Smart Health (2019)

Contents lists available at ScienceDirect

Smart Health

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

Multi-task Osteoporosis Pre-screening Using Dental Panoramic Radiographs with Feature Learning

Sijia Yu[a], Peng Chu[a], Jie Yang[b], Bingyao Huang[a], Fan Yang[a], Vasileios Megalooikonomou[c], Haibin Ling[a,*]

[a]*Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122, United States*
[b]*Division of Oral and Maxillofacial Radiology, School of Dentistry, Temple University, Philadelphia, PA 19122, USA*
[c]*Computer Engineering and Informatics Department, University of Patras, 26504, Patras, Greece*

ARTICLE INFO

Communicated by xxx

Osteoporosis
Octuple Siamese Network
Dental Panoramic Radiography

ABSTRACT

Known as the most common bone disease, osteoporosis is affecting millions of people every year. Clinical diagnostic methods for osteoporosis are expensive and therefore have limited availability in population. Recent studies have shown that Dental Panoramic Radiography (DPR) images can provide the bone density change clues in trabecular bone structure analysis. We collected 108 DPR images from 108 different subjects with 52 osteoporosis and 56 normal subjects, and aim to use the powerful deep neural network (DNN) on these DPR images for osteoporosis classification. However, DNN, such as AlexNet, requires millions of images for model learning. For many clinical related tasks, the available amount of data is far from adequate to train a DNN from scratch. One common solution is to use a pre-trained DNN and fine-tune it on the target dataset. Due to limited number of images and the big difference in image styles between ImageNet and our DPR dataset, direct fine-tuning of DNN can hardly achieve promising performance. In this paper, we propose a multi-task scheme to transfer the network pre-trained on ImageNet to our DPR dataset and improve the accuracy in osteoporosis pre-screening. Specifically, the AlexNet pre-trained on ImageNet is first fine-tuned on the task of (patient) identity classification using the abundant patches extracted from 108 DPR images. Features learned from such task are more dental-data friendly and are better for our main task, *i.e.*, osteoporosis pre-screening. For this purpose, the Octuplet Siamese Network (OSNet) is built on top of such features and trained on eight regions of interest (ROIs) in each DPR image for the osteoporosis classification task. Finally, the leave-one-out cross-validation shows that the highest overall accuracy of osteoporosis pre-screening is 92.6%, showing significant advantage over previous methods.

1. Introduction

Osteoporosis is a metabolic bone disease characterized by low bone mass and disruption in bone microarchitecture. It can lead to bone fragility and increase fracture risks Deng et al. (2018). According to National Osteoporosis Foundation's report, osteoporosis affects 55% of Americans over 50 years old and has more than 3 million US cases per year Wright et al. (2014). Osteoporosis patients have no symptoms until having a bone fracture, which may lead to significant morbidity and mortality Oleksik et al. (2000, 2005); Bliuc et al. (2009); Ensrud et al. (2000). Early diagnosis of osteoporosis plays an important role in reducing bone fracture risks and enhancing the quality of life for these patients. One of the most precise diagnostic method of osteoporosis in current clinical treatments is dual-energy x-ray absorptiometry (DXA), which determines patients'

*Corresponding author: Haibin Ling Tel.: +1-215-204-6973; fax: +1-215-204-5082;
e-mail: haibin.ling@gmail.com (Haibin Ling)

http://dx.doi.org/10.1016/j.smhl.xxxx.xx.xxx
Received 19 March 2019; Received in final form 05 July 2019; Accepted 16 June 2019; Available online xxx

osteoporosis conditions by measuring their bone mineral density (BMD) van Casteren-Messidoro et al. (2014). However, the DXA assessment equipment is not widely available and the cost of DXA is high Moyad (2003).

Panoramic radiography is a relatively cheap and more convenient diagnostic method for potential osteoporosis patients Taguchi et al. (2008). In fact, dental panoramic radiography (DPR) images are frequently taken by dentists in routine clinical dental examinations with no additional cost Taguchi (2010). In this case, studies have been conducted on discovering the diagnostic potential of DPR images on osteoporosis. Recently, increasing studies have shown that DPR images comprise trabecular bone patterns which can be used to predict undetected osteoporosis Geraets et al. (2007); Devlin et al. (2007); Cakur et al. (2008); Taguchi (2009); Božič & Hren (2013). Those works investigated multiple image features in these dental images, such as intensity and texture. However, in practice, those pre-defined low-level feature descriptors are vulnerable to various imaging imperfections. Failing to integrate the comprehensive image features also restricts the efficiency and effectiveness to discriminate desired trabecular bone patterns from other overlapping structures in DPR images White (2002); Devlin & Horner (2008). Recent deep learning methods, especially the convolutional neural network (CNN), achieve great success in learning task-specific features from image data. With the deep and hierarchical features, the performance in many medical image classification tasks are greatly improved. However, training a powerful deep neural network (DNN) requires sufficient labeled training samples which are usually hard to acquire in many medical tasks. A common solution is the fine-tuning scheme where a DNN is initialized by pre-training and task-specific data is used for minor adjustment on the pre-trained model. This strategy works promisingly when the images in pre-training and target tasks share enough low-level features and the amount of task data is sufficient to prevent over-fitting. However, neither of them are satisfied in our DPR dataset.

In our study, we propose a multitask scheme to train a pre-trained DNN for osteoporosis classification task using limited labeled DPR images. It is accomplished by introducing an auxiliary task to classify patients' IDs using their DPR images. The network learns features from sufficient patches extracted from DPR images in the auxiliary task and utilizes those features in the final osteoporosis classification task with limited labeled data. The whole paper is arranged as follow: Section 1. Introduction: In the introduction section, we briefly describe the background information of our study including shortcomings of current osteoporosis diagnostic methods and the potential of using features learnt from DPR images for osteoporosis classification. Section 2. Related Work: In this section, we introduce previous work on osteoporosis screening using DPR images. We also include a basic review of machine learning, CNN, multi-task network and their applications in medical research. Section 3. Method: In this section, we first describe the composition of our dataset. Then we present the design of the network and ordinary fine-tuning scheme, followed by the details of our multi-task scheme. Section 4. Experiments and results: In this section, we state the steps of our experiments and our results. Section 5. Conclusion and discussion: In this section, we discuss the contributions and future directions of our work.

