 ©2020 Kvantovaya Elektronika and Turpion Ltd

https://doi.org/10.1070/QEL17177

Application of deep neural networks to improve diagnostic accuracy of rheumatoid arthritis using diffuse optical tomography

Yangqin Feng, D. Lighter, Lei Zhang, Yan Wang, H. Dehghani

Abstract. **A set of deep neural network models for rheumatoid arthritis (RA) classification using a highway network, a convolutional neural network and a residual network is proposed based on the data of diffuse optical tomography (DOT) utilising near-infrared light, which ensures early diagnosis of pathophysiological changes resulting from inflammation. A numerical model of the finger is used to generate images to overcome the inherent problem of insufficient clinical DOT images available. The proposed deep neural network models are applied to automatically classify simulated DOT images of inflamed and non-inflamed joints and transfer learning is also used to improve the performance of the classification. The results demonstrate that all three deep neural network methods improve the diagnostic accuracy as compared to the widely applied support vector machine (SVM), especially for high inter-subject variability databases. In cases of distinct modelled severity of disease, residual network achieved the highest accuracy (> 99%), and both of highway and convolutional neural networks reached 99%, respectively. However, as the severity of the modelled disease is reduced, this accuracy is reduced to 75.2% for residual networks. The results indicate that transfer learning can improve the performance of deep neural network methods on RA classification from DOT data and highlight their potential as a computer aided tool in DOT diagnostic systems.**

Keywords: *rheumatoid arthritis diagnosis, diffuse optical tomography, finger joints, deep neural networks, medical image classification.*

1. Introduction

Rheumatoid arthritis (RA) is a chronic condition associated with significant pain and disability [1], which is the most common type of inflammatory arthritis and affects between 0.5% to 1% of the world's population [2]. Joints most commonly affected by RA are the wrists, metatarsophalangeal and proximal interphalangeal (PIP) joints [1]. The first three to four months of symptoms provide a window of therapeutic opportunity [3], during which aggressive therapy leads to improved long-term patient outcome [4]. Development of medical imaging technologies in recent decades has allowed a more accurate detection of inflammation in patients with RA at early stages of disease progression, for example, with ultrasound [5, 6], or magnetic resonance imaging (MRI) [7, 8]; however, the need for experienced staff to operate both these modalities has led to high cost and limited availability. Therefore, the need for low cost, non-invasive and objective evaluation methodologies which can be operated by non-clinical staff is desired, making diffuse optical tomography (DOT) an attractive proposition for diagnosis and longitudinal monitoring of patients with RA.

DOT is an imaging technique where near-infrared (NIR) light is injected and detected at multiple locations on the boundary of biological tissue, to allow recovery of the underlying distribution of optical properties [9]. Recently, DOT systems have shown progress in many applications, including breast cancer detection, functional brain imaging and arthritic joint diagnosis [10–12]. Hielscher et al. [12] presented a single wavelength, frequency-domain DOT system capable of recovering absorption and scattering maps in finger joints. From analysis of basic statistical features from these images, both sensitivities and specificities of up to 85% were demonstrated for classifying inflamed joints of RA patients. Montejo et al. [13] further introduced approaches for extracting more complex heuristic features from DOT images of PIP joints and a method for using such derived features to diagnose RA. They also presented a comprehensive analysis of techniques for classification of these extracted features from the DOT image, including k-nearest-neighbours, linear and quadratic discriminant analysis, self-organising maps, and support vector machine (SVM) [14]. More recently, Lighter et al. [15] introduced a multispectral continuous wave (CW) DOT system to detect pathophysiological changes in inflamed RA finger joints. Recovered maps of oxygen saturation, total haemoglobin, water and scattering amplitude in healthy subjects were reported, with significantly greater inter-subject variability as compared to the variability observed within fingers from the same subject. Previously, studies of RA classification using single wavelength CW data have shown a weak discrimination using statistical features as extracted by Montejo et al. [12], where it has been shown that the accuracy of classification with CW-DOT images was not high (64% sensitivity and 55% specificity). The challenge faced is to discriminate between inflamed and non-inflamed cases, which can be due to noise, high inter-subject variability [15] and the ill-posed and underdetermined nature of the imaging as associated with

Y.Feng College of Computer Science, Sichuan University, Chengdu 610065, China; Imaging Lab of the School of Computer Science, University of Birmingham, Birmingham B152TT, UK;
D.Lighter, H.Dehghani Imaging Lab of the School of Computer Science, University of Birmingham, Birmingham B152TT, UK; e-mail: H.Dehghani@cs.bham.ac.uk;
L.Zhang, Y.Wang College of Computer Science, Sichuan University, Chengdu 610065, China

Received 31 October 2019
Kvantovaya Elektronika **50** (1) 21–32 (2020)
Submitted in English

single wavelength CW-DOT. Basic statistical features therefore are not sufficient for discriminating inflamed and non-inflamed conditions [12], and hence new methods for feature extraction from DOT RA images to allow disease classification are necessary, which is the topic of this presented work.

In recent years, deep neural network (DNN) approaches [16, 17], deep belief networks [18], and recurrent neural networks [19–22] have been successfully applied to the fields of image classification, speech recognition, visual tracking [23] and natural language processing [24]. In the domain of image classification, AlexNet [16], VGGNet [25] and GoogLeNet [26] are a few representative examples. These DNN models have shown great accomplishments in image classification tasks; however, they are all trained on a large set of labelled data. Although the medical image datasets are usually much smaller than the typical databases widely used, attempts have been made in the medical imaging domain as inspired by the success of DNN models in image classification. For example, to obtain a better representation of input data, patch-based deep neural network models have been proposed for breast cancer classification [27, 28], in which an AlexNet-based variant was used to extract the features and classify breast cancer histopathological images as either benign or malignant.

Tajbakhsh et al. [29] has investigated the performance of convolutional neural network (CNN) models in four distinct medical imaging applications, within three specialties (radiology, cardiology, and gastroenterology) involving transfer learning for classification, detection and segmentation. They found that a pre-trained CNN with adequate fine-tuning outperformed or performed as well as a CNN trained from scratch, and fine-tuned CNNs were more robust to the size of training sets than CNNs trained from scratch.

While there exist some efforts for computer-aided diagnosis of RA and some DNN models for medical image classification, little or no exploration into the application of DNN models with DOT images for diagnosis of RA has been carried out. In this work, we apply three deep learning methods, i.e. highway network [30], residual networks (ResNet) [31] and CNN, to achieve RA classification by using DOT images. Specifically, the basic concepts and the architectures of the three deep learning methods are presented. Also, the advantages and limitations of different methods are evaluated and analysed. The novelty and main contributions of this work include three points: 1) it is the first one that adopts and customises the deep neural networks to achieve the classification of RA using DOT images; 2) it applies the transfer learning to improve the performance of the models when the clinical data are not sufficient; and 3) to overcome the shortage of clinical images at the first stage of DOT imaging system, we propose a numerical model of the finger exhibiting inflammation to generate simulation data for the research on RA classification. This work has verified the possibility of adopting deep neural networks to achieve RA classification by using DOT images. The promising performance of the proposed methods demonstrates these deep models can contribute a lot to developing computer-aided systems for the clinical diagnosis of RA.

The remainder of this paper is organised as follows: Section 2 introduces the basic concepts of three deep neural networks and presents the developed models for RA classification. Section 3 introduces the utilised DOT images, while results are reported in Section 4 and Section 5 concludes this paper.

2. Methods

In this section, we introduce the concepts of three DNN approaches, which have achieved great success in image recognition, each of which shows distinctive advantages and limitations in different image classification tasks. Furthermore, the design of the DNN models based on these three approaches for RA classification and the details of how to use transfer learning to improve the classification accuracy are presented.

