Turkish Journal of Electrical Engineering and Computer Sciences

Volume 28 | Number 5 Article 39

1-1-2020

Detection of hand osteoarthritis from hand radiographs using convolutionalneural networks with transfer learning

KEMAL ÜRETEN

HASAN ERBAY

HADİ HAKAN MARAŞ

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2020) 28: 2968 – 2978
© TÜBİTAK
doi:10.3906/elk-1912-23

Detection of hand osteoarthritis from hand radiographs using convolutional neural networks with transfer learning

Kemal ÜRETEN[1,2,*], **Hasan ERBAY**[3], **Hadi Hakan MARAŞ**[2]
[1]Department of Rheumatology, Faculty of Medicine, Kırıkkale University, Kırıkkale, Turkey
[2]Department of Computer Engineering, Faculty of Engineering, Çankaya University, Ankara, Turkey
[3]Department of Computer Engineering, Faculty of Engineering, University of Turkish Aeronautical Association, Ankara, Turkey

Received: 03.12.2019 • **Accepted/Published Online:** 05.06.2020 • **Final Version:** 25.09.2020

Abstract: Osteoarthritis is the most common type of arthritis. Hand osteoarthritis leads to specific structural changes in the joints, such as asymmetric joint space narrowing and osteophytes (bone spurs). Conventional radiography has traditionally been the primary method of visualizing these structural changes and diagnosing osteoarthritis. We aimed to develop a computerized method that is capable of determining the structural changes seen in radiography of the hand and to assist practitioners in interpreting radiographic changes and diagnosing the disease.

In this retrospective study, transfer-learning-based convolutional neural networks were trained on a randomly selected dataset containing 332 radiography images of hands from an original set of 420 and were validated with the remaining 88. Multilayer convolutional neural network models were designed based on a transfer learning method using pretrained AlexNet, GoogLeNet, and VGG-19 networks. The accuracies of the models were 93.2% for AlexNet, 94.3% for GoogLeNet, and 96.6% for VGG-19. The sensitivities of these models were 0.9167 for AlexNet, 0.9184 for GoogLeNet, and 0.9574 for VGG-19, while the specificity values were 0.9500, 0.9744, and 0.9756, respectively. The performance metrics, including accuracy, sensitivity, specificity, and precision, of our newly developed automated diagnosis methods are promising in the diagnosis of hand osteoarthritis. Our computer-aided detection systems may help physicians in interpreting hand radiography images, diagnosing osteoarthritis, and saving time.

Key words: Hand osteoarthritis, convolutional neural networks, transfer learning, conventional hand radiography, classification

1. Introduction

Osteoarthritis (OA) is the most common type of arthritis, and its incidence has been rising due to increases in life expectancy [1]. OA of the hand leads to specific structural distortions in the joints, and most commonly affects the first carpometacarpal joints (1.CMC), distal interphalangeal joints (DIP), and proximal interphalangeal joints (PIP). The radiographic hallmarks of OA include nonuniform joint space loss, osteophyte formation, and subchondral sclerosis [2–5]. Notable risk factors for OA include age, obesity, sex, smoking, genetics, diet, and occupation [3]. Hand OA is common in manual workers such as textile employees [6], and previous trauma may speed up the development of the disease. The diagnosis of hand OA is based primarily on the patient's history, physical examination, and radiographic findings. A radiological examination may reveal instability and ankylosis of the joints, Heberden nodes (enlargement of DIP), Bouchard nodes (enlargement of PIP), and joint

*Correspondence: kemalureten@yahoo.com

deformities. Levels of acute-phase reactants such as C-reactive protein and erythrocyte sedimentation rate are typically within the reference range in patients with OA [3, 4].

Conventional radiographs (CRs) of the hands are the gold standard diagnostic method for OA. CRs are also successful in demonstrating the severity of the disease, monitoring its development, and differentiating between erosive and nonerosive patterns of OA [7, 8]. CRs are widely used since they are low-cost, quick, and easily accessible imaging methods. Osteophytes, joint space narrowing, subchondral sclerosis, pseudocystic areas, subchondral erosions, subluxations, and loss of joint space can be seen with a CR. Various scoring methods have been proposed to assess the severity of OA, and the Kellgren-Lawrence scoring method is the most universally accepted of these [9].