2. Related Work

2.1. Overview of machine learning in medical research

Machine learning (ML) is a scientific study that utilizes mathematical algorithms to improve the performance of computer systems on predicting outcomes in target tasks. The basic process of machine learning is that the computer uses the sample data and statistical algorithms to build a characteristic mathematical model. This model can be further used to make predictions on cases which are not in the sample dataset for this task. As big data analysis growing, machine learning is closely related to many fields including data mining, optimization and statistics. Especially, machine learning provides several indispensable and relatively inexpensive tools for large data analysis in medical research and clinical diagnosis Kononenko (2001), including cell activity study Zhang et al. (2018), health monitoring system Chen et al. (2017), medical prognose Chan et al. (2002), missing data handling in medical research Quinlan (1989) and complementary medicine Zhao et al. (2015).

There are generally three types of machine learning algorithms: supervised learning, unsupervised learning and reinforcement learning. In supervised learning, the training data have both inputs and desired outputs, known as labels. A supervised algorithm is first trained using given dataset to make predictions as the given labels. After the training step, the optimized supervised learning algorithm should have the capability to provide reliable predictions on new inputs Mohri et al. (2018). One of the most common supervised learning tasks is classification. In a classification task, there are limited numbers or sets of outputs, known as classes. Each input is classified by the supervised learning algorithm model into one class. In our study, we have two output classes, which are osteoporosis and normal. Our aim is to develop a network scheme that uses DPR images as the input and outputs a proper prediction of the osteoporosis condition shown in each DPR image.

2.2. CNN and AlexNet in medical image analysis

With the booming of health-care and medical research, the accumulation of medical records offers a promising source for big data analysis. Highly efficient techniques such as machine learning and artificial intelligence have been applied on medical data analysis in many aspects, including medical image processing and computer-aided diagnosis Razzak et al. (2018). These techniques are built on conventional and deep learning algorithms. Among them, one of the most popular algorithms is CNN. Different types of CNNs are Alexnet Krizhevsky et al. (2012), ResNet He et al. (2016), VGGNet Simonyan & Zisserman (2014), ZFNet Zeiler & Fergus (2014) and so on. Due to its strong learning ability and good performance, CNN has caught lots of attention in digital imaging processing and computer vision.

Gulshan et al. (2016) proposed a deep CNN algorithm for diabetic retinopathy and diabetic macular edema detection in retinal images. They applied CNN on Eye Picture Archive Communication System (EyePACS-1) dataset and Messidor-2 dataset which contain 9963 and 1748 images respectively. They achieved 97.5% sensitivity and 93.4% specificity on EyePACS-1 dataset and 96.1% sensitivity and 93.9% specificity on Messidor-1 dataset. CNN is also used in histological analysis. In Bayramoglu & Heikkilä (2016), the author used CNN transfer learning and fine-tuning for cell nuclei classification. Quinn et al. (2016)employed CNN on shape feature learning for malaria diagnosis with samples from blood smears, sputum and stool. All three cases shown high classification accuracy and better performance than traditional medical imaging techniques. Spanhol et al. (2016) proposed an AlexNet variant for breast cancer image classification using histopathological images from BreaKHis dataset. They generated patches from these images for network training and developed strategies to handle high-resolution textured images. Their results

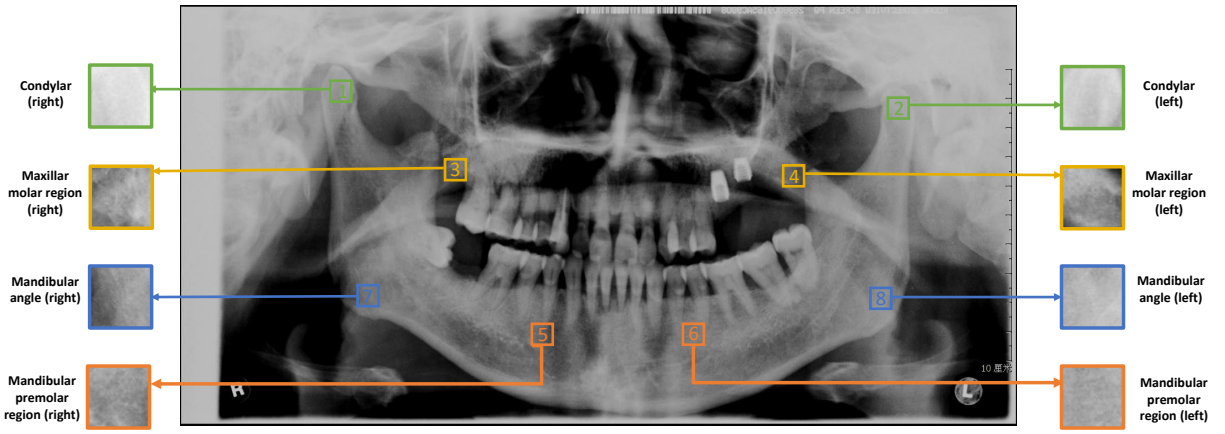

Fig. 1: DPR image with eight ROIs. Eight ROIs are separated in 4 groups according to symmetry. 1 and 2: the condylar (green); 3 and 4: the maxillar molar region (yellow); 5 and 6: the mandibular premolar region (orange); 7 and 8: the mandibular angle (blue).

shown improved accuracy using CNN than previous machine learning models on the same dataset. Besides, CNN has a variety of applications in medical data analysis including cardiac imaging analysis, tumor detection and gastrointestinal diseases diagnosis Razzak et al. (2018).