2.1. Highway network

Highway networks [30], inspired by long short-term memory (LSTM) recurrent neural networks, are methods for constructing feedforward networks with hundreds or thousands of layers. They are trained directly using stochastic gradient descent with a variety of activation functions and learned gating mechanisms to regulate information flow. The gating mechanisms allow neural networks to have paths for information to follow across different layers on information highways. The building block of the highway network is shown in Fig. 1. For a highway network, we define $a_i(l)$ as the output of the ith unit ($i \in \{1,2,...,N\}$) in the lth hidden layer ($l \in \{1,2,...,L-2\}$) such that,

$$a_i(l) = \mathrm{H}(\mathrm{net}_i^{\mathrm{H}}(l))\mathrm{T}(\mathrm{net}_i^{\mathrm{T}}(l)) + a_i(l-1)\mathrm{C}(\mathrm{net}_i^{\mathrm{C}}(l)), \qquad (1)$$

where N is the layer size, L is the total layer of a highway based model, H is a nonlinear transform followed by a nonlinear activation function, T is the transform gate and C is the carry gate. The two gates express how much of the output is produced by transforming the input and carrying it, respectively; $\mathrm{net}_i^{\mathrm{H}}(l)$, $\mathrm{net}_i^{\mathrm{T}}(l)$ and $\mathrm{net}_i^{\mathrm{C}}(l)$ are the total inputs of the nonlinear transform (H), transform gate (T) and carry gate (C) of the ith unit in the lth layer. They can compute as follows:

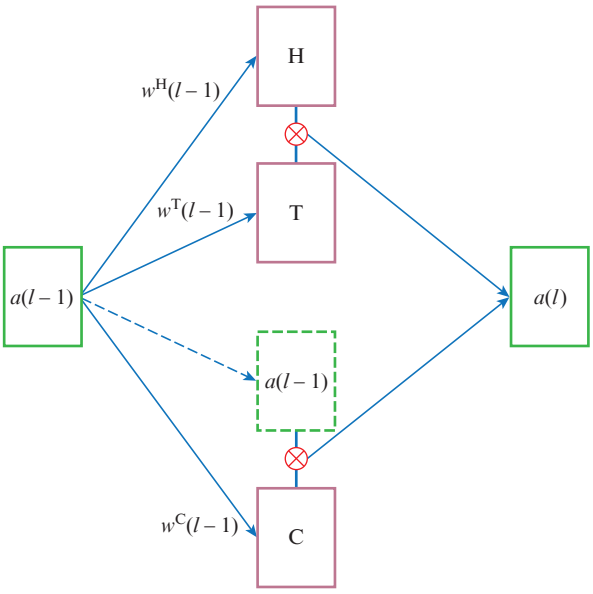

Figure 1. Building block of the highway network: H is a nonlinear transform and two additional nonlinear transforms (T and C) corresponding to the transform gate and the carry gate, respectively, and $a(l)$ is the output of the lth layer.

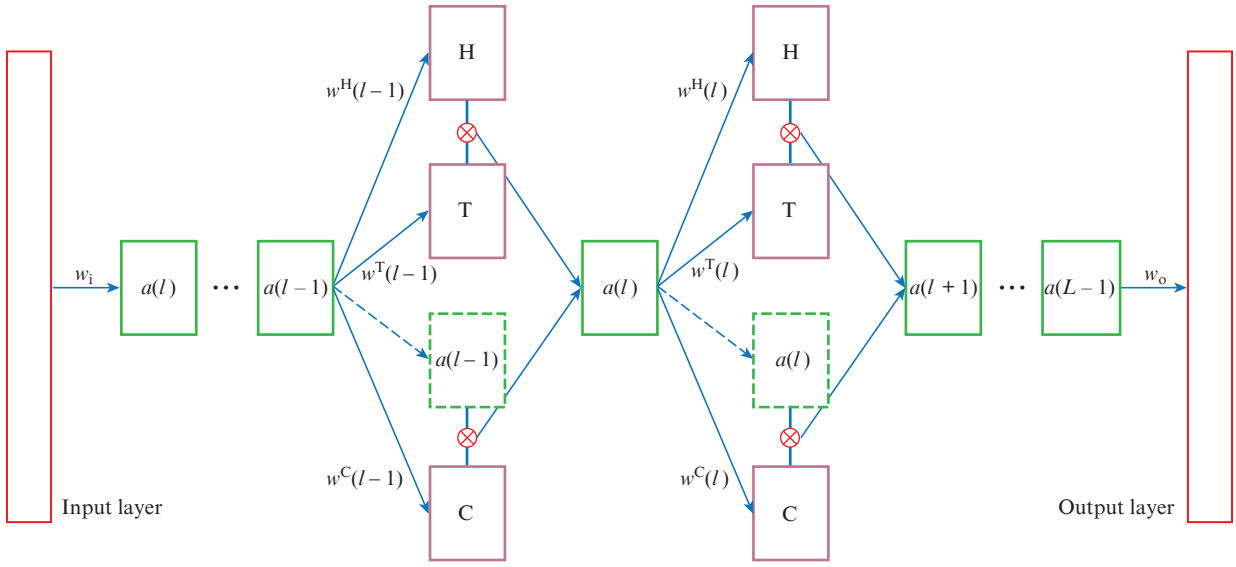

Figure 2. Proposed architecture of the highway network for RA classification. A highway based feedforward network is stacked with 18 hidden layers and 150 units in each layer; w_i is the weight of the input layer and w_o denotes the weight of the output layer.

$$\text{net}_i^{H}(l) = \sum[w_{ij}^{H}(l-1)a_j(l-1)], \quad (2)$$

$$\text{net}_i^{T}(l) = \sum[w_{ij}^{T}(l-1)a_j(l-1)], \quad (3)$$

$$\text{net}_i^{C}(l) = \sum[w_{ij}^{C}(l-1)a_j(l-1)]. \quad (4)$$

A customised highway-based feedforward network for RA classification is applied in this work, which consists of $L-2$ hidden layers (highway layers), an input layer and an output layer. The architecture of the proposed network for RA is shown in Fig. 2. Given a training set as $\boldsymbol{X} = \{\boldsymbol{x}(1), \boldsymbol{x}(2), ..., \boldsymbol{x}(M)\}$ and a testing image set as $\boldsymbol{T} = \{\boldsymbol{t}(1), \boldsymbol{t}(2), ..., \boldsymbol{t}(K)\}$, we use the training set to optimise the weights of the model and then use the trained weights to test the samples in the testing set.

Algorithm 1. The training and testing procedure of the highway network for our RA classification.

Input:
Training data set: $\boldsymbol{X} = \{\boldsymbol{x}(1), \boldsymbol{x}(2), ..., \boldsymbol{x}(M)\}$.
Testing dataset: $\boldsymbol{T} = \{\boldsymbol{t}(1), \boldsymbol{t}(2), ..., \boldsymbol{t}(K)\}$.
Iteration number: I_t.

Output:
Prediction for each testing image: $\boldsymbol{P} = \{\boldsymbol{p}(1), \boldsymbol{p}(2), ..., \boldsymbol{p}(K)\}$.

// Initialisation
1. Initialise $\{\boldsymbol{w}(l)^{T}, \boldsymbol{w}(l)^{H}, \boldsymbol{w}(l)^{C}\}_{i=1}^{(L-2)}$, $\boldsymbol{w}^{i}$ and $\boldsymbol{w}^{o}$.

// Highway network training
2. For each $t \in [1, I_t]$ do
3. Reshape each training image into a vector.
4. Do forward propagation for training images.
5. Update $\{\boldsymbol{w}(l)^{T}, \boldsymbol{w}(l)^{H}, \boldsymbol{w}(l)^{C}\}_{i=1}^{(L-2)}$, $\boldsymbol{w}^{i}$ and $\boldsymbol{w}^{o}$.
6. End for.
7. Reshape each testing image into a vector.
8. Do forward propagation for the testing image to compute the output of the network: $\boldsymbol{P} = \{\boldsymbol{p}(1), \boldsymbol{p}(2), ..., \boldsymbol{p}(K)\}$.
9. Return the predictions of the testing images: $\boldsymbol{P} = \{\boldsymbol{p}(1), \boldsymbol{p}(2), ..., \boldsymbol{p}(K)\}$.

In this work, the network parameters are experimentally set as $L = 20$ and $N = 150$. An input layer and a single softmax (n-normalised vector which is normalised into a probability distribution) output layer are included, meaning there are 18 highway layers in the proposed model. The training and the testing procedure is shown in Algorithm 1.

2.2. Convolutional neural networks (CNNs)

CNNs [32] are biologically-inspired variants of fully connected networks, designed to recognise visual patterns directly from 2D raw images. This is achieved with local connections and tied weights followed by some form of pooling which results in translation and shift of invariant features. CNNs are easier to train since they have much fewer parameters than fully connected networks with the same number of hidden units. Furthermore, images have a strong 2D local structure: Pixels that are spatially nearby are highly correlated, whilst the topology of the input is entirely ignored in fully connected networks. The extraction of local features is well considered by CNNs by restricting the receptive fields of hidden units to be local.

A CNN consists of a number of convolutional and pooling layers optionally followed by fully connected layers, with the convolutional layer consisting of several learnable filters for feature extraction. Each filter is convolved with the input data volume resulting in a 2D feature map that gives the responses of that filter at each spatial position. All of the feature maps are then stacked along the depth dimension to produce the output volume of the convolutional layer. An example of how a filter in a convolutional layer is used to process the input is shown in Fig. 3a. The pooling layer is used to reduce the spatial size of the representation, thus reduce the number of parameters and computation in the network. It operates independently on every depth slice of the input and resizes this slice spatially, using the MAX (for maximum) or AVG (for average) operation. An example of a pooling operation is shown in Fig. 3b. The fully connected layer consists of neurons that have full connections to all activations in the previous layer and output activations computed with a matrix multiplication followed by a bias offset.