Computer-aided detection (CAD) is a term for systems that assist doctors in the interpretation of medical images. Image processing and computer vision in healthcare have been expanding following recent technological advancements, and the increased resolution of images has created a rich data pool, making it easier to identify pathological markers. The successful classification of heterogenous images relies on the availability of large datasets. Computerized methods, and especially image analysis and machine learning, can facilitate the identification of pathological findings, help in making a correct diagnosis, and support the physician's workflow [10]. A convolutional neural network (CNN) is a deep learning method based on a multilayer structure of artificial neural networks. A typical CNN architecture for image processing consists of stacks of multiple convolution filter layers and a series of data reduction or accumulation layers. Convolution layers are used as small-scale detectors to explore the features of an image, and a feature in a given location within the image can be computed by convolution. The number of layers, neurons, iterations, and samples to be used for training and validating should be determined before train the network. CNN applications offer a promising approach and provide state-of-the-art results for image processing. There is a growing interest in CNNs, and encouraging results are being reported [10]. Successful results have been obtained in studies using CNN as a method for diagnosing breast cancer [11], brain tumors [12], lung lesions [13], and rheumatoid arthritis [14].

CNNs are particularly successful when applied to feature extraction and classification. A great deal of data is required for the implementation of a CNN method, and the quality of the images, the availability of sufficient data, and an appropriate design of the CNN are important factors [10, 15, 16] in creating a successful model. When not enough data are available, data augmentation and transfer learning can be applied. ImageNet is a project containing 1.2 million images in 1000 separate categories, and networks such as AlexNet [17], SqueezeNet [18], Resnet [19], VGG [20], GoogLeNet [21], and Inceptionv3 [22] have been developed based on this dataset. These are relatively popular networks with different features and levels of performance: AlexNet and VGG have traditional sequential network architectures, while ResNet is an "exotic architecture" based on microarchitecture modules (also called a "network in a network"). GoogLeNet contains an inception module [23].

Hand OA is a very common health problem, and patients with hand pain may seek medical advice from primary care doctors or a broad range of different specialists. The accurate diagnosis of OA leads to a reduction in the number of unnecessary examinations and healthcare costs, and can also contribute to the diagnosis of inflammatory arthritis via the exclusion of OA. In this context, we aim to develop a CNN-based model using CR to diagnose hand OA.

2. Materials and methods

An ethical approval certificate was obtained from the Noninterventional Clinical Research Ethics Board at Kırıkkale University. Certificate date: Sep 12th 2018; Certicate no: 2018.09.01.

2.1. Image acquisition

Images were obtained from patients examined at the rheumatology outpatient clinic of Kırıkkale University, between Jan 1st, 2012, and Oct 1st, 2018. The dataset consisted of 210 X-ray images of patients' both hands, with various dimensions. The radiographs were evaluated by an experienced rheumatologist, who classified each radiograph based on whether or not it was compatible with osteoarthritis. Of these 210 radiographs, 103 were compatible with hand OA. The images were cut in two to give left- and right-hand images and the left-hand radiographs were flipped, giving a dataset containing 214 normal and 206 OA radiographs.

2.2. Data preprocessing and splitting

The radiographs were in RGB image format. All of the images were resized to 224×224 pixels. Thus, $224 \times 224 \times 3$ (3 means RGB image format) dimension images were obtained, and the dataset was randomly split into a training set and a validation set, as shown in Table 1.

Table 1. Training and validation statistics after splitting.

Images	Training	Validation
Normal	168	46
Osteoarthritis	164	42
Total	332	88

2.3. Data augmentation

Image data augmentation helps to prevent overfitting and improves the overall performance of CNN. We, therefore, applied online data augmentation by randomly translating images by up to three pixels horizontally and vertically, and rotating images through an angle of up to 20°.

2.4. Transfer learning

AlexNet, GoogLeNet, and VGG-19 networks used in this study were developed using the Imagenet dataset. AlexNet and VGG-19 have a sequential network structure. AlexNet has 8 layers, VGG-19 has 19 layers. These layers are convolution layers, max-pooling layers, and fully connected layers. In these networks, the rectified linear unit (ReLU) was used as the activation function, which shows improved training performance over tanh and sigmoid. GoogLeNet has an inception network architecture and has 22 layers. The Inception network model consists of modules. Each module consists of different dimensional convolution and max-pooling processes. Nowadays these pretrained networks are used successfully for transfer learning. Figure 1 shows the overview of the transfer model architecture. The characteristics of these pretrained networks are shown in Table 2. In this study, a transfer learning method was applied using pretrained AlexNet, GoogLeNet, and VGG-19 networks, and a binary classifier was added instead of the original classifier. During the training, fine-tuned same hyperparameters were used for each model, as shown in Table 3.