2.3. DPR in osteoporosis diagnosis

It is known that variations in trabecular bone textures are closely related to changes of bone density. Therefore, trabecular bone structures have the potential to assist bone disease analysis including osteoporosis diagnosis. According to recent medical research, dental radiography can provide valuable information in trabecular bone structures for osteoporosis pre-screening studies Kavitha et al. (2012); Roberts et al. (2013); Ling et al. (2014); Li et al. (2014). Kavitha et al. (2012) developed a kernel-based support vector machine learning method to identify women over 50 years old with osteoporosis or low bone mineral density. Their method discriminated osteoporosis from normal subjects using the difference in cortical width of the mandible on DPR images. Their results suggested that DPR could be a useful source to get bone density information in computer-aided osteoporosis diagnosis. Moreover, in Roberts et al. (2013), the authors reported that using additional texture features from DPR images in osteoporosis classification led to highly correlated results with clinical diagnostic records. In this case, the authors suggested that image features in DPR could be useful in bone quality measurement and thus could serve as a biomarker for osteoporosis diagnosis. Li et al. (2014) proposed a multi-feature approach to analyze trabecular patterns from CBCT volumes. They extracted fractal dimension, multi-fractal spectrum and gradient based features from regions of interest (ROIs) and used multi-kernel learning to fuse these features for classification. Their results provided evidences that using multiple features and ROIs could benefit the outcome in computer-aided osteoporosis classification tasks. Ling et al. (2014) investigated the correlation between multiple trabecular texture features and patients' gender and age using dental cone beam CT (CBCT) images. They showed that features in these dental CBCT images, especially those in the mandible, were statistically significantly related to the changes of trabecular patterns. Therefore, they concluded that dental CBCT images had the potential for trabecular structure analysis.

However, the accuracy and reliability of using DPR to determine osteoporosis conditions is dissatisfactory. Bo et al. (2017) constructed a two-stage classification framework using different support vector machines (SVMs) for osteoporosis classification task. In the training process, each SVM in the first stage handled a given feature and a group of ROIs. The probability outputs from the first stage were combined by an additional liner SVM in the second stage for classification. Their best results with HOG features reached 72.5% in overall accuracy with p-value 0.0164. Chu et al. (2018) proposed to use Octuplet Siamese Network (OSNet) to classify patients on their osteoporosis status using DPR images. In their work, the network was trained with augmented texture dataset and fine-tuned on target DPR images. Due to differences between the images in augmented texture dataset and DPR images, the best accuracy achieved was 89.8%. Therefore, in this work, we propose a multitask scheme to enhance the feature learning process and finally improve the classification accuracy of patients' osteoporosis status using OSNet.

2.4. Multi-task network in medical image analysis

Typically, in a machine learning or deep learning task, we focus on training our single model and fine-tuning the model to its best performance. However, it might be the case that we overlook useful information that could be acquired from some auxiliary tasks. Multi-task learning is a strategy that further optimizes a model's performance by introducing information of the training signals from related tasks Ruder (2017). According to previous studies, multi-task learning can benefit the model training in many aspects including data augmentation, attention focusing and balancing the representation bias.

Due to these advantages, multi-task learning has been used successfully in many machine learning tasks, from natural language processing Collobert & Weston (2008) to computer vision Girshick (2015) and drug discovery Ramsundar et al. (2015). Hunt et al. (2018) used multi-task learning with LASSO and graph regularization to analyze plug-in estimators for missing data handling in healthcare records. They tested their model on Alzheimer's disease progression dataset and showed effectiveness for disease prediction when missing data was present. Ramsundar et al. (2015) dug deeper in the advantages of multi-task neural architectures in drug discovery. They mentioned that, comparing with single-task methods, multi-task networks could improve predictive accuracies and this predictive power was strengthened when additional data and tasks were

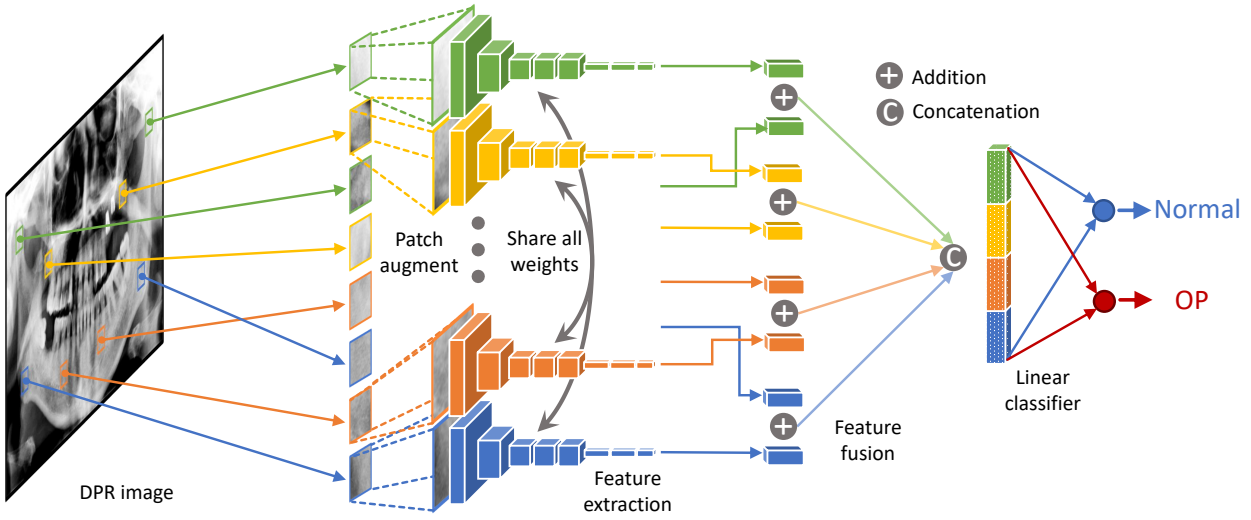

Fig. 2: Overview of Octuplet Siamese Network (OSNet) Chu et al. (2018). OSNet has eight identical subnets and each of them is an AlexNet. OSNet has three parts: deep feature extraction, feature fusion and linear classifier.

involved. Also, features learnt by multi-task networks showed some extent of transferability to tasks that were not trained with the network before. These studies provide evidences that multi-task learning can benefit tasks which have limitations in data and require more powerful network learning abilities.

