Prediction of Osteoporosis Through Deep Learning Algorithms on Panoramic Radiographs

Melek Tassoker (iałehkmelek@gmail.com)  
Necmettin Erbakan University

Muhammet Usame Ozic  
Necmettin Erbakan University

Fatma Yuce  
Necmettin Erbakan University

Research Article

Keywords: Artificial intelligence, Osteoporosis, Panoramic Radiography, Deep Learning, Transfer Learning

Posted Date: January 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1245777/v1

License: ☑️️ This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Objective: The aim of the present study was to predict osteoporosis on panoramic radiographs of women over 50 years of age through deep learning algorithms.

Method: Panoramic radiographs of 744 female patients over 50 years of age were labeled as C1, C2, and C3 depending on mandibular cortical index (MCI). According to this index; C1: presence of a smooth and sharp mandibular cortex (normal); C2: resorption cavities at endosteal margin and 1 to 3-layer stratification (osteopenia); C3: completely porotic cortex (osteoporosis). The data of the present study were reviewed in different categories including C1-C2-C3, C1-C2, C1-C3 and C1-(C2+C3) as two-class and three-class prediction. The data were separated as 20% random test data; and the remaining data were used for training and validation with 5-fold cross-validation. AlexNET, GoogleNET, ResNET-50, SqueezeNET, and ShuffleNET deep learning models are trained through the transfer learning method. The results were evaluated by performance criteria including accuracy, sensitivity, specificity, F1-score, AUC and training duration.

Findings: The dataset C1-C2-C3 has an accuracy rate of 81.14% with AlexNET; the dataset C1-C2 has an accuracy rate of 88.94% with GoogleNET; the dataset C1-C3 has an accuracy rate of 98.56% with AlexNET; and the dataset C1-(C2+C3) has an accuracy rate of 92.79% with GoogleNET.

Conclusion: The highest accuracy was obtained in differentiation of C3 and C1 where osseous structure characteristics change significantly. Since the C2 score represent the intermediate stage (osteopenia), structural characteristics of the bone present behaviors closer to C1 and C3 scores. Therefore, the data set including the C2 score provided relatively lower accuracy results.

1. Introduction

Osteoporosis is a bone disease that develops when bone mineral density (BMD) and bone mass decrease or the quality or structure of the bone changes. Osteoporosis results in an increased risk of fractures in the vertebrae, hip and forearm bones along with decreased strength and increased fragility of the bone. Since hip and spine fractures can develop without symptoms, patients do not seek treatment and their mortality is higher. Therefore, it is vital to diagnose this group of patients who may be asymptomatic. Early diagnosis of this disease which causes pathological fractures especially in post-menopausal women over 50 years of age is very important.

Osteoporosis may be diagnosed through imaging techniques, biochemical markers and bone biopsy. The most practical diagnosis is established by imaging methods. The dual x-ray absorptiometry (DXA) is the gold standard for BMD measurement, and single photon absorptiometry, dual photon absorptiometry (DPA), quantitative computed tomography (QCT), high resolution magnetic resonance imaging (MRI) may also be used; however, the World Health Organization bases on BMD values measured by DXA for the diagnosis of osteoporosis. Accordingly, the T score above -1 is defined as normal whereas a T score between -1 and -2.5 is defined as osteopenia, and a T score below -2.5 is defined as osteoporosis.

Although DXA is a widely accepted diagnostic method, it is not a triage method in the diagnosis of osteoporosis. The frequent use of panoramic and intraoral radiographs directed researchers to focus on detecting lower BMD with these methods due to their cost-efficiency and usefulness. The mandible cortex has been particularly studied by researchers in order to detect lower BMD. A number of mandibular cortical indexes including the mandibular cortical index (MCI), mandibular cortical width (MCW), and panoramic mandibular index (PMI), have been developed to assess and measure the quality of mandibular bone mass and to observe signs of resorption. The MCI is one of the most commonly used osteoporosis risk indexes in the diagnosis of osteoporosis with a sensitivity ranging from 48.7–100% and specificity up to 88.89% in detecting BMD, and with a sensitivity between 35.9% and 90.9% and specificity up to 93.9% for diagnosis of osteoporosis.