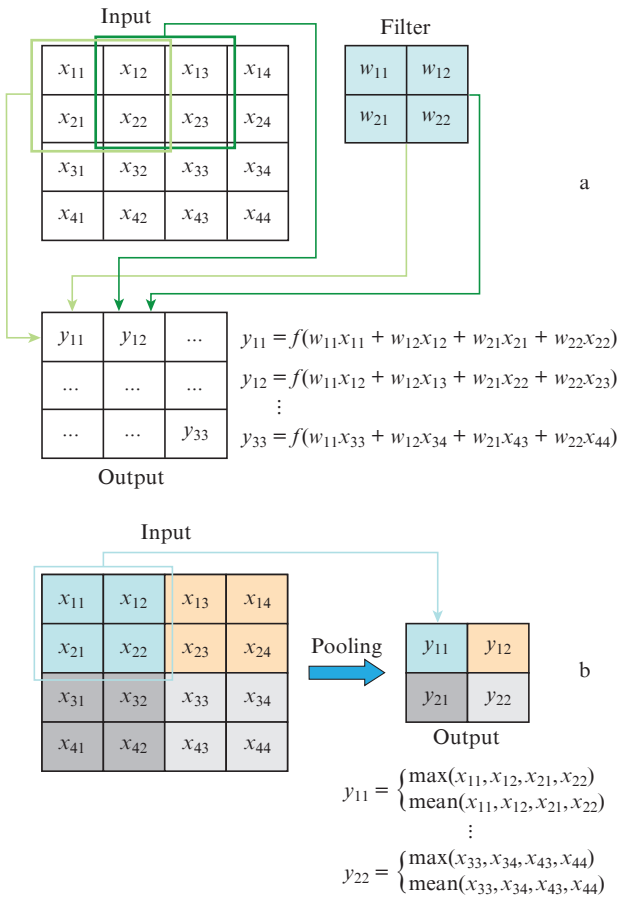

Figure 3. Example of (a) a convolutional operation and (b) a pooling operation: f is the activation function.

A customised convolutional network [32] is employed to achieve the RA classification. The architecture of the customised convolutional network is shown in Fig. 4. It comprises of nine layers, including the input layer. The second layer is a convolutional layer consisting of 32 filters with a size of 3×3 to produce 32 feature maps. The third layer is the same as the second layer, followed by a max pooling layer (size of 2×2, stride of 2) for the fourth layer which down samples every depth slice by 2 along both width and height, discarding 75% of the activations. The fifth and the sixth layers are both convolutional layers, consisting of 64 filters with the filter size of 3×3 to produce 64 feature maps. The seventh layer is a max-pooling layer with the same setting as the fourth layer. The last two layers are both fully connected layers with 512 and 2 output units, respectively. The activation function of all the convolutional layers and the seventh layer is a function of the rectified linear units (ReLu) [33], and the activation function of the last layer is a softmax function. When an image is loaded into this CNN, it will classify the input image as one of two classes, i. e. either inflamed or non-inflamed. The training and the testing procedure of CNN is shown in Algorithm 2.

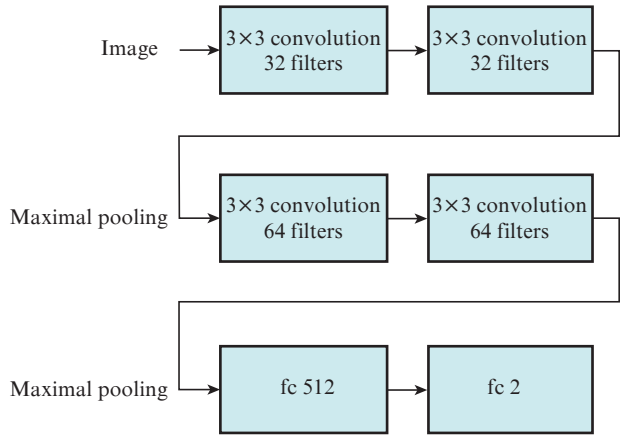

Figure 4. Architecture of the convolutional neural networks for RA classification. The max-pooling operator has a size of 2×2, and fc 512 and fc 2 layers denote fully connected layers with 512 output units and 2 output units, respectively.

Algorithm 2. The training and testing procedure of the CNN for our RA classification.

Input:
Training dataset: $\boldsymbol{X} = \{\boldsymbol{x}(1), \boldsymbol{x}(2), ..., \boldsymbol{x}(M)\}$.
Testing dataset: $\boldsymbol{T} = \{\boldsymbol{t}(1), \boldsymbol{t}(2), ..., \boldsymbol{t}(K)\}$.
Iteration number: I_t.

Output:
Prediction for each testing image: $\boldsymbol{P} = \{\boldsymbol{p}(1), \boldsymbol{p}(2), ..., \boldsymbol{p}(K)\}$.

// Initialisation
1. Initialise the weights of filters and the fully-connected layers.

// Training the CNN
2. For each $t \in [1, I_t]$ do
3. Do forward propagation for training images.
4. Update the weights in all the filters and fully connected layers.
5. End for.
6. Do forward propagation for testing image to compute the output of the network: $\boldsymbol{P} = \{\boldsymbol{p}(1), \boldsymbol{p}(2), ..., \boldsymbol{p}(K)\}$.
7. Return the predictions of the testing images: $\boldsymbol{P} = \{\boldsymbol{p}(1), \boldsymbol{p}(2), ..., \boldsymbol{p}(K)\}$.

2.3. Residual network (ResNet)

ResNet [31] adopts residual learning to address the degradation problem that the accuracy gets saturated and then degrades rapidly if the depth of a network increases. Instead of aiming that each few stacked layers directly fit a desired underlying mapping, residual learning explicitly allows these layers to fit a residual map instead. Although both forms are able to approximate the desired functions with the universal approximation theorem [34], the learning in residual mapping is much easier if the desired function is closer to an identity mapping than to a zero mapping. As introduced in [31], if the added layers can be constructed as identity mappings, a deeper model should have a training error no greater than its shallower counterpart. From this point of view, the residual learning allows the very deep networks to potentially avoid the degradation problem, and hence has achieved state-of-the-art performance on a number of visual tasks [31].

The basic building block of ResNet is shown in Fig. 5, from which we can see that the input $\boldsymbol{x}$ is firstly transformed by a residual mapping with two layers, then the sum of the input and the result of the residual mapping is taken as the output. Formally, the residual building block can be defined as

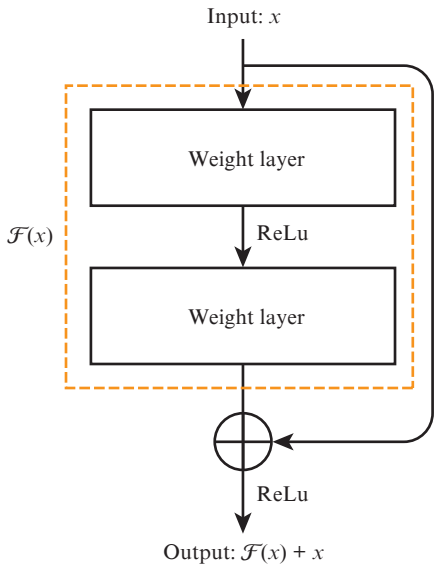

Figure 5. Building block of ResNet [31]. The operation inside the dash box represents the residual mapping process.

$$\boldsymbol{y} = \mathcal{F}(\boldsymbol{x}) + \boldsymbol{x}, \tag{5}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the input and output vectors of the layers considered, and the function $\mathcal{F}(\boldsymbol{x})$ denotes the residual mapping. As there are two layers in Fig. 5, the calculation of the residual mapping is referred to $\mathcal{F}(\boldsymbol{x}) = \boldsymbol{w}(2)\sigma(\boldsymbol{w}(1)\boldsymbol{x})$ in which σ denotes the function of the rectified linear units (ReLu) [33], $\boldsymbol{w}(1)$ and $\boldsymbol{w}(2)$ are the weight matrices of the first and the second layers, and the biases are omitted for simplifying notations. Finally, the second nonlinear mapping σ is conducted on the addition $\boldsymbol{y}$ to be the output of the building block.

As can be seen the dimensions of $\boldsymbol{x}$ and $\mathcal{F}$ must be equal in Eqn (5). If this is not the case, a linear projection $\boldsymbol{P}$ can be performed on $\boldsymbol{x}$ to match the dimension as

$$\boldsymbol{y} = \mathcal{F}(\boldsymbol{x}) + \boldsymbol{P}\boldsymbol{x}. \tag{6}$$

Furthermore, the above formulations are about fully connected layers for simplicity; they are also applicable to convolutional layers.

In our work, by stacking the residual learning building blocks, a deep residual network can be constructed to extract the discriminative features from the raw images. The architecture of ResNet for our task is shown in Fig. 6. Specifically, inspired by the philosophy of VGGnets, ResNet is constructed with convolutional layers that mostly have 3×3 filters and following two simple design rules: 1) for the same output feature map size, the layers have the same number of filters; and 2) if the feature map size is halved, the number of filters is doubled to preserve the time complexity per layer. The down sampling is performed directly by convolutional layers that have a stride of 2. The network ends with a global average pooling layer and a g-way fully connected layer with softmax, where g denotes the number of image classes in the visual task to classify the image into either inflamed or non-inflamed. The training and the testing procedure of ResNet is shown in Algorithm 3.

Algorithm 3. The training and testing procedure of ResNet for our RA classification.