2.5. Methods

This study was implemented in MATLAB® programming environment. The networks were trained on a dataset containing 332 images of hands, of which 168 were normal and 164 OA, and validated them on 88 radiographs, of which 46 were normal and 42 OA. Hand joints are different from the knee and hip joints; they are small and

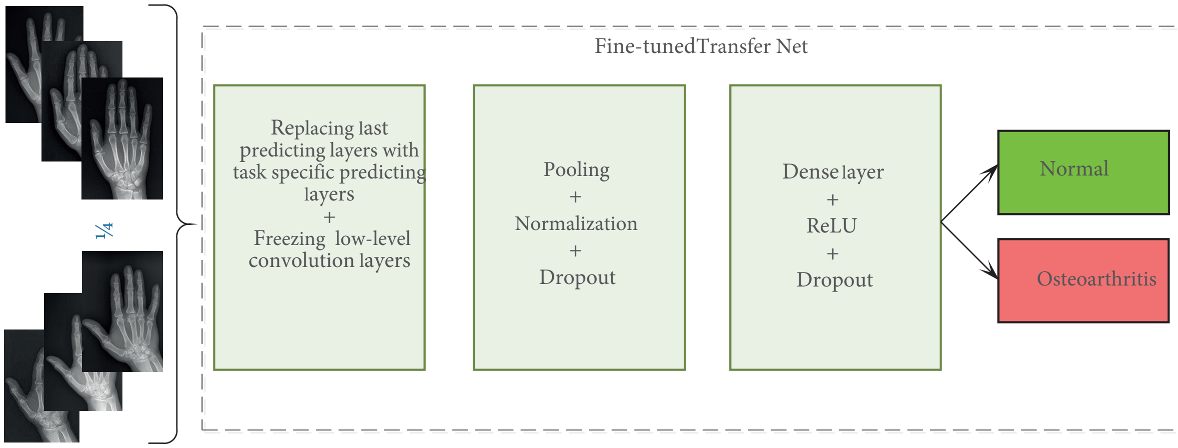

Figure 1. The overview of the transfer model.

Table 2. Properties of CNN models used in this study.

	AlexNet	GoogLeNet	VGG-19
Model	Sequential	Inception modul	Sequential
Depth	8	22	19
Number of layers	25	144	47
Number of filters	96	64	64
Filter size	11x11	7x7	3x3
Stride	4x4	1x1	1x1
Dropout		50%	
Activation function		ReLU	
Output size		2	

Table 3. Hyperparameters used in this study.

	AlexNet	GoogLeNet	VGG-19
Optimizer		sgdm	
Momentum		0.900	
Initial learn rate		$3.0000e-04$	
L2 Regularization		$1.0000e-04$	
Gradient threshold method		L2 Norm	
Gradient threshold		Inf	
Validation frequency		3	
Validation patience		Inf	
Maximum epochs		30	
Frozen weights (layers)	1:6	1:10	1:6
Total learnables	56,876,418	5,975,602	139,578,434
Mini batch size		16	
Shuffle		Every epoch	

numerous. Depending on the etiology, different joints can be affected at different degrees. Therefore, during preprocessing, in this study, each hand image was cut separately on the radiographs and the left-hand image was turned to the right. Thus, the number of images for training was also increased. Figure 2 shows some sample images from our dataset.

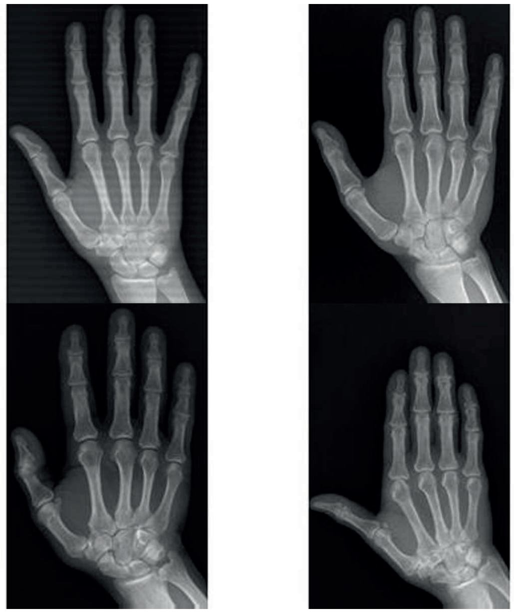

Figure 2. Image samples from the dataset (left: patient without osteoarthritis; right: patient with osteoarthritis).

