for low-experienced clinicians in the third world countries. This detection system can be integrated with the MR imaging machines. This proposed method is easily extensible to several image modalities such as radiography to detect the joint dislocation or fractured bones, elastography to diagnose the presence or status of disease in soft tissues.

It can be found from Table 3 that those baseline models are not so powerful to extract effective features from the MR images and to detect the synovial fluid of human knee joint in comparison with the proposed model according to the specificity and complexity. It even finds out that the proposed model improves the accuracy by a factor of 10%, and decrease as the evaluation time by a factor of 9% compared with the baseline model-3 in the experiments.

These results provide strong support to that the deep learning approach is able to play a key role in assisting doctors and healthcare systems, but more research is required to evaluate the deep models in clinical setting. It is worth indicating the limitations of this deep learning-based detection system. As mentioned before, experiments are conducted with 89 (training: 49, development: 15, test: 25) subjects. This number of subjects is relatively small. Therefore, to obtain better generalisability, it is essential to increase the number of subjects in the future. Secondly, due to the limited computational power and memory, the training time is very high. The small size of the panel of medical professionals, two co-authors, is also main limitation of this research.

5 Conclusion

A deep learning-based method for automatic detection of synovial fluid from MR images of human knee joint has been proposed in this research. The specialised deep CNN architecture is designed for the complicated synovial fluid detection task and the transfer learning is implemented from the pre-trained weights of COCO dataset. The experimental results demonstrate that the proposed deep learning approach with transfer learning mechanism can detect the synovial fluid of human knee joint with high accuracy. While these results are encouraging, further validation and improvement are still necessary before it can be entirely executed in medical practice. MR imaging datasets of human knee joint for synovial fluid detection are introduced to facilitate future research.

6 Acknowledgments

The authors are thankful to 'Shanghai Jiaotong University' and 'PC Hospital Liaoning' for providing them Knee MRI datasets. This work was supported by the National Key Research and Development Program of China under grant 2018AAA0100205.

7 References

[1] Everingham, M., Van Gool, L., Williams, C.K.I., *et al.*: 'The PASCAL visual object classes (VOC) challenge', *Int. J. Comput. Vis.*, 2010, **88**, (2), pp. 303–338

[2] Hinton, G.E., Salakhutdinov, R.R.: 'Reducing the dimensionality of data with neural networks', *Science*, 2006, **313**, (5786), pp. 504–507

[3] Krizhevsky, A., Sutskever, I., Geoffrey, H.E.: 'Imagenet classification with deep convolutional neural networks'. Advances in Neural Information Processing Systems 25 (NIPS2012), Lake Tahoe, NV, USA, 2012, pp. 1097–1105

[4] Russakovsky, O., Deng, J., Su, H., *et al.*: 'Imagenet large scale visual recognition challenge', *Int. J. Comput. Vis.*, 2015, **115**, (3), pp. 211–252

[5] Szegedy, C., Liu, W., Jia, Y., *et al.*: 'Going deeper with convolutions'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015

[6] Bramlett, K.: 'Knee joint', 2018. Available at: http://www.bramlettorthopedics.com/conditions/knee. [Accessed: 04-Dec-2018]

[7] West, S.: '*Rheumatology secrets*' (Elsevier Mosby, Philadelphia, 2015, 3rd edn.)

[8] Lam, P., Marcin, J., Felman, A.: 'What to know about MRI scans', 2018. Available at: https://www.medicalnewstoday.com/articles/146309.php. [Accessed: 10-Dec-2018]

[9] Reicher, M., Rauschning, W., Gold, R., *et al.*: 'High-resolution magnetic resonance imaging of the knee joint: normal anatomy', *Am. J. Roentgenol.*, 1985, **145**, pp. 895–902

[10] Stoller, D.W., Genant, H.K.: 'Magnetic resonance imaging of the knee and hip', *Arthritis Rheum.*, 1990, **33**, (3), pp. 441–449

[11] Shanghai Hospital.: 'Department of orthopedics, shanghai key laboratory of orthopedic implants, shanghai ninth people's hospital, shanghai jiaotong university, school of medicine, shanghai, China'

[12] Hospital, P.C.: 'Department of orthopaedic surgery, PC hospital liaoning, China'

[13] Breheret, A.: 'Pixel annotation tool', 2017. Available at: https://github.com/abreheret/PixelAnnotationTool

[14] Lin, T.-Y., Maire, M., Belongie, S., *et al.*: 'Microsoft COCO: common objects in context'. European Conf. on Computer Vision, Zürich, Switzerland, 2014, pp. 740–755

[15] Esteva, A., Robicquet, A., Ramsundar, B., *et al.*: 'A guide to deep learning in healthcare', *Nat. Med.*, 2019, **25**, pp. 24–29doi:

[16] Acharya, U.R., Oh, S.L., Hagiwara, Y., *et al.*: 'A deep convolutional neural network model to classify heartbeats', *Comput. Biol. Med.*, 2017, **89**, pp. 389–396

[17] Javadi, S., Mirroshandel, S.A.: 'A novel deep learning method for automatic assessment of human sperm images', *Comput. Biol. Med.*, 2019, **109**, pp. 182–194