Input:
Training dataset: $\boldsymbol{X} = \{\boldsymbol{x}(1), \boldsymbol{x}(2), ..., \boldsymbol{x}(M)\}$.
Testing dataset: $\boldsymbol{T} = \{\boldsymbol{t}(1), \boldsymbol{t}(2), ..., \boldsymbol{t}(K)\}$.
Iteration number: I_t.

Output:
Prediction for each testing image: $\boldsymbol{P} = \{\boldsymbol{p}(1), \boldsymbol{p}(2), ..., \boldsymbol{p}(K)\}$.

// Initialisation
1. Initialise the weights of filters and the fully connected layers.

// Training ResNet
2. For each $t \in [1, I_t]$ do
3. Do forward propagation for training images.
4. Update the weights in all the filters and fully connected layers.
5. End for.
6. Do forward propagation for testing image to compute the output of the network: $\boldsymbol{P} = \{\boldsymbol{p}(1), \boldsymbol{p}(2), ..., \boldsymbol{p}(K)\}$.
7. Return the predictions of the testing images: $\boldsymbol{P} = \{\boldsymbol{p}(1), \boldsymbol{p}(2), ..., \boldsymbol{p}(K)\}$.

2.4. Transfer learning

In the machine learning domain, transfer learning is the ability of a system to recognise and apply knowledge and skills learned in previous tasks to new tasks. It explores how models would learn from one task and apply these learning skills to other similar task. In recent decades, researchers have applied

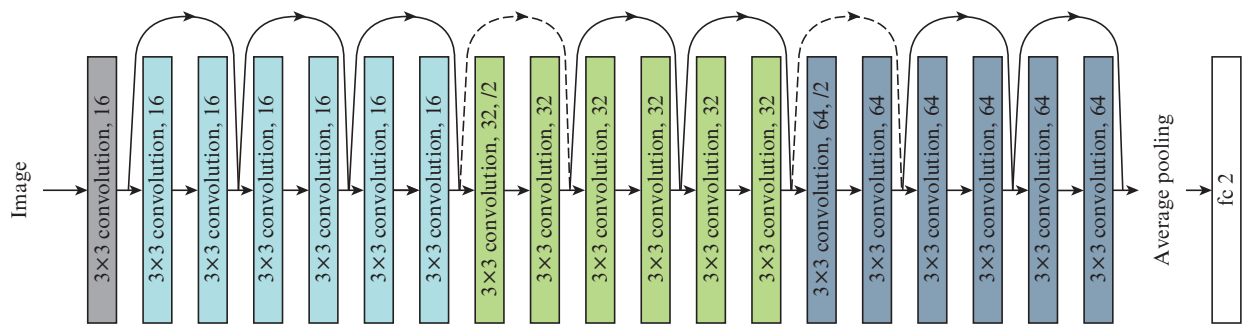

Figure 6. Architecture of the residual network with 20 layers designed for RA classification. The dotted shortcuts increase dimensions. The input image data flow through a number of residual network blocks, then are operated with an average pooling (pool size of 2×2) and fully connected to the output layer with softmax activation function.

techniques for transfer learning in text classification [35–37], face recognition [38, 39], speech recognition [40, 41] and reinforcement learning [42].

In medical diagnostic tasks, obtaining medical image samples is expensive and time consuming, especially when accurate ground truth labels for images are required. Generating simulated images which are similar to the clinical data and using transfer learning may be a good way to improve the performance of a deep learning model. To do so, simulation data can be generated and used to pre-train the DNNs, i. e., highway, ResNet and CNN and the network parameters can be fine-tuned with real-world clinical data. If the generated images and the real-world clinical images share some similar features, the DNNs are potentially able to apply the knowledge learned from the simulation data to improve their performance on real-world clinical data.

2.5. Comparing with traditional machine learning algorithms

There are many traditional machine learning algorithms that have been used in the medical image classification. For instance, support vector machine (SVM) [43], decision tree (DT) [44] and random forest classifier (RFC) [45]. SVM constructs a hyperplane in high-dimensional space to separate different classes. The core idea of SVM is to find a maximum marginal hyperplane that divides the data samples into correct classes. DT is a non-parametric supervised learning method. Its goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. RFC is a meta-estimator that fits several decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. However, these traditional methods require human expertise for feature extraction to achieve high accuracy on medical image classification tasks. Alternatively, deep neural networks can achieve great power and flexibility automatically by learning to represent the medical images as a nested hierarchy of concepts.

3. Clinical finger images

Between March and August in 2018, a controlled pilot study was carried out, in which patients were recruited through the outpatient clinic in the University of Birmingham Research Laboratories, Queen Elizabeth Hospital, Birmingham. Ethical approval was obtained as part of the University Hospitals Birmingham "Prediction of outcomes in patients with inflammatory arthritis" (RRK4678) research study, with all subjects providing written informed consent prior to participating. Specifications and acquisition protocols of the multispectral, non-contact DOT system have been outlined previously [15].

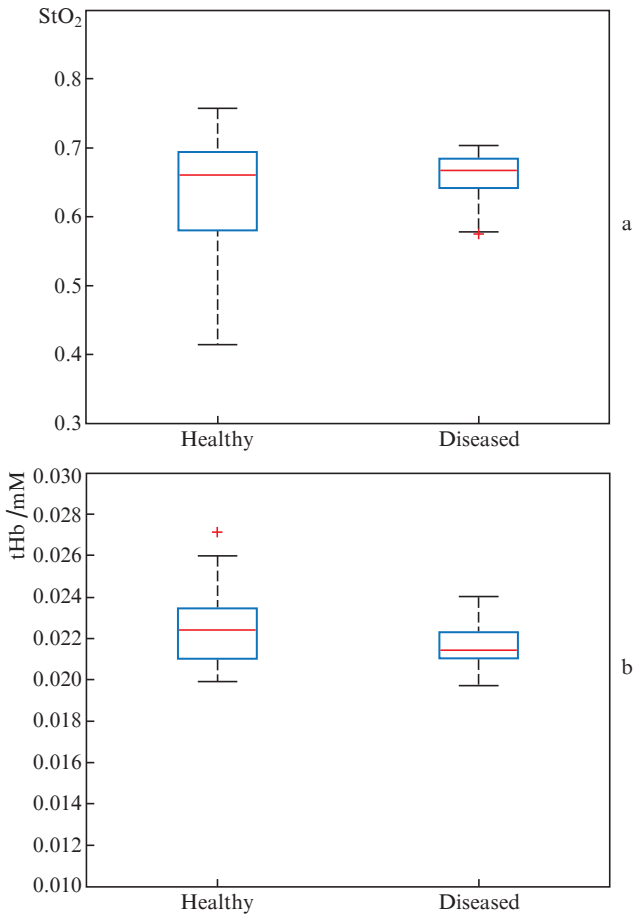

Figure 8. Plots of StO_2 and tHb from 88 imaged fingers: a total of 74 non-inflamed and 14 inflamed fingers were enrolled.

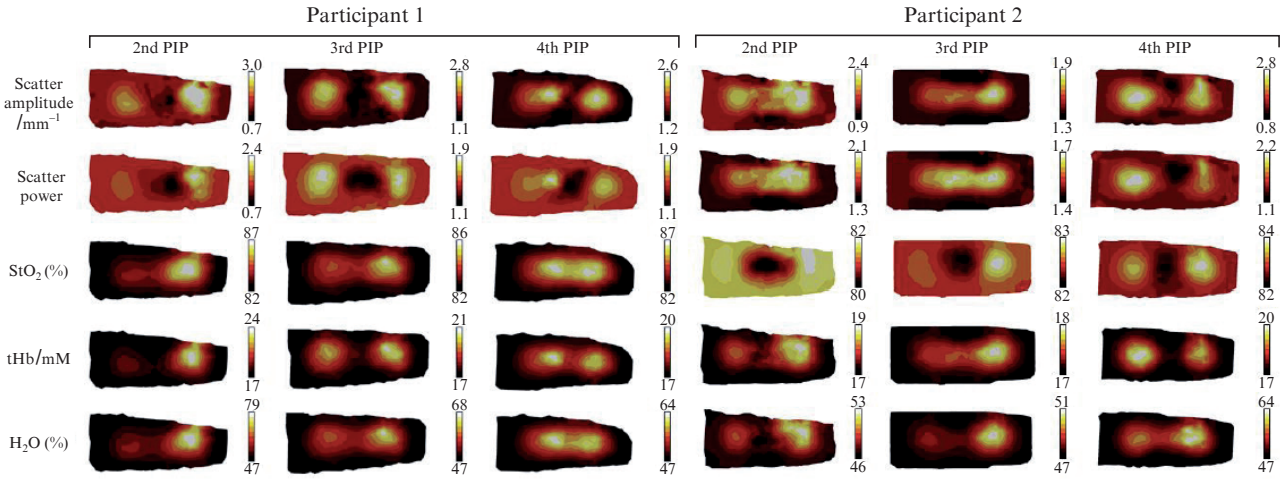

Figure 7. Central transverse slices of reconstructed images of pathophysiologic parameters from three PIP joints of two healthy participants.