3. Results

Figures 3–5 show the accuracy of training and loss plots for the transfer learning models using AlexNet, GoogLeNet, and VGG-19, respectively. We observed that all of the models behaved similarly. The accuracies of training (blue line) and validation (dashed black line) converged to each other during the training process for each model. We observed that both the training and the validation losses gradually decreased further in the learning process for each model.

Figure 6 shows the confusion matrices for the AlexNet, GoogLeNet, and VGG-19 models. We used these confusion matrices to evaluate the efficacy based on performance metrics such as accuracy, sensitivity, specificity, and precision. The accuracies of the models were 93.2% for AlexNet, 94.3% for GoogLeNet, and 96.6% for VGG-19, while the sensitivities of the models were 0.9167 for AlexNet, 0.9184 for GoogLeNet, and 0.9574 for VGG-19, and the specificity values were 0.9500, 0.9744, and 0.9756, respectively (Table 4).

These three models were tested with the images which had been used neither training nor validating. The test set included 22 normal both hand radiographs and 26 hand OA. With a total of 96 images consisting of 44 normal and 52 hand OA images, 0.9375, 0.8958 and 0.8958 accuracies were obtained for AlexNet, GoogLeNet, and VGG-19, respectively.

4. Discussion

CR assessment is important in the final diagnosis of hand OA, as acute phase reactants are typically within the reference range and there are no clear biomarkers. Experience is very important in detecting the early radiographic changes in hand OA. Our CAD methods can assist physicians irrespective of their experience and

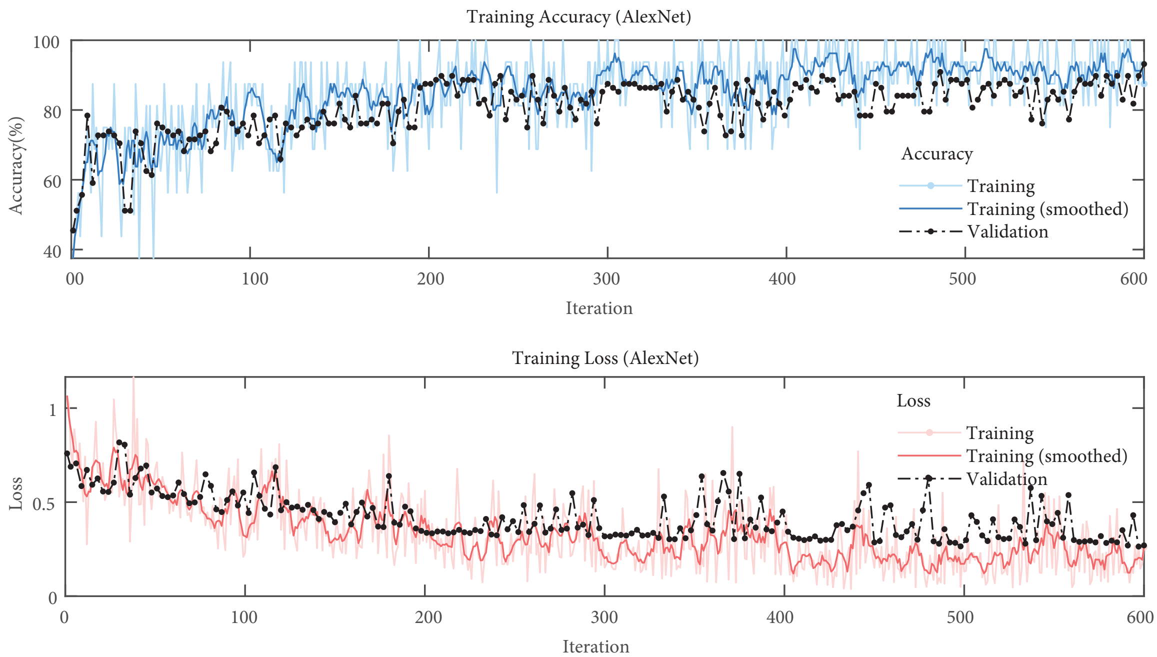

Figure 3. Transfer learning using a pretrained AlexNet: training accuracy (upper image), training loss (lower image).