3. Methods

3.1. Dataset

Our dataset consists of 108 DPR images from 108 different subjects, among which 52 are diagnosed as osteoporosis and 56 are normal. The subjects are classified according to the World Health Organization as normal (T-score ≥ -1.0) or osteoporotic (T-score < -1.0). In practice, due to the abundance and homogeneity of trabecular bone structure distribution in oral cavity, we investigate multiple ROIs in the mandible and the maxilla area for robustness. Specifically, eight ROIs are manually picked by dentists in each DPR. As shown in Figure 1, the locations of ROIs are in the condylar (green), angle (blue) and premolar region (orange) of the mandible, and the maxillar molar region (yellow). The 8 ROIs in each DPR image are divided into 4 groups according to symmetry: Group 1: ROIs 1, 2; Group 2: ROIs 3, 4; Group 3: ROIs 5, 6; and Group 4: ROIs 7, 8. Each ROI is normalized into 50 x 50 pixels and ROIs in the same group are labeled with the same color. We choose these four sets of ROIs because the trabecular bone structure in these four regions are less likely to be affected by other dental diseases and are more closely related to bone density of the observational subject. Thus, these eight ROIs are more effective and accurate in indicating the osteoporosis level of the subject.

3.2. Octuple Siamese Network

In order to apply powerful CNN on a dataset with limited DPR images on osteoporosis classification, Chu et al. (2018) designed an efficient OSNet for this binary classification task, as shown in Figure 2. The OSNet mainly consists of three parts: deep feature extraction, feature fusion and linear classifier. In the feature extraction part, Siamese-style network with eight identical subnets (AlexNet) are used and each subnet corresponds to one of the ROI inputs in a DPR image. While from different locations, the eight input patches are extracted from the same DPR image so they share the same image features. Since the network training is accomplished by updating weights in each layer of the neural network based on information learnt from input features, the eight subnets are weights sharing. In this way, features in each ROI can be used to update weights in all the eight subnets which is equivalent to increasing the training samples by seven times. The feature fusion part is built on the fact that human face is symmetric. In this part, features from the same ROI group are added together while features from different ROI groups are concatenated. These schemes ensure that features from different parts of a DPR image are passed independently into the classifier where decisions can be made in parallel. At the same time, this design also keeps the feature compact so that a simple classifier can be adapted afterwards and it further reduces the complexity for training. Finally, a single fully-connected layer with two output neurons is used to generate the classification decision of each input DPR image. The two outputs indicate the probability of an input DPR image being osteoporosis or normal.

3.3. Fine-Tuning

Training a powerful CNN requires a huge dataset, such as ImageNet Deng et al. (2009), which contains one million labeled images for training. However, available training data in ordinary tasks is usually very limited and is not enough to train deep CNNs, such as AlexNet, from scratch. One practical method to handle this situation is fine-tuning. The procedure of fine-tuning usually involves two steps. Firstly, the deep CNN is trained on some large and comprehensive databases such as ImageNet, where the CNN can be trained on enough data with very limited levels of overfitting.

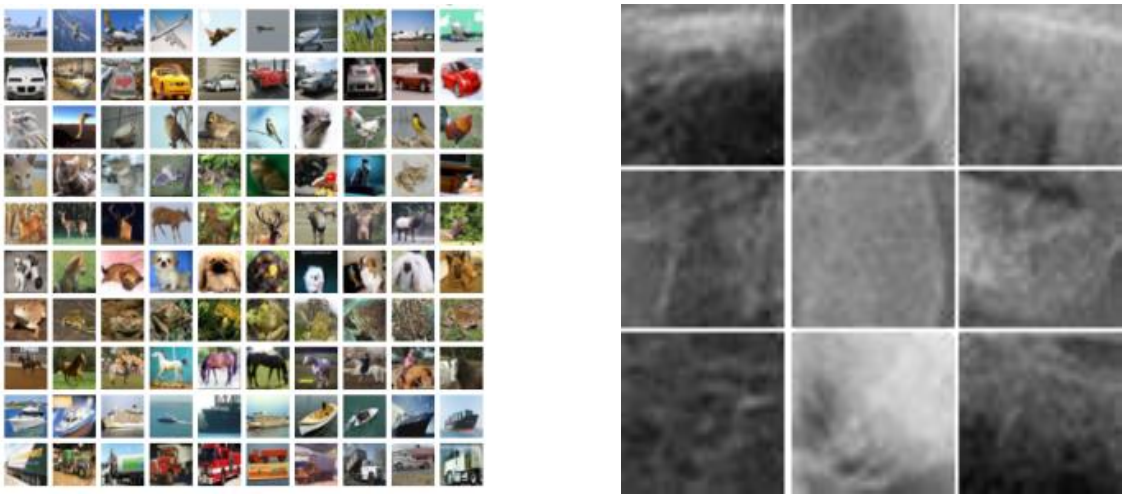

Fig. 3: ImageNet Deng et al. (2009) Images(left) *vs* DPR Images(right).

After that, the CNN obtains powerful feature representation abilities and can serve as a pre-trained model for general image learning tasks. In the second step, the task-specific dataset, which is usually much smaller than ImageNet, is used to further train the pre-trained model. This training is usually performed with very small learning rates, such as tenth or even less of the original learning rate in the initial network settings. In this way, the pre-trained model is making only slight modifications to fit the target dataset. By these two steps, a task-specified CNN can be obtained even when the task has a relatively small number of training data. The reason that fine-tuning works is that the feature representation in CNN is hierarchical. While images from different domains have different appearances, they all share the same low-level graphic characteristics. Through the classification process, these low-level features contribute to the domain specific clues and these clues, known as high-level features, are used for object classification. This feature representation peculiarity in CNN makes it possible for us to apply the feature extracting network which is trained in one domain on other domains. Moreover, pre-training on big datasets helps the network loss get close to global minimum. When this network is further applied on the target task, it is easier to converge with limited training data. Therefore, fine-tuning is an effective and efficient strategy in DNN training.

In many medical image classification tasks where labeled data is hard to collect, the number of data required for fine-tuning is still hard to reach. In this case, fine-tuning with insufficient data cannot improve the feature representation in the target domain. Moreover, medical images share less low-level features with natural images, especially for those captured by X-ray or in MRI. All these factors attribute to the need of more domain data for fine-tuning the CNN pre-trained on natural images. To meet this requirement, we introduce an auxiliary task to improve the network performance.

3.4. Multi-task Scheme

In order to improve the efficiency and accuracy of the network in our proposed osteoporosis classification task, we need to pre-train it before applying on the final classification task. However, we have two main challenges.