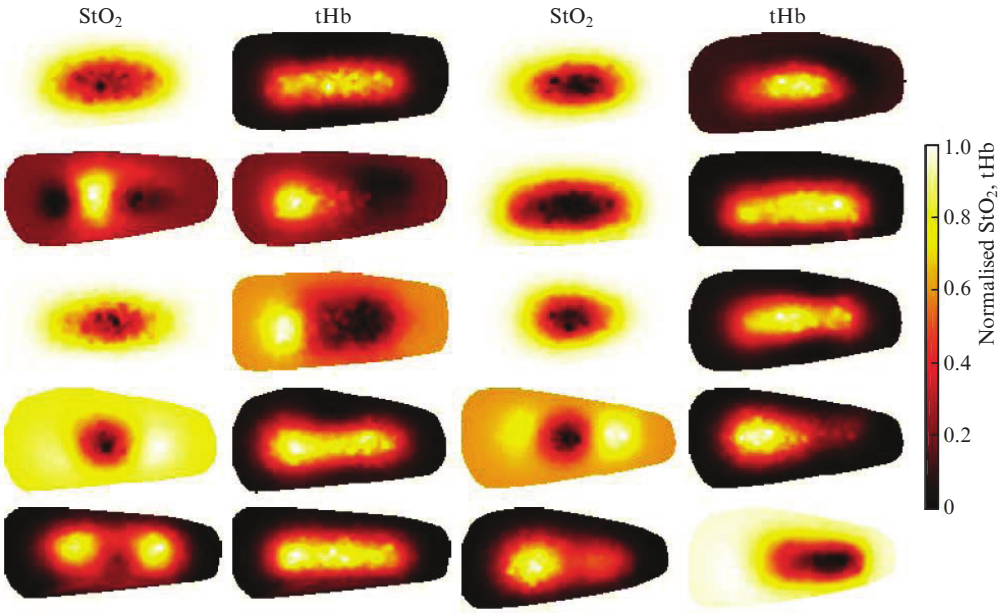

Figure 9. Clinical images to illustrate weak discrimination. The first two columns are normalised healthy StO_2 and tHb images, each row of them sampled from the same finger, and the last two columns are normalised diseased StO_2 and tHb images, each row of them sampled from the same finger as well.

In brief, patients placed their hands on the platform of the system, which collected optical transmission images of the finger using an air-cooled, charged couple device camera, whilst broadband, point source light was injected into the opposing side of the joint. Images were collected at 14 source positions in a straight line along the sagittal direction of the finger, repeated at five wavelengths (650, 710, 730, 830, 930 nm) by spectral decoupling using a filter wheel preceding the camera objective lens. 3D tomographic maps of oxygen saturation (StO_2), total haemoglobin (tHb), water concentration (H_2O), scatter power (SP) and scatter amplitude (SA) were recovered in finger joints using NIRFAST [46].

In this work, data are collected from 13 healthy and RA volunteers. Imaged joints included the II, III, IV and for some later participants the V PIP joints, on each hand, giving a total of 88 finger joints. A typical example of healthy PIP joint for the transverse is displayed in Fig. 7, where the 2D slices were taken along the central transverse plane of the joint, which shows the reconstructed image maps corresponding to the clinically relevant parameters StO_2, tHb, SP, SA and H_2O. In an inflamed joint of a patient with RA, known pathophysiological changes include lower StO_2 (hypoxia) [47] and higher tHb (synovial angiogenesis) [48], as compared to healthy joints. Therefore, StO_2 and tHb images were used to diagnose the inflamed and non-inflamed fingers in this work.

It is challenging to classify the images of StO_2 and tHb since there exists a high inter subject variability, which is consistent with the results as published by Lighter et al. [15]. To illustrate the variance in these imaged features for the healthy and diseased subject data, plots of the recovered StO_2 and tHb distributions for all of the imaged 88 fingers are shown in Fig. 8. High inter-subject variability will lead to weak discrimination of inflamed and non-inflamed subjects and Fig. 9 shows an example of normalised diseased StO_2 and tHb images to illustrate this weak discrimination. All the images shown in Fig. 9 are a centred 2D slice from each PIP joints, and self-normalised by the maximum value for each image.

4. Simulation experiments

In an inflamed joint the multispectral CW DOT-based imaging system outlined above aims to recover optical parameters as spatial maps for each finger which are the input of the outlined classification methods. The process of collecting labelled clinical data is time-consuming; therefore, to investigate the behaviour of the DNN models a set of data from simulated models have been generated, which includes both the healthy and diseased finger samples from the experiment. Additionally, the experiments on simulated data were specifically conducted to investigate the behaviour of the DNN algorithms since the discrimination level of the simulated data can be controlled and exact classification is known. Without this methodology, DNN will not be possible due to the shortage and otherwise unavailable data.

4.1. Simulation finger joint image

In order to generate a set of simulated data for this work, a two-dimensional model of a finger joint, as shown in Fig. 10, was created consisting of skin, bone, joint and muscle, with the estimates of StO_2, tHb and H_2O of different finger tissues based on literature values [49, 50] as shown in Table 1. These values were assigned to a finite element model mesh with 2747 nodes and 5280 linear tetrahedral elements. An array of 8 sources below the finger joint and 8 detectors on the top of the finger joint, evenly spaced with 2.86 mm separation, was assigned to this model. For each model, a set of value for both StO_2 and tHb has been defined for the joint space, depending on whether the tissue was healthy (non-inflamed) or diseased (inflamed) to reflect these distinctions. Data was generated for each simulated finger model, with 1% Gaussian noise added from which images of StO_2 and tHb were reconstructed using an iterative Levenberg–Marquardt procedure. All simulations were carried out using open source finite element met-

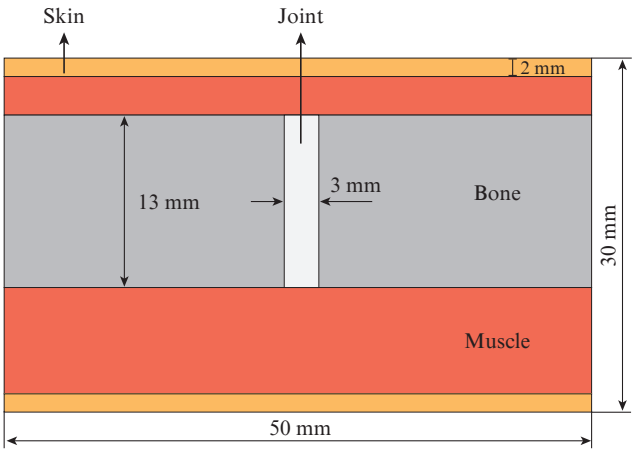

Figure 10. Two dimensional model of a finger.

Table 1. Parameters for different finger tissues.

Finger tissue	StO_2	tHb/mM L^{-1}	Water
Skin	0.75	0.06	0.50
Muscle	0.80	0.10	0.50
Bone	0.80	0.08	0.40
Joint	Def	Def	0.50

Note: StO_2 and water are measured in fractions. For the joint, StO_2 and tHb were defined based on the levels of hypoxia and angiogenesis between healthy and diseased joints, denoted by Def. For different finger tissue, SA was calculated based on the data from [51].

Algorithm 4. The procedure of generating StO_2 and tHb images.

Input:
Value of StO_2.
Value of tHb.
Noise level.

Output:
One StO_2 image and one tHb image.

// For each healthy finger to generate an image:
1. Generate a physically realistic model (2D) based on the StO_2 value, tHb value and other parameters in Table 1.
2. Add the Gaussian noise percentage to the model.
3. Reconstruct the image from the model.
4. Return one StO_2 image and one tHb image.

hod package (NIRFAST) for modelling light propagation [46]. The procedure of generating an StO_2 image and a tHb image is shown in Algorithm 4.