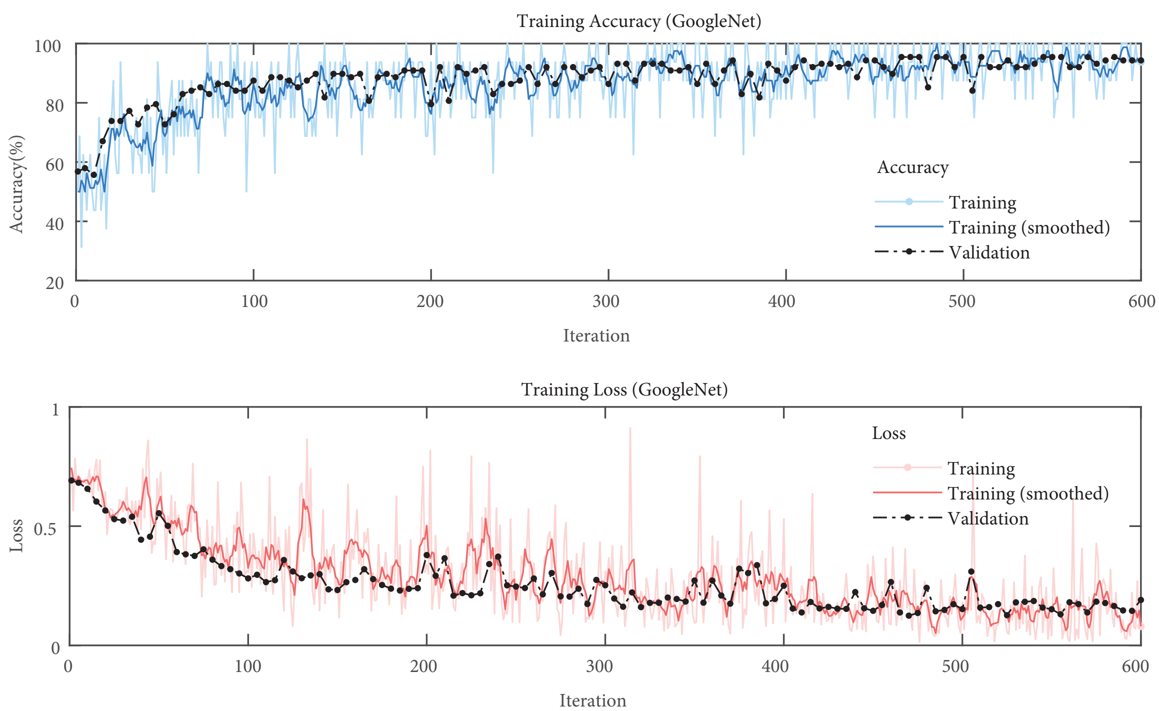

Figure 4. Transfer learning using a pretrained GoogLeNet: training accuracy (upper image), training loss (lower image)

could be used by a range of types of physicians, from those who are not familiar with hand radiographs to rheumatologists and radiologists, in their daily practice or research activities.

Recent developments in computer technology have facilitated the acquisition of high-resolution images and have increased the ability of systems to process these. These developments have revolutionized medical