In the first step of training, we use the network trained on ImageNet as our pre-trained model. The ImageNet dataset is a large image database which contains more than 14 million images from over 20,000 classes. About one million images in ImageNet have bounding box annotations which makes it a useful resource for network pre-training in a variety of deep learning tasks including object localization and classification. However, in our task, the images for classification are DPR images which contain trabecular bone structures. As shown in Figure 3, patterns in ImageNet are mostly from colorful natural objects whereas patterns in our dental images are porous, spongy or honeycombed. Due to the huge difference in image styles between ImageNet and DPR images, the model pre-trained on ImageNet cannot offer qualified performance on our classification task. Thus, we need to further fine-tune our model on images that have features close to those in DPR images for better classification performance.

Secondly, AlexNet pre-trained on ImageNet is powerful and needs lots of training samples to accommodate it from ImageNet to dental images. Insufficient data can easily lead to overfitting and losing generalizability, which severely degrades the performance of network on testing data. However, we have only 108 images and each of the image has eight ROIs. Even though we can make the most use of our DPR images by leave-one-out training strategy, the amount of data is still insufficient to accomplish this task. Thus, we need more data to ensure that the network can learn the features in DPR images well. Chu et al. (2018) used an augmented texture dataset to facilitate the transition from natural images to trabecular bone radiographs. In this dataset, there are images showing natural or artificial textures. Using an intermediate dataset of different

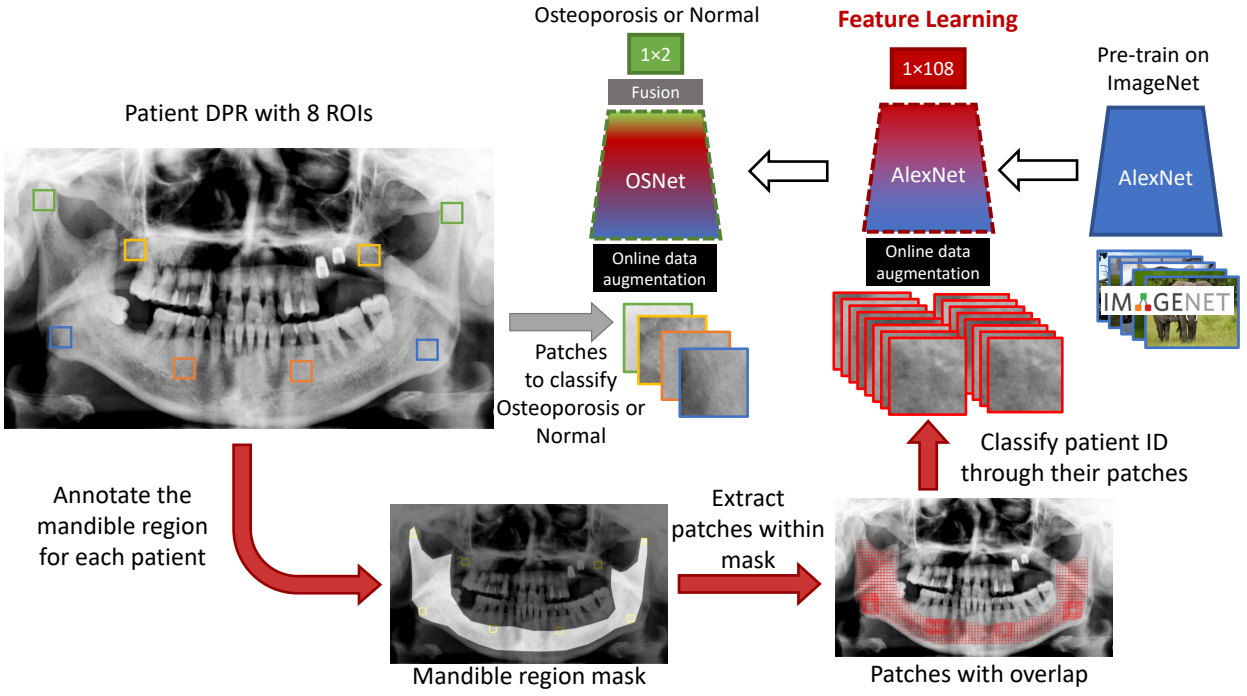

Fig. 4: Process flow of multitask osteoporosis classification.

domains may introduce overfitting on other irrelevant features and lose the discriminability on potential useful features. Considering this issue, we believe that learning features using patches directly extracted from DPR images should be the optimal solution to further enhance the performance of classification. Thus, we propose a multi-task feature learning setting to include DPR feature learning in our classification task.

Instead of directly fine-tuning AlexNet (the backbone network of OSNet) on 108 patients' osteoporosis status, we use the 108 patients' IDs as the training label set to enhance feature learning for our model. In detail, multiple patches are extracted from each DPR image. Those patches are used to train AlexNet to classify which patient a patch comes from. Since the patches are directly derived from DPR images, they share the same low-level features with those in eight ROIs. Through the 108-classes classification task, AlexNet gets adapted on DPR images through the feature learning and is more sensitive on the image details that discriminates each DPR patch. We use cross-entropy loss for training in this step and the loss can be calculated as:

$$Loss = \sum_{c=1}^{M} y_{o,c} \log(p_{o,c}), \tag{1}$$

where:

- M - number of classes (in our case: 108)
- log - the natural log
- $y_{o,c}$ - binary indicator (0 or 1) if class label c is the correct classification for observation o
- $p_{o,c}$ - predicted probability observation o is of class c

4. Experiments and results

4.1. Process flow

The whole task consists of three parts: pre-training, feature learning and fine-tuning for osteoporosis classification. As shown in Figure 4, we firstly use AlexNet pretrained on ImageNet as the initialization. Then, in the feature learning step, the pretrained AlexNet is trained with patient ID classification task using patches generated from DPR images. In this step, sufficient DPR patches are extracted from the mandible region mask and used for the network to learn the low-level features in DPR images. In the final step, the AlexNet after feature learning is served as the backbone network of OSNet. The OSNet is trained for the osteoporosis status classification task using patches in the eight ROIs.