4.2. Simulation database

Four different simulation databases were generated for the experiment, each of which contained 800 simulated healthy fingers and 800 diseased fingers with two images (StO_2 image and tHb image) for each. Therefore, overall 1600 StO_2 images

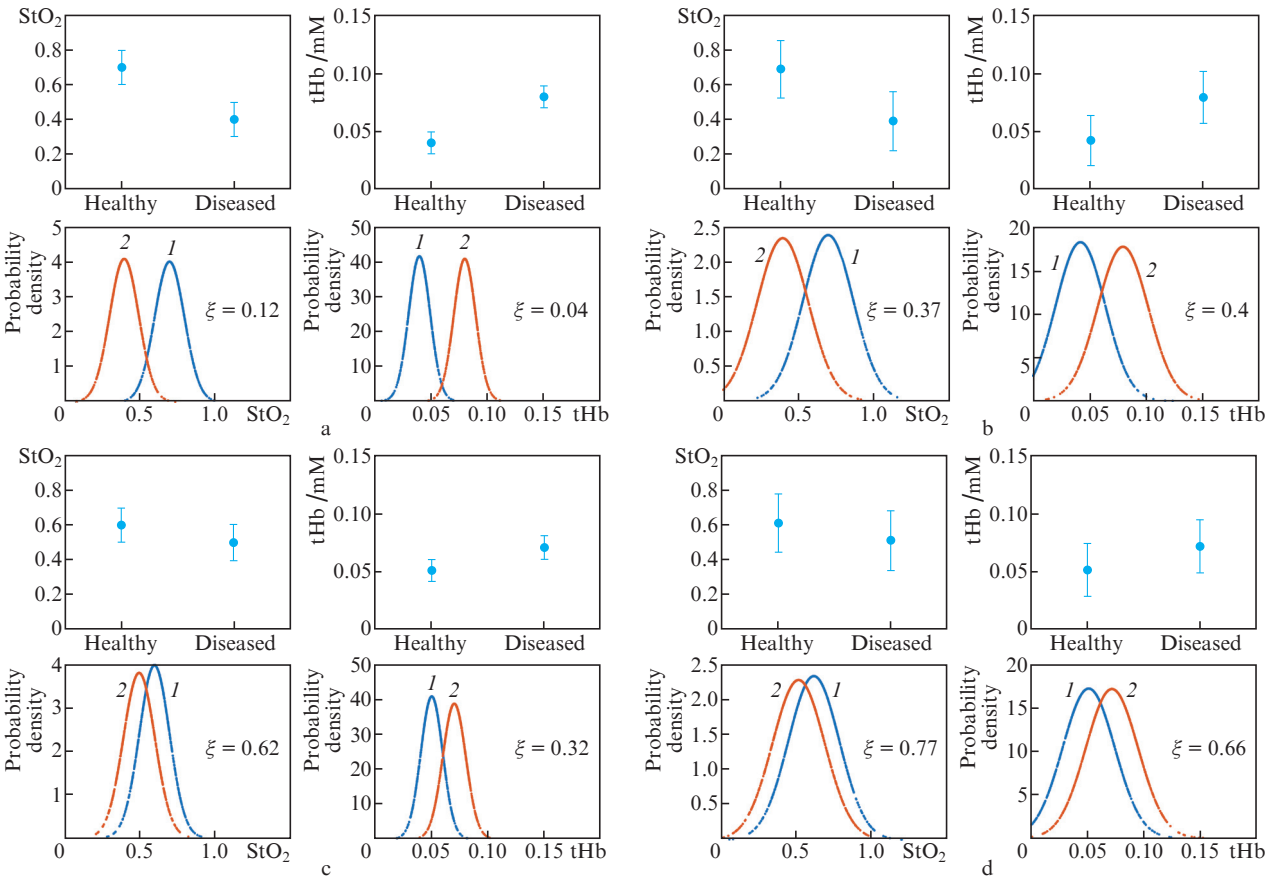

Figure 11. Four data distributions of physiological parameters assigned to the joint space denoted by Def for simulation: (*1*) healthy and (*2*) diseased fingers; ξ is the overlapping coefficient.

and 1600 tHb images are included in each simulation database. Figure 11 shows the distribution of StO_2 and tHb values for each simulation database, from which images of StO_2 and tHb are reconstructed. All the reconstructed images are normalised by

$$z_i = (x_i - \min\boldsymbol{P})/(\max\boldsymbol{P} - \min\boldsymbol{P}),$$

where z_i is the normalised image from the image x_i, and $\max\boldsymbol{P}$ and $\min\boldsymbol{P}$ is the max and min values of the database $\boldsymbol{P}$. The four simulation databases were controlled at different levels of statistical discrimination by overlapping coefficient; the overlapping coefficient is defined as the measure of agreement between probability distributions and point estimation of the overlap of two normal densities. The greater the value of the overlapping coefficient, the greater the inseparability of the healthy finger and the diseased finger. The mean value of StO_2 for healthy joints is higher than that for diseased joints, and the mean value of tHb for healthy joints is lower than that for diseased joints. Database 1 (hdls) and database 2 (hdhs) have the same mean value of StO_2 and tHb, while database 1 has a lower standard deviation than database 2. Database 3 (ldls) and database 4 (ldhs) have the same mean value of StO_2 and tHb, while database 3 has lower standard deviation than data database 4. Compared to database 1 and database 2, the difference between the mean value of healthy values and diseased values are diminished for database 3 and database 4. In addition, the four simulation databases are controlled at different levels of difficulty by the overlapping coefficient (a higher overlapping coefficient will lead to a weaker discrimination) as it is not clear at this stage how different these populations (i. e. healthy and diseased) will be in large studies and the efficacious of the DNN models for different population distributions need to be investigated. The details of the settings for the four databases are listed in Table 2 and shown in Fig. 11.

There are two challenges for the classification task based on these four defined simulation databases.

1. Weak discrimination: All four different databases are based on overlapping tissue parameters, which leads to weak discrimination between healthy and diseased fingers. The details have been shown in Fig. 11, for example, where there is 37% overlap for StO_2 and 40% overlap for tHb values between healthy and diseased joints. This implies that there exists StO_2 and tHb images exhibiting the same distribution but they belong to the different classes (healthy and diseased).

2. Noise: All of the simulations have 1% Gaussian noise which is used to reconstruct images to allow for a more realistic experimental scenario, with an example shown in Fig. 12. It is evident that there will exist image artefacts in both healthy and diseased joints, which will lead to weak discrimination between healthy and diseased fingers.

Table 2. Settings for the four databases.

Database	Healthy joints (StO_2)	Diseased joints (StO_2)	Healthy joints (tHb)	Diseased joints (tHb)
1 (hdls)	0.70 ± 0.10	0.40 ± 0.10	0.04 ± 0.01	0.08 ± 0.01
2 (hdhs)	0.70 ± 0.17	0.40 ± 0.17	0.04 ± 0.023	0.08 ± 0.023
3 (ldls)	0.60 ± 0.10	0.50 ± 0.10	0.05 ± 0.01	0.07 ± 0.01
4 (ldhs)	0.60 ± 0.17	0.50 ± 0.17	0.05 ± 0.023	0.07 ± 0.023

Note: hd and ld denote a high difference and a low difference between mean values for healthy and diseased joints (both StO_2 and tHb mean values); hs and ls denote a high standard deviation and a low standard deviation for healthy and diseased joints (both StO_2 and tHb mean values).

4.3. Results and analysis

In the first experiment, a new image is generated for each finger by simply connecting (adding) the StO_2 and tHb image which are treated as a new 'optical index' parameter. The influence on RA classification is then investigated by using different images, such as only the StO_2 or tHb or StO_2 + tHb images as inputs of the neural networks. The results are presented as the mean accuracy and the standard deviation of 5-fold cross-validation, with Table 3 showing the results of highway, ResNet, CNN, as well as the results using SVM [43], DT [44] and RFC [45], which are taken as the baseline (a multi-class linear SVM method with the one-versus-all strat-

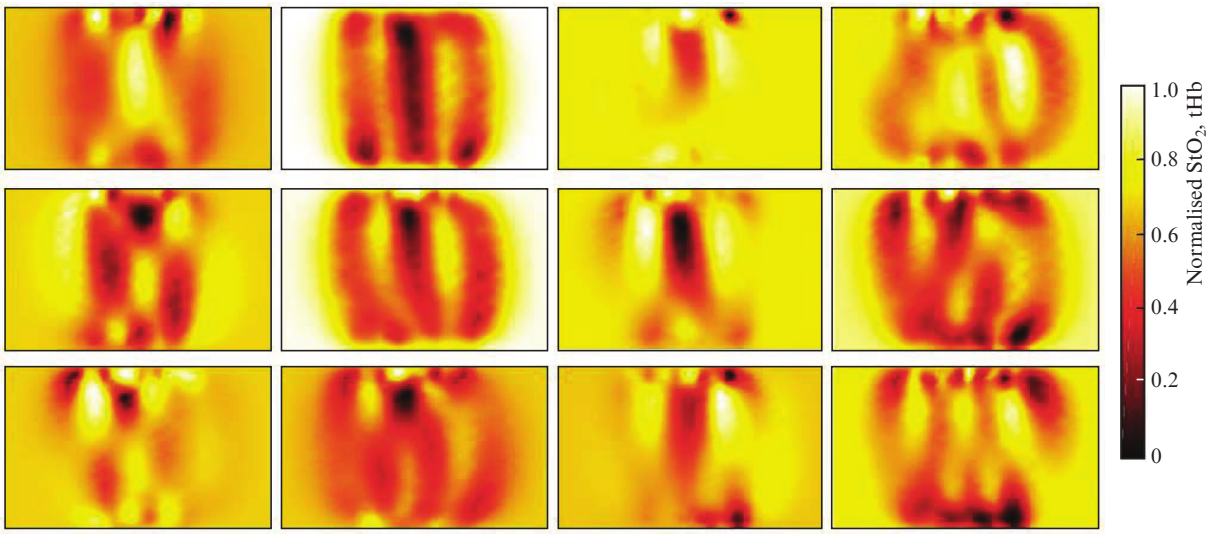

Figure 12. Example images chosen from simulation database 1 (hdls) to demonstrate the image artefacts. The first two columns are normalised healthy StO_2 and tHb images, each row of them sampled from the same finger, and the last two columns are normalised diseased StO_2 and tHb images, each row of them sampled from the same finger as well.