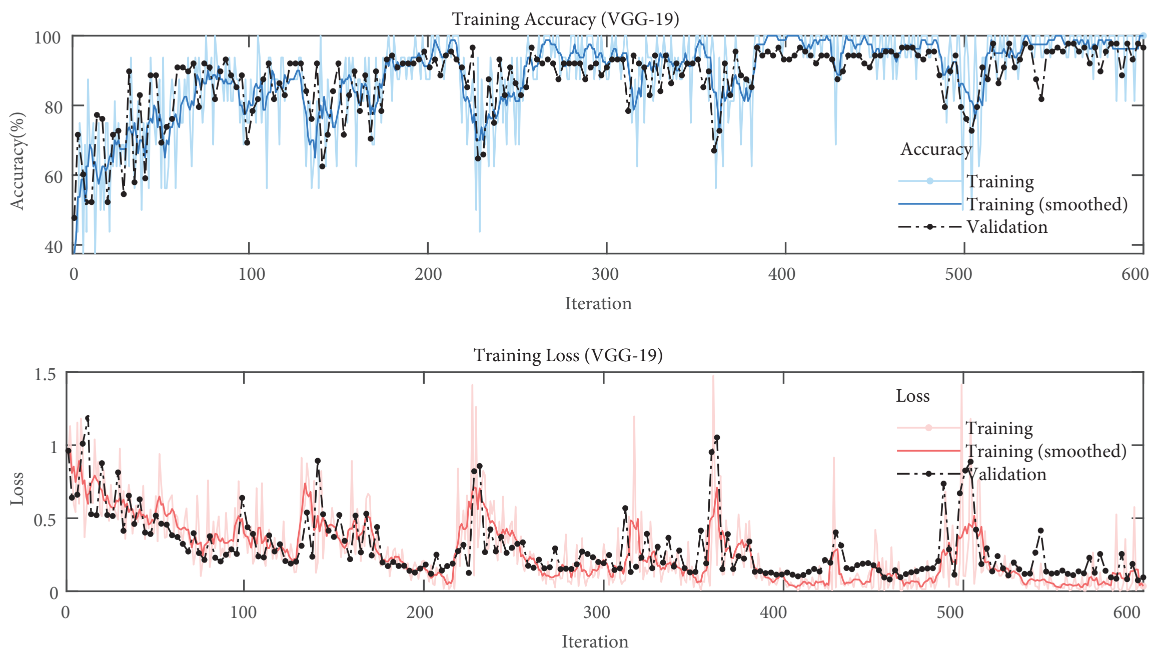

Figure 5. Transfer learning using a pretrained VGG-19: training accuracy (upper image), training loss (lower image)

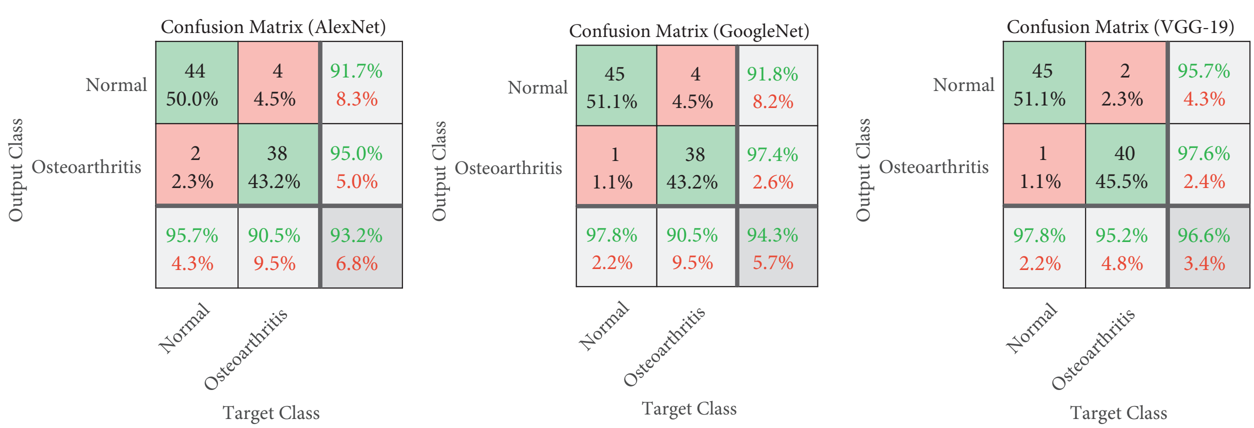

Figure 6. Confusion matrix for transfer learning using AlexNet (left), GoogLeNet (middle), and VGG-19 (right).

Table 4. Performance statistics.

Measure	Value		
	AlexNet	GoogLeNet	VGG-19
Sensitivity	0.9167	0.9184	0.9574
Specificity	0.9500	0.9744	0.9756
Precision	0.9565	0.9783	0.9783
Accuracy	0.9318	0.9432	0.9659

imaging and artificial intelligence, resulting in numerous successful studies that have reported the efficacy of CNN applications in the diagnosis of several conditions [12, 13, 24].

Training a CNN network from scratch requires a great deal of data (images) and time. When insufficient data are available, data augmentation and transfer learning can be applied. Transfer learning is a way of applying a previously trained model to a new problem and allows us to train deep neural networks with relatively small amounts of data. This represents a new era in deep learning since it would be hard to find large, labeled datasets to train complex models. In particular, finding sufficient datasets for medical research purposes is extremely difficult, due to the cost of annotating images and the scarcity of disease lesions. With transfer learning, we can use previously trained CNN models for new medical tasks [24, 25].