4.2. Feature learning

To improve the performance of our network, we need to generate more data with features in DPR images and train our network on those data before doing the osteoporosis classification. To generate patches from each DPR image for feature learning, we firstly draw a polygon mask

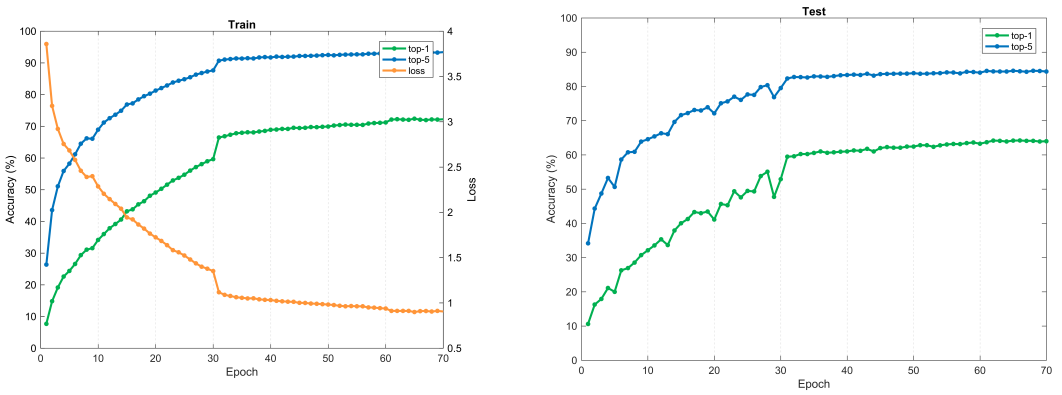

Fig. 5: Learning curves from the training (left) and testing (right) processes are shown. In detail, left figure shows training loss (orange), top-1 accuracy (green) and top-5 accuracy (blue) in the training and right figure shows top-1 accuracy (green) and top-5 accuracy (blue) in the testing. The shift the 30^{th} epoch accords with the first-time learning rate change.

Table 1: Accuracy of osteoporosis classification with different number of epochs on feature learning.

# of Epoch	0	1	3	4	5	6	7	8	9	10
Accuracy of classification (%)	87.96	87.96	92.59	90.74	92.59	90.74	91.66	87.96	87.04	87.04

covering the trabecular bone structure in mandible region, as shown in Figure 4 bottom left. We do not expend the mask into the maxilla area because there is only limited trabecular bone structure shown in that region. Each of the mask starts from the left condyle and goes down to cover the whole mandible area all the way to the right condyle. All six ROIs in the mandible region are included in the mask and the rest area in the mask share the same bone texture pattern with the six ROIs. Patches are extracted by using sliding-widow across the whole DPR image with a bounding box starting from the left top corner of each image. Since DPRs are taken with different scaling, the size of the bounding box is normalized according to DPI (Dots Per Inch) of each image and resized into 100×100 pixels. A patch is accepted if its overlapping with the mask is larger than 90%. DPR patches are generated with step length of 25 pixels and for patches whose distances to any of the six ROIs are smaller than 50 pixels, the steps are reduced to 10 pixels. The distance is calculated between the center of the bounding box and the center of a ROI. All these steps and distances are also normalized according to DPI. As a result, we extract 69,850 patches from 108 DPR images in total and on average each class (patient ID) contains about 650 patches.

Then, we use these patches to train AlexNet for patient identification. Patches are randomly split into 10 folds, with equal distribution in each patient class. We take 9 folds for training and 1 fold for testing. We also use online data augmentation methods including random crop and horizontal flip to reduce overfitting. In order to adapt AlexNet for ID classification task, we make some adjustments on network configuration. Firstly, since AlexNet has 1000 output classes by default, we add an extra layer to fit 1000 outputs into 108 classes. In this layer, there are 108 nodes and each node corresponds to one patient. After a patch is processed through the network, each node in this layer can give a probability if an input patch belongs to one of the 108 classes. Based on those probabilities, the output layer gives a prediction which patient class should this patch come from. Here, we pick both a tight and a loose way to evaluate the classification accuracy of the network and they are called the top-1 and top-5 accuracy respectively. The top-1 accuracy is calculated by picking the node with the highest probability and if the node's class is the same as the ground truth of this patch, we consider the prediction is true (mark as 1), otherwise false (mark as 0). For top-5 accuracy, the top 5 nodes with the highest probabilities are picked. If one of them is the ground truth of the patch, the prediction is 1 and if the ground truth is not in these five classes, then the prediction is 0. Also, to meet the default setting of input in AlexNet, we resize patches into 224×224 pixels.

The learning curve of the first 70 epochs is shown in the left in Figure 5. The shift in curve at the 30^{th} and 60^{th} epoch correspond to the change in learning rate. The starting learning rate is 0.001. Every time the training finishes 30 epochs, the learning rate is decreased to one tenth. We measure both the top-1 and top-5 prediction on every patch and the average accuracy of all patients in each epoch is shown in the figure. The highest top-1 accuracy on training data is around 70% and the highest accuracy on testing data is 60%.

4.3. Osteoporosis classification

After feature learning, we use the network as the feature extraction sub-network of OSNet and apply it on osteoporosis classification. The network trained in the feature learning part with different numbers of epochs is used to train the osteoporosis classification task. Their accuracies are shown in Tabel 1. The highest accuracy is around 92.6% with 3 or 5 epochs of feature learning. We also compare our results with other methods on the same task, shown in Tabel 2. We report the accuracy on non-osteoporosis subjects (NOPA), osteoporosis subjects (OPA) and the overall accuracy (OA). Our model increases 2.7% in OA and remains top in OPA and NOPA compared with the best previous model.

Table 2: Accuracy of osteoporosis classification compared with other methods.