[18] Simonyan, K., Zisserman, A.: 'Very deep convolutional networks for large-scale image recognition'. ICLR 2015, San Diego, CA, USA, 2015, pp. 1–14

[19] Hernandez, V., Rezzoug, N., Gorce, P., *et al.*: 'Engineering force feasible set prediction with artificial neural network and musculoskeletal model', *Comput. Methods Biomech. Biomed. Engin.*, 2018, **21**, (14), pp. 740–749

[20] Esteva, A., Kuprel, B., Novoa, R.A., *et al.*: 'Dermatologist-level classification of skin cancer with deep neural networks', *Nature*, 2017, **542**, (7639), pp. 115–118

[21] Szegedy, C., Vanhoucke, V., Ioffe, S., *et al.*: 'Rethinking the inception architecture for computer vision'. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2818–2826

[22] Gulshan, V., Peng, L., Coram, M., *et al.*: 'Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs', *J. Am. Med. Assoc.*, 2016, **316**, (22), pp. 2402–2410

[23] Poplin, R., Varadarajan, A.V., Blumer, K., *et al.*: 'Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning', *Nat. Biomed. Eng.*, 2018, **2**, (3), pp. 158–164

[24] Chang, V., Garcia, A., Hitschfeld, N., *et al.*: 'Gold-standard for computer-assisted morphological sperm analysis', *Comput. Biol. Med.*, 2017, **83**, pp. 143–150

[25] Fix, E., Hodges, L.J.: 'Discriminatory analysis. Nonparametric discrimination: consistency properties'. Tech. Rep. 4, USAF School of Aviation Medicine, Randolph Field, Texas, USA, 1951

[26] Kononenko, I.: 'Comparison of inductive and naive Bayesian learning approaches to automatic knowledge acquisition, in: current trends in knowledge adquisition', *Front. Artif. Intell. Appl.*, 1990, **8**, pp. 190–197

[27] Quinlan, J.R.: 'Induction of decision trees', *Mach. Learn.*, 1986, **1**, (1), pp. 81–106

[28] Boser, V.V.B., Guyon, I.: 'A training algorithm for optimal margin classifiers'. Fifth Annual Workshop on Computational Learning Theory, Pittsburgh, PA, USA, 1992, pp. 144–152

[29] Riordon, J., Mccallum, C., Sinton, D.: 'Deep learning for the classification of human sperm', *Comput. Biol. Med.*, 2019, **111**, p. 103342doi:

[30] He, K., Zhang, X., Ren, S., *et al.*: 'Deep residual learning for image recognition'. Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016

[31] Chollet, F.: 'Xception: deep learning with depthwise separable convolutions'. IEEE Conf. on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017, pp. 1800–1807

[32] Szegedy, C., Ioffe, S., Vanhoucke, V., *et al.*: 'Inception-v4, inception-ResNet and the impact of residual connections on learning'. AAAI, San Francisco, CA, USA, 2017, pp. 4278–4284

[33] Nair, V., Hinton, G.E.: 'Rectified linear units improve restricted Boltzmann machines'. Proc. of the 27th Int. Conf. on Machine Learning, Haifa, Israel, 2010, no. 3, pp. 807–814

[34] Ioffe, S., Szegedy, C.: 'Batch normalization: accelerating deep network training by reducing internal covariate shift', 2015

[35] Thrun, S., Pratt, L.: '*Learning to learn*' (Kluwer Academic Publishers, Boston, Mass, 1998)

[36] Shin, H.-C., Roth, H.R., Gao, M., *et al.*: 'Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning', *IEEE Trans. Med. Imaging*, 2016, **35**, (5), pp. 1285–1298

[37] Huang, J., Rathod, V., Sun, C., *et al.*: 'Speed/accuracy trade-offs for modern convolutional object detectors'. in Proc. – 30th IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 2017, pp. 3296–3305

[38] Abadi, M., Agarwal, A., Barham, P., *et al.*: 'Tensorflow: large-scale machine learning on heterogeneous distributed systems', 2016

[39] He, K., Gkioxari, G., Dollar, P., *et al.*: 'Mask R-CNN'. Proc. of the IEEE Int. Conf. on Computer Vision, Venice, Italy, 2017

[40] Yu, F., Koltun, V.: 'Multi-scale context aggregation by dilated convolutions'. ICLR, San Juan, Puerto Rico, 2016

[41] Kingma, D.P., Ba, J.L.: 'Adam: A method for stochastic optimization'. Proc. of the 3rd Int. Conf. on Learning Representations, San Diego, CA, USA, 2015

[42] Qian, N.: 'On the momentum term in gradient descent learning algorithms', *Neural Netw.*, 1999, **12**, (1), pp. 145–151

[43] Srivastava, N., Hinton, G., Krizhevsky, A., *et al.*: 'Dropout: A simple way to prevent neural networks from overfitting', *J. Mach. Learn. Res.*, 2014, **15**, (1), pp. 1929–1958

[44] McHenry, M.C., Easley, K.A., Locker, G.A.: 'Vertebral osteomyelitis: long-term outcome for 253 patients from 7 Cleveland-area hospitals', *Clin. Infect. Dis.*, 2002, **34**, (10), pp. 1342–1350