Few studies using pretrained networks developed using Imagenet dataset images have reported successful applications in the medical field. Successful results were obtained with a CNN method using transfer learning for the diagnosis of hip and knee OA, and hand fractures [24, 26–28]. Studies on osteoarthritis on plain radiographs using CNN and transfer learning are shown in Table 5. To the best of our knowledge, the present work is the first to detect hand OA using transfer learning with a CNN. In this study, transfer learning was performed using three different pretrained networks to detect hand OA. The architectures and features of AlexNet, GoogLeNet, and VGG-19 are different, and we found that the performance of VGG-19 was slightly better than the others; however, all three networks successfully diagnosed OA of the hand with high levels of accuracy, sensitivity, and specificity.

Table 5. Comparison of the proposed study with the existing studies on osteoarthritis in the literature.

	Dataset	Studied on	Number of arthritic images	Used CNN models
Tiulpin et al. [27]	Osteoarthritis Initiative (OAI) and Multicenter Osteoarthritis Study (MOST)	Knee osteoarthritis	27.293	Deep siamese CNN model
Gorriz et al. [28]	Osteoarthritis Initiative (OAI) and Multicenter Osteoarthritis Study (MOST)	Knee osteoarthritis	4.446	End-to-end CNN model
Xue et al. [26]	Beijing Chaoyang Hospital, China	Hip osteoarthritis	201	Transfer learning using VGG-16 network
Proposed study	Kırıkkale University Hospital, Turkey	Hand osteoarthritis	206	Transfer learning using AlexNet, GoogLeNet, VGG-19 networks

Overfitting is an important problem in training CNN networks. We used data augmentation, dropout, and L2 regularization, all of which are validated methods for avoiding overfitting [29, 30]. In addition, normalization, pooling, and dropout implementation in CNN middle layer has been shown to increase performance [31]. Dropout is the random selection and deactivation of a particular portion of neurons during training. Neurons are forced to learn, thus enhancing the network's generalization capability. L2 regularization improves generalization in linear models by penalizing weights in proportion to the sum of squares of weights [32]. We observed no signs of overfitting in the training graphs for the networks.

This study does have certain limitations: the severity of osteoarthritis can be determined by Kellgren Lawrence severity score. Osteophytes, joint space narrowing and subchondral sclerosis are evaluated in the Kellgren-Lawrence scoring system. Our dataset contained images of both early and late OA, and we did not code our CAD method to assess the severity of OA using the Kellgren-Lawrence scoring system, meaning that we cannot comment with a high degree of accuracy on the reliability of our system in recognizing early OA [33]. Further studies using more data should be conducted to assess whether this CAD method is capable of diagnosing early OA. We are collecting more patient data to improve our CAD model for this purpose.

This was a preliminary study performed in a single center, and state-of-the-art methods such as transfer learning and data augmentation were used to attain better accuracy. The accuracy, sensitivity, and specificity of all of the networks studied here were found to be promising. Our CAD model can support a clinician in the decision-making process towards the correct diagnosis of hand OA. The key interventions in treating OA are related to education and lifestyle modification for the patient, both of which lead to a better quality of life and course of disease in early OA patients. Our results suggest that the model is promising for clinical usage; our CAD system can allow primary care physicians to diagnose OA accurately and to take an active role in its management.

Conflict of interest

The authors declare that they have no conflicting interests.

Ethics approval

An ethical approval certicate was obtained from the Non-interventional Clinical Research 11 Ethics Board at Kırıkkale University. Certicate date: Sep 12th 2018; Certicate no: 2018.09.01.

Author contributions

All authors were fully involved in the preparation of this manuscript and approved the final version.

Funding

None.

References

[1] Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Clinics in Geriatric Medicine 2010; 26(3): 355-369.

[2] Hayashi D, Roemer FW, Guermazi A. Imaging for osteoarthritis. Annals of Physical and Rehabilitation Medicine 2016; 59(3): 161-169.

[3] Leung GL, Rainsford KD, Kean WF. Osteoarthritis of the hand i: aetiology and pathogenesis, risk factors, investigation and diagnosis. Journal of Pharmacy and Pharmacology 2014; 66(3): 339-346.

[4] Ramonda R, Frallonardo P, Musacchio E, Vio S, Punzi L. Joint and bone assessment in hand osteoarthritis. Clinical Rheumatology 2014; 33(1): 11-19.