Methods	TSL[5]	SFTA[6]	OSN[6]	proposed
NOPA(%)	67.86	71.43	89.29	89.29
OPA(%)	67.31	82.69	90.38	96.15
OA(%)	67.59	76.85	89.81	92.59

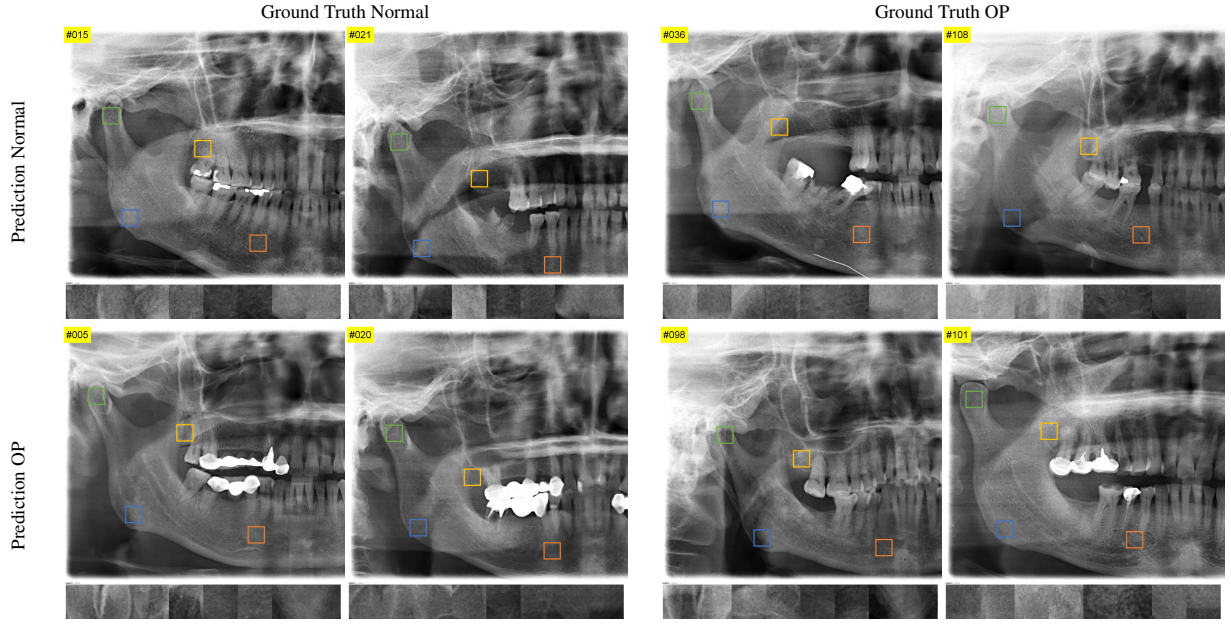

Fig. 6: Visualization of selected results. Each cell shows two sample cases with their left faces. The eights patches in the ROIs are visualized at the bottom of each case. Top left and bottom right are correctly classified cases and the rest two are misclassified cases. The T-score of misclassified cases are -3.45 (#036), -1.25 (#108), -0.98 (#05), -0.37 (#20).

5. Conclusion and Discussion

In this project, we present a study that uses DPR images for osteoporosis condition evaluation and propose an image-based osteoporosis pre-screening method. To improve the classification accuracy, we conduct a multi-task network scheme with feature learning to enhance the network performance in osteoporosis classification. Thanks to the powerful features learned in the multi-task scheme, we improve classification accuracy to 92.59%. In this work, we firstly use the ImageNet dataset for pretraining on AlexNet. Then we extract patches from the mandible region in DPR images for feature learning and use patches in eight ROIs from each DPR image for osteoporosis classification. This multi-step training scheme can be adopted in clinical osteoporosis condition analysis using DPR images as a fast and easy diagnostic method.

For further understanding the behavior of the proposed method, some misclassified samples are shown in Figure 6 (top right and bottom left) from our results. Compared with correctly classified samples (top left and bottom right in Figure 6), we can see that these are visually very hard cases, and our network may not be sensitive enough for handling them. Potential future directions along the study include 1) Include more areas in the DPR image, such as the maxilla region. 2) Use more data for osteoporosis classification. This can be done by either gather more DPR images or use patches extracted in feature learning part for osteoporosis classification. 3) Take other related factors into consideration. For example, since one very correlated indicator of osteoporosis is tooth loss, we can potentially include the bone structure in the tooth region for training. Moreover, we need to not only include more relative factors for learning but also look at the correlations between different factors to further increase the model's sensitivity to these factors. While there is much more future work to be done, our tentative work shows improvements in osteoporosis classification accuracy even when the data is scarce and hopefully can inspire more efficient models in the future.

Overall, the improvement shown in classification accuracy in this work encourages future work using DPR image for inexpensive osteoporosis pre-screening. This method also indicates that using effective training strategies and carefully designed feature augmentation can help improve network performance on other screening tasks.

Acknowledgments

This work was supported in part by US National Science Foundation (No. 1407156 and No. 1350521).

References

Bayramoglu, N., & Heikkilä, J. (2016). Transfer learning for cell nuclei classification in histopathology images. In *European Conference on Computer Vision* (pp. 532–539). Springer.

Bliuc, D., Nguyen, N. D., Milch, V. E., Nguyen, T. V., & Eisman, J. A. (2009). Mortality risk associated with low-trauma osteoporotic fracture and subsequent fracture in men and women. *Jama*, *301*, 513–521.

Bo, C., Liang, X., Chu, P., Xu, J., Wang, D., Yang, J., Megalooikonomou, V., & Ling, H. (2017). Osteoporosis prescreening using dental panoramic radiographs feature analysis. In *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on* (pp. 188–191). IEEE.

Božič, M., & Hren, N. I. (2013). A novel method of dental panoramic tomogram analysis: a perspective tool for a screening test for osteoporosis. *Journal of Cranio-Maxillofacial Surgery*, *41*, 808–815.

Cakur, B., Şahin, A., Dagistan, S., Altun, O., Caglayan, F., Miloglu, Ö., & Harorli, A. (2008). Dental panoramic radiography in the diagnosis of osteoporosis. *Journal of International Medical Research*, *36*, 792–799.

van Casteren-Messidoro, C., Huisman, A., Birnie, E., van Gelder, M., van de Geijn, F., & Hamberg, P. (2014). Quantitative ultrasound of the heel as triage test to measure bone mineral density compared with dual energy x-ray absorptiometry in men with prostate cancer commencing with androgen deprivation therapy. *Combination therapy of GLP-1 analogues and insulin: do the benefits outweigh the costs?*, (p. 528).

Chan, K., Lee, T.-W., Sample, P. A., Goldbaum, M. H., Weinreb, R. N., & Sejnowski, T. J. (2002). Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Transactions on Biomedical Engineering*, *49*, 963–974.

Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*, *5*, 8869–8879.