Table 3. Comparison of the mean accuracy (%) and standard deviation with all the three deep learning methods and SVM, DT and RFC.

Database	Input image	Highway	ResNet	CNN	SVM	DT	RFC
1	StO_2	97.6 ± 1.8	98.6 ± 0.6	95.3 ± 1.9†	96.8 ± 1.4†	90.6 ± 1.5†	92.9 ± 1.8†
	tHb	98.8 ± 0.5	98.9 ± 0.6	98.7 ± 0.8	97.9 ± 0.7	96.3 ± 1.0†	97.3 ± 1.4
	StO_2 + tHb	99.3 ± 0.5	99.3 ± 0.4	99.1 ± 0.7	98.0 ± 0.8†	97.1 ± 0.8†	98.3 ± 1.0
2	StO_2	88.3 ± 1.8	88.8 ± 1.9	87.8 ± 2.3	79.1 ± 10.7†	79.3 ± 2.6†	84.6 ± 2.7
	tHb	88.4 ± 0.8	89.1 ± 1.8	88.4 ± 1.6	72.9 ± 17.9†	77.7 ± 1.7†	83.3 ± 1.3†
	StO_2 + tHb	89.4 ± 1.6	90.2 ± 1.4	89.0 ± 1.6	79.4 ± 14.4†	80.6 ± 1.4†	87.3 ± 1.6†
3	StO_2	80.1 ± 6.0†	86.3 ± 3.2	78.8 ± 5.1†	77.3 ± 5.5†	66.9 ± 3.3†	72.2 ± 4.1†
	tHb	86.3 ± 1.8	87.0 ± 4.4	86.3 ± 2.0	83.4 ± 1.2	75.9 ± 2.7†	82.1 ± 1.8
	StO_2 + tHb	86.8 ± 2.0	88.4 ± 2.0	86.6 ± 1.9	75.9 ± 13.7†	77.1 ± 1.4†	82.1 ± 2.0†
4	StO_2	68.6 ± 2.6	70.1 ± 3.3	68.6 ± 2.7	54.2 ± 3.6†	57.4 ± 2.7†	61.6 ± 1.4†
	tHb	70.5 ± 1.5	73.8 ± 5.3	70.4 ± 2.7	59.8 ± 5.1†	57.6 ± 2.8†	64.8 ± 2.1†
	StO_2 + tHb	71.4 ± 2.7	75.2 ± 8.8	71.6 ± 1.7	55.0 ± 8.6†	57.1 ± 3.6†	65.1 ± 1.7†

Note: The '†' indicates that the accuracy of the method is significantly different from the accuracy of ResNet at a 0.05 level by the Wilcoxon's rank sum test.

egy of the penalty parameter of 1.0 and a RFC method with 10 trees are tested).

The results show that all four methods perform well on the images from database 1. Specifically, they all achieve a mean classification accuracy higher than 95%. ResNet obtains the highest mean accuracy of 98.6% on StO_2 images, and it also obtains the highest mean accuracy of 98.9% on tHb images, and highway and ResNet performs best on StO_2 + tHb images with the mean accuracy of 99.3%. For all six algorithms, as the overlap of the simulated healthy and diseased classes increased, the performance of all methods decrease, especially for the SVM and DT methods. In terms of algorithms, the ResNet method outperforms the other three methods on the images from all four different databases. The performance gap between ResNet and the other three methods increases along with the increase in the overlap between the samples of the healthy and diseased classes. The three deep neural network models, highway, ResNet and CNN, obtain the best results on StO_2 + tHb images, and better results on tHb images as compared to StO_2 images.

In the second experiment, the impact of transfer learning is investigated using the four databases which have different overlaps for the deep neural networks. Specifically, the weight parameters trained on the database 1 are used to initialise the parameters of the networks for classification tasks on the other three databases (databases 2–4).

Table 4. Comparison of the mean accuracy (%) and standard deviation with all the three deep learning methods with transfer learning.

Database	Input image	Highway	ResNet	CNN
2	StO_2	88.6 ± 1.0	90.4 ± 1.5	88.1 ± 2.3
	tHb	88.6 ± 1.1	90.1 ± 1.7	88.5 ± 1.3
	StO_2 + tHb	91.9 ± 1.4†	92.8 ± 0.9†	90.7 ± 1.5
3	StO_2	82.6 ± 1.1	88.1 ± 3.4	81.8 ± 2.6
	tHb	87.9 ± 1.9	88.6 ± 1.4	86.5 ± 1.4
	StO_2 + tHb	87.6 ± 1.5	90.3 ± 2.5	87.8 ± 1.8
4	StO_2	69.3 ± 1.4	74.8 ± 2.5†	69.4 ± 2.9
	tHb	73.2 ± 1.7†	75.9 ± 4.8	74.1 ± 2.1†
	StO_2 + tHb	73.9 ± 1.4	77.3 ± 4.5	75.8 ± 1.9†

Note: The '†' indicates that the accuracy of the method is significantly different from the accuracy in Table 3 at a 0.05 level by the Wilcoxon's rank sum test.

The 5-fold cross validation method is employed to obtain the statistical results, as summarised in Table 4. The results show that all the three deep neural network models have a better performance when use is made of transfer learning. For example, ResNet obtains the highest mean accuracy of 90.4% on StO_2 images on database 2, and it is also performs best on tHb and StO_2 + tHb images with the mean accuracy of 90.1% and 92.8%. It is shown that transfer learning can help improve the performance on DOT images. This highlights the benefit of using transfer learning through different simulation for the prediction of class when there is a lack of experimental data.

5. Conclusions

Achieving a high diagnostic accuracy of inflammation in rheumatoid arthritis does not only rely on the separate-ability of data (low variation within each class and a large discrepancy between the two classes), but also on an accurate and objective classification algorithm. In this paper, a method for generating simulated healthy and diseased DOT images of finger joints are presented, with three DNN models for classification. The goal is then to classify each PIP joint as an inflamed or non-inflamed with RA based on the analysis of the image features which are extracted by DNN models automatically. Specifically, three state-of-the-art DNN models have been investigated, including highway networks, CNN and ResNet, to extract discriminative features from DOT images to improve the diagnostic accuracy. To help understand the classification algorithms, four databases are defined which are controlled in terms of overlapping coefficients for StO_2 and tHb. The DNN models were then investigated on their efficacy to improve the diagnostic accuracy in studies including patients with arthritis.

It was first demonstrated through patient and healthy human subject data that the images recovered for StO_2 and tHb from inflamed and non-inflamed finger joints demonstrate a high inter-subject variability. The underlying challenge is that (1) the healthy and diseased finger DOT images share similar patterns which would lead to weak discrimination of healthy and diseased fingers, (2) the degree of the overlap between

different subjects are unknown in a clinical setting, and (3) not enough clinical data exists for a detailed evaluation of DNN based algorithms. Therefore, a model of a finger joint has been utilised to generate databases which have allowed the investigation of the influence of weak discrimination when using computer aided diagnostic methods.

Some researchers have used the traditional methods, e. g., SVM, random forest regressor for recognition of the spatial frequency domain imaging data [52, 53]. Also, the SVM method is applied to understand finger DOT images [14]. In this work, we aim to apply deep learning methods to help diagnose finger DOT images. Furthermore, raw DOT images are hard to interpret and so we developed the simulated model to control the difficulty levels of the RA classification tasks to investigate the capability of the tested methods. It helps us to understand what algorithms work better than others in different conditions. In the simulation experiments, it shows that all the three DNN models, SVM and RFC are performing better on StO_2 + tHb images as compared to StO_2 images or tHb images individually. It indicates that both the measurements of oxygen saturation (to indicate hypoxia) and blood content (vascularity) can provide discrimination information, and combining them would give better performance for detecting RA. Although using other recoverable measurements such as water content and scattering should be subject of future studies, the experimental results also show that all the three DNN models outperform the traditional machine learning methods (SVM, DT and RFC). With the increase of the overlap in the distribution of StO_2 and tHb, the accuracy of the DNN models become higher than traditional machine learning methods which indicates that the DNN based models are more robust than the traditional machine learning methods. Lastly, transfer learning has been used to improve the classification accuracy with the results showing that transfer learning can improve classification accuracy, which should be used in future with clinical data to demonstrate direct applicability.

To conclude, the classification results of three different DNN models have been presented, demonstrating their capability for the diagnosis of inflamed and non-inflamed finger joints for the detection of rheumatoid arthritis. The analysis of the presented results indicate that DNN models are robust with direct applications in DOT even when the recovered optical maps show large overlap between two different functional states and are typically less differentiated for different classes. These results underscore the potential for DNN models to be used as a computer aided tool in DOT diagnostic systems, and warrants larger prospective trials to conclusively demonstrate the ultimate clinical utility of this approach.