[5] Schaefer LF, McAlindon TE, Eaton CB, Roberts MB, Haugen IK et al. The associations between radiographic hand osteoarthritis definitions and hand pain: data from the osteoarthritis initiative. Rheumatology International 2018; 38(3): 403-413.

[6] Rossignol M, Leclerc A, Allaert FA, Rozenberg S, Valat JP et al. Primary osteoarthritis of hip, knee, and hand in relation to occupational exposure. Occupational and Environmental Medicine 2005; 62(11): 772-777.

[7] Mathiessen A, Cimmino MA, Hammer HB, Haugen IK, Iagnocco A et al. Imaging of osteoarthritis (oa): What is new? Best Practice & Research Clinical Rheumatology 2016; 30(4): 653-669.

[8] Roemer FW, Eckstein F, Hayashi D, Guermazi A. The role of imaging in osteoarthritis. Best practice & Research Clinical Rheumatology 2014; 28(1): 31-60.

[9] Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Annals of The Rheumatic Diseases 1957; 16(4): 494.

[10] Greenspan H, Ginneken BV, Summers RM. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging 2016; 35(5): 1153-1159.

[11] Li H, Giger ML, Huynh BQ, Antropova NO. Deep learning in breast cancer risk assessment: evaluation of convolutional neural networks on a clinical dataset of full-field digital mammograms. Journal of Medical Imaging 2017; 4(4): 041304.

[12] Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in mri images. IEEE Transactions on Medical Imaging 2016; 35(5): 1240-1251.

[13] Setio AAA, Ciompi F, Litjens G, Gerke P, Jacobs C et al. Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. IEEE Transactions on Medical Imaging 2016; 35(5): 1160-1169.

[14] Üreten K, Erbay H, Maraş HH. Detection of rheumatoid arthritis from hand radiographs using a convolutional neural network. Clinical Rheumatology 2020; 39(4): 969-974

[15] Dreyer KJ, Geis JR. When machines think: radiology's next frontier. Radiology 2017; 285(3): 713-718.

[16] Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annual Review of Biomedical Engineering 2017; 19:221-248.

[17] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems; Red Hook, New York, USA; 2012. pp. 1097-1105.

[18] Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ et al. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. arXiv preprint arXiv: 2016; 1602.07360.

[19] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; Las Vegas, NV, USA; 2016. pp. 770-778.

[20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 2014; 1409.1556.

[21] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S et al. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition; Boston, MA, USA; 2015. pp. 1-9.

[22] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017; 542(7639): 115.

[23] Rawat, W and Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review In: Neural Computation 2017; 29(9): 2352-2449.

[24] Kim DH, MacKinnon T. Artificial intelligence in fracture detection: transfer learning from deep convolutional neural networks. Clinical Radiology 2018; 73(5): 439-445.

[25] Anwar SM , Majid M, Qayyum A, Awais M, Alnowami M et al. Medical image analysis using convolutional neural networks: a review. Journal of Medical Systems 2018; 42(11): 226.

[26] Xue Y, Zhang R, Deng Y, Chen K, Tao Jiang T. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. PloS one 2017; 12(6): e0178992.

[27] Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic knee osteoarthritis diagnosis from plain radiographs: A deep learning-based approach. Scientific Reports 2018; 8(1): 1727.

[28] Górriz M, Antony J, McGuinness K, Giró-i-Nieto X, O'Connor NE. Assessing Knee OA Severity with CNN attention-based end-to-end architectures, arXiv preprint arXiv: 2019; 1908.08856.

[29] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 2014; 15(1): 1929-1958.

[30] Wong SC, Gatt A, Stamatescu V, McDonnell MD. Understanding data augmentation for classification: when to warp? In: 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA); New York, USA; 2016. pp. 1-6.

[31] Tobias H, Barros P, Wermter S. The effects of regularization on learning facial expressions with convolutional neural networks. In : International Conference on Artificial Neural Networks; Barcelona, Spain; 2016. pp. 80-87.

[32] Phaisangittisagul, E. An analysis of the regularization between L2 and dropout in single hidden layer neural network In : 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS); Bangkok (Thailand); 2016. pp:174-179.

[33] Kohn MD, Sassoon AA, Fernando ND. Classifications in brief: Kellgren-Lawrence classification of osteoarthritis. Clinical Orthopaedics and Related Research 2016; 474(8): 1886-93.