Chu, P., Bo, C., Liang, X., Yang, J., Megalooikonomou, V., Yang, F., Huang, B., Li, X., & Ling, H. (2018). Using octuplet siamese network for osteoporosis analysis on dental panoramic radiographs. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 2579–2582). IEEE.

Collobert, R., & Weston, J. (2008). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning* (pp. 160–167). ACM.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.

Deng, J., Feng, Z., Li, Y., Pan, T., Li, Q., & Zhao, C. (2018). Efficacy and safety of recombinant human parathyroid hormone (1–34) are similar to those of alendronate in the treatment of postmenopausal osteoporosis. *Medicine*, *97*, e13341.

Devlin, H., & Horner, K. (2008). Diagnosis of osteoporosis in oral health care. *Journal of oral rehabilitation*, *35*, 152–157.

Devlin, H., Karayianni, K., Mitsea, A., Jacobs, R., Lindh, C., van der Stelt, P., Marjanovic, E., Adams, J., Pavitt, S., & Horner, K. (2007). Diagnosing osteoporosis by using dental panoramic radiographs: the osteodent project. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, *104*, 821–828.

Ensrud, K. E., Thompson, D. E., Cauley, J. A., Nevitt, M. C., Kado, D. M., Hochberg, M. C., Santora, A. C., & Black, D. M. (2000). Prevalent vertebral deformities predict mortality and hospitalization in older women with low bone mass. *Journal of the American Geriatrics Society*, *48*, 241–249.

Geraets, W., Verheij, J. G., van der Stelt, P. F., Horner, K., Lindh, C., Nicopoulou-Karayianni, K., Jacobs, R., Harrison, E., Adams, J., & Devlin, H. (2007). Prediction of bone mineral density with dental radiographs. *Bone*, *40*, 1217–1221.

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440–1448).

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J. et al. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, *316*, 2402–2410.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hunt, X. J., Emrani, S., Kabul, I. K., & Silva, J. (2018). Multi-task learning with incomplete data for healthcare. *arXiv preprint arXiv:1807.02442*, .

Kavitha, M. S., Asano, A., Taguchi, A., Kurita, T., & Sanada, M. (2012). Diagnosis of osteoporosis from dental panoramic radiographs using the support vector machine method in a computer-aided system. *BMC medical imaging*, *12*, 1.

Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, *23*, 89–109.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Li, P., Yang, X., Xie, F., Yang, J., Cheng, E., Megalooikonomou, V., Xu, Y., & Ling, H. (2014). Trabecular texture analysis in dental cbct by multi-roi multi-feature fusion. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)* (pp. 846–859). IEEE.

Ling, H., Yang, X., Li, P., Megalooikonomou, V., Xu, Y., & Yang, J. (2014). Cross gender–age trabecular texture analysis in cone beam ct. *Dentomaxillofacial Radiology*, *43*, 20130324.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. MIT press.

Moyad, M. A. (2003). Osteoporosis: a rapid review of risk factors and screening methods. In *Urologic Oncology: Seminars and Original Investigations* (pp. 375–379). Elsevier volume 21.

Oleksik, A., Lips, P., Dawson, A., Minshall, M. E., Shen, W., Cooper, C., & Kanis, J. (2000). Health-related quality of life in postmenopausal women with low bmd with or without prevalent vertebral fractures. *Journal of bone and mineral research*, *15*, 1384–1392.

Oleksik, A. M., Ewing, S., Shen, W., van Schoor, N. M., & Lips, P. (2005). Impact of incident vertebral fractures on health related quality of life (hrqol) in postmenopausal women with prevalent vertebral fractures. *Osteoporosis international*, *16*, 861–870.

Quinlan, J. R. (1989). Unknown attribute values in induction. In *Proceedings of the sixth international workshop on Machine learning* (pp. 164–168). Elsevier.

Quinn, J. A., Nakasi, R., Mugagga, P. K., Byanyima, P., Lubega, W., & Andama, A. (2016). Deep convolutional neural networks for microscopy-based point of care diagnostics. In *Machine Learning for Healthcare Conference* (pp. 271–281).

Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., & Pande, V. (2015). Massively multitask networks for drug discovery. *arXiv preprint arXiv:1502.02072*, .

Razzak, M. I., Naz, S., & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. In *Classification in BioApps* (pp. 323–350). Springer.

Roberts, M. G., Graham, J., & Devlin, H. (2013). Image texture in dental panoramic radiographs as a potential biomarker of osteoporosis. *IEEE Trans. Biomed. Engineering*, *60*, 2384–2392.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*, .

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, .

Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). Breast cancer histopathological image classification using convolutional neural networks. In *2016 International Joint Conference on Neural Networks (IJCNN)* (pp. 2560–2567). IEEE.

Taguchi, A. (2009). Panoramic radiographs for identifying individuals with undetected osteoporosis. *Japanese Dental Science Review*, *45*, 109–120.

Taguchi, A. (2010). Triage screening for osteoporosis in dental clinics using panoramic radiographs. *Oral diseases*, *16*, 316–327.

Taguchi, A., Asano, A., Ohtsuka, M., Nakamoto, T., Suei, Y., Tsuda, M., Kudo, Y., Inagaki, K., Noguchi, T., Tanimoto, K. et al. (2008). Observer performance in diagnosing osteoporosis by dental panoramic radiographs: results from the osteoporosis screening project in dentistry (ospd). *Bone*, *43*, 209–213.

White, S. (2002). Oral radiographic predictors of osteoporosis. *Dentomaxillofacial radiology*, *31*, 84–92.

Wright, N. C., Looker, A. C., Saag, K. G., Curtis, J. R., Delzell, E. S., Randall, S., & Dawson-Hughes, B. (2014). The recent prevalence of osteoporosis and low bone mass in the united states based on bone mineral density at the femoral neck or lumbar spine. *Journal of Bone and Mineral Research*, *29*, 2520–2526.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818–833). Springer.

Zhang, X., Foderaro, G., Henriquez, C., & Ferrari, S. (2018). A scalable weight-free learning algorithm for regulatory control of cell activity in spiking neuronal networks. *International journal of neural systems*, *28*, 1750015.

Zhao, C., Li, G.-Z., Wang, C., & Niu, J. (2015). Advances in patient classification for traditional chinese medicine: a machine learning perspective. *Evidence-Based Complementary and Alternative Medicine*, *2015*.