Acknowledgements. This work is supported by the National Natural Science Foundation of China (Grant Nos 61772353 and 61332002), the Foundation for Youth Science and Technology Innovation Research Team of Sichuan Province (Grant No. 2016TD0018), the Fok Ying Tung Education Foundation (Grant No. 151068) and by EPSRC through a studentship from the Sci-Phy-4-Health Centre for Doctoral Training (EP/L016346/1).

References

1. Majithia V., Geraci S.A. *Am. J. Med.*, **120** (11), 936 (2007).
2. Helmick C.G., Felson D.T., Lawrence R.C., Gabriel S., Hirsch R., Kwoh C.K., Liang M.H., Kremers H.M., Mayes M.D., Merkel P.A., Pillemer S.R., Reveille J.D., Stone J.H., Workgrp N.A.D. *Arthritis Rheum.*, **58** (1), 15 (2008).
3. Nell V.P.K., Machold K.P., Eberl G., Stamm T.A., Uffmann M., Smolen J.S. *Rheumatology*, **43** (7), 906 (2004).
4. Landewe R.B.M., Boers M., Verhoeven A.C., Westhovens R., van de Laar M.A.F.J., Markusse H.M., van Denderen J.C., Westedt M.L., Peeters A.J., Dijkmans B.A.C., Jacobs P., Boonen A., van der Heijde D.M.F.M., van der Linden S. *Arthritis Rheum.*, **46** (2), 347 (2002).
5. Scheel A.K., Hermann K.G.A., Ohrndorf S., Werner C., Schirmer C., Detert J., Bollow M., Hamm B., Muller G.A., Burmester G.R., Backhaus M. *Ann. Rheum. Dis.*, **65** (5), 595 (2006).
6. Wakefield R.J., O'Connor P.J., Conaghan P.G., McGonagle D., Hensor E.M.A., Gibbon W.W., Brown C., Emery P. *Arthrit. Rheum. Arthr.*, **57** (7), 1158 (2007).
7. Klarlund M., Ostergaard M., Jensen K.E., Madsen J.L., Skjodt H., Lorenzen I., Grp T. *Ann. Rheum. Dis.*, **59** (7), 521 (2000).
8. Haavardsholm E.A., Boyesen P., Ostergaard M., Schildvold A., Kvien T.K. *Ann. Rheum. Dis.*, **67** (6), 794 (2008).
9. Durduran T., Choe R., Baker W.B., Yodh A.G. *Rep. Prog. Phys.*, **73** (7), 076701 (2010).
10. Tromberg B.J., Pogue B.W., Paulsen K.D., Yodh A.G., Boas D.A., Cerussi A.E. *Med. Phys.*, **35** (6), 2443 (2008).
11. Giacometti P., Diamond S.G. *Bioanalysis Adv. Mat.*, **3**, 57 (2013).
12. Hielscher A.H., Kim H.K., Montejo L.D., Blaschke S., Netz U.J., Zwaka P.A., Illing G., Muller G.A., Beuthan J. *IEEE Trans. Med. Imaging*, **30** (10), 1725 (2011).
13. Montejo L.D., Jia J.F., Kim H.K., Netz U.J., Blaschke S., Muller G.A., Hielscher A.H. *J. Biomed. Opt.*, **18** (7), 076001 (2013).
14. Montejo L.D., Jia J.F., Kim H.K., Netz U.J., Blaschke S., Muller G.A., Hielscher A.H. *J. Biomed. Opt.*, **18** (7), 076002 (2013).
15. Lighter D., Hughes J., Styles I., Filer A., Dehghani H. *Biomed. Opt. Express*, **9** (4), 1445 (2018).
16. Krizhevsky A., Sutskever I., Hinton G.E. *Commun. ACM*, **60** (6), 84 (2017).
17. Chen Y.Y., Zhang L., Yi Z. *Inform. Sci.*, **424**, 27 (2018).
18. Mohamed A.R., Dahl G.E., Hinton G. *IEEE Trans. Audio Speech*, **20** (1), 14 (2012).
19. Zhang L., Yi Z. *Chaos Soliton Fract.*, **33** (3), 979 (2007).
20. Zhang L., Yi Z. *IEEE Trans. Neural Networks*, **22** (7), 1021 (2011).
21. Zhang L., Yi Z., Amari S. *IEEE Trans. Neur. Net. Lear.*, **29** (11), 5242 (2018).
22. Zhang L., Yi Z., Yu J.L. *IEEE Trans. Neural Networks*, **19** (1), 158 (2008).
23. Wang L.T., Zhang L., Yi Z. *IEEE Trans Cybernetics*, **47** (10), 3172 (2017).
24. LeCun Y., Bengio Y., Hinton G. *Nature*, **521** (7553), 436 (2015).
25. Simonyan K., Zisserman A. arXiv preprint arXiv:1409.1556 (2014).
26. Szegedy C., Liu W., Jia Y.Q., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. *Proc. IEEE CVPR* (Boston, 2015) pp 1–9.
27. Spanhol F.A., Oliveira L.S., Petitjean C., Heutte L. *Proc. IEEE IJCNN* (Vancouver, 2016) pp 2560–2567.
28. Spanhol F.A., Cavalin P.R., Oliveira L.S., Petitjean C., Heutte L. *Proc. 2017 IEEE Int. Conf. Trans. Syst. Man. Cyb.* (Banff, 2017) p. 1868.
29. Tajbakhsh N., Shin J.Y., Gurudu S.R., Hurst R.T., Kendall C.B., Gotway M.B., Liang J.M. *IEEE Trans. Med. Imaging*, **35** (5), 1299 (2016).
30. Srivastava R.K., Greff K., Schmidhuber J. arXiv preprint arXiv:1505.00387 (2015).
31. He K.M., Zhang X.Y., Ren S.Q., Sun J. *Proc. IEEE CVPR* (Las Vegas, 2016) 770–778.
32. Lecun Y., Bottou L., Bengio Y., Haffner P. *Proc. IEEE*, **86** (11), 2278 (1998).
33. Nair V., Hinton G.E. *Proc. 27th Int. Conf. on Machine Learning (ICML-10)* (Haifa, 2010) pp 807–814.
34. Hornik K. *Neural Networks*, **4** (2), 251 (1991).

35. Quattoni A., Collins M., Darrell T. *Proc. IEEE CVPR* (Anchorage, 2008) pp 2300–2307.
36. Oquab M., Bottou L., Laptev I., Sivic J. *Proc. IEEE CVPR* (Columbus, 2014) pp 1717–1724.
37. Zhu Y., Chen Y., Lu Z., Pan S.J., Xue G.R., Yu Y., Yang Q. *Proc. 25th Conf. on Artificial Intelligence* (San Francisco, 2011) p. 4057.
38. Ahmed A., Yu K., Xu W., Gong Y.H., Xing E. *Lect. Notes Comput. Sci.*, **5304**, 69 (2008).
39. Cao X.D., Wipf D., Wen F., Duan G.Q., Sun J. *Proc. IEEE ICCV* (Sydney, 2013) pp 3208–3215.
40. Deng J., Zhang Z.X., Marchi E., Schuller B. *Proc. IEEE ACII* (Geneva, 2013) pp 511–516.
41. Huang J.T., Li J.Y., Yu D., Deng L., Gong Y.F. *Proc. IEEE ICA SSP* (Vancouver, 2013) pp 7304–7308.
42. Taylor M.E., Stone P. *J. Mach. Learn. Res.*, **10**, 1633 (2009).
43. Vapnik V.N., Vapnik V. *Statistical Learning Theory* (New York: Wiley, 1998).
44. Safavian S.R., Landgrebe D. *IEEE Trans. Syst. Man Cyb.*, **21** (3), 660 (1991).
45. Breiman L. *Mach. Learn.*, **45** (1), 5 (2001).
46. Dehghani H., Eames M.E., Yalavarthy P.K., Davis S.C., Srinivasan S., Carpenter C.M., Pogue B.W., Paulsen K.D. *Commun. Numer. Meth. Eng.*, **25** (6), 711 (2009).
47. Ng C.T., Biniecka M., Kennedy A., McCormick J., FitzGerald O., Bresnihan B., Buggy D., Taylor C.T., O'Sullivan J., Fearon U., Veale D.J. *Ann. Rheum. Dis.*, **69** (7), 1389 (2010).
48. Falchuk K.H., Goetzl E.J., Kulka J.P. *Am. J. Med.*, **49** (2), 223 (1970).
49. Ash H.E., Unsworth A. *Proc. Instn. Mech Engrs*, **211H** (5), 377 (1997).
50. Yuan Z., Zhang Q.Z., Sobel E.S., Jiang H.B. *Biomed. Opt. Express*, **1** (1), 74 (2010).
51. Klose A.D., Hielscher A.H. *Med. Phys.*, **26** (8), 1698 (1999).
52. Rowland R., Ponticorvo A., Baldado M., Kennedy G.T., Burmeister D.M., Christy R.J., Bernal N.P., Durkin A.J. *J. Biomed. Opt.*, **24** (5), 056007 (2019).
53. Panigrahi S., Gioux S. *J. Biomed. Opt.*, **24** (7), 071606 (2019).