The Computer-Assisted Diagnosis (CAD) systems that may help differential diagnosis by using medical images take place in the literature day by day along with the recent development of artificial intelligence (AI) methods and computer hardware. The image classification, segmentation, detection and recognition processes appear with various applications in every clinic of dentistry through development of deep learning algorithms which is a sub-branch of AI since 2012. Some of these include detection of the tooth number, detection of caries, detection of apical pathologies, detection of periodontal bone loss. There are studies on prediction and classification of osteoporosis with machine learning and fuzzy systems by obtaining features from panoramic images in the literature. However, deep learning methods reveal superior implementations when compared to traditional machine learning methods. Machine learning methods take the features obtained by gray-level matrix occurred concomitantly, wavelet local binary pattern, and other texture analyzes as input for the classification of an image and classify them with feature selection methods. Since there are numerous feature extraction and selection methods, and the search for suitable methods for the machine learning method to be used is a time-consuming and costly process. In deep learning methods, the image is taken as a direct input and features are automatically obtained and selected along the created layers.
Gradient-weighted Class Activation Mapping (Grad-CAM) method visually enables to locate the weights obtained as a result of deep learning gather the feature from which image regions. 21

The aim of the present study was to predict the radiological changes compatible with osteoporosis by using different deep learning algorithms on panoramic radiographs in women over 50 years of age.

2. Methods And Materials

2.1. The Study Design and Sampling

This study was carried out retrospectively through panoramic radiographs taken for various diagnostic reasons at Necmettin Erbakan University, Faculty of Dentistry, Department of Oral and Maxillofacial Radiology between 2014 and 2021. The study protocol was approved by Necmettin Erbakan University Faculty of Dentistry Research Ethics Committee (No: 10-84) and conducted in accordance with the principles defined in the Declaration of Helsinki, including all revisions. In case of retrospective study, Necmettin Erbakan University Faculty of Dentistry waived the need for informed consent.

Panoramic radiographs of female patients over 50 years of age without history of systemic disease affecting the bone metabolism and medical treatment and maxillofacial trauma were reviewed. Panoramic radiographs with artifacts which are not sufficient in terms of diagnosis, especially in which the hyoid bone is superimposed to the mandibular cortical region were excluded from the study. Seven hundred and forty-four panoramic radiographs were evaluated in accordance with the specified criteria.

2.2. Image Acquisition and Radiographic Evaluation

All panoramic radiographs reviewed were taken by a single device and acquired with a 2D Veraviewpocs (J MORITA MFG corp, Kyoto, Japan) digital panoramic x-ray device with 70 kVp, 5 mA and 15 sec irradiation time in accordance with the exposure protocols determined by the manufacturer. Studies were conducted by the same single observer through i-Dixel (J Morita MFG Corp., Kyoto, Japan) software. Intra-observer correlation coefficient Kappa value = 1.

Categorization of panoramic radiographs was performed by Mandibular Cortex Index (MCI) developed by Klemetti et al.22 depending on the cortical bone morphology on the posterior of mental foramen. The MCI is a cortical index that examines the region extending from the distal mental foramen to the antegonial region. Accordingly; C1 (Normal Mandibular Cortex): the endosteal margins of the cortex are even and sharp on both sides (Fig. 1a); C2 (Mildly or Moderately Eroded Cortex): Endosteal margins present crescent-shaped defects (lacunar resorption) (Fig. 1b); C3 (Severely Eroded Cortex): The cortical layer contains intense endosteal remnants and the cortex is severely porous (Fig. 1c).

2.3. Dataset and Image Pre-Processing

Each panoramic image cut as right-left was labeled as C1,C2,C3 by an oral radiologist. The data were exposed to the three phased image pre-processing in order to predict by deep learning algorithms. In the first stage, the right and left ROI areas were cut and recorded in *.tiff image format by considering the existence of the region extending from the distal to the antegonial region of the mental foramen.23 ImageJ v1.52 for Windows which is a version of the National Institutes of Health (NIH) Image software was used for determination of the ROIs. A dataset (744 individuals) was thereby created with a total of 1488 images including 597 images tagged with C1, 581 images tagged with C2, and 310 images tagged with C3.

Since anatomical structures and image sizes of the individuals are different, the cropped images have different width and height pixel sizes. Since standard size data will be input to the input of deep learning algorithms, standardization is required in image sizes. The second stage included resizing of images to 224x224 by using the bilinear interpolation method for this procedure. The deep learning algorithms used process with three-channel image input. Therefore, the images were reconverted into three channels in the third step. This procedure does not convert the image to color format, it only adds the same image one after the other, allowing the algorithms to process in three channels. The flow chart of right and left cutting of the images, the standardization process, and conversion into three channels are presented in Figure 2.

2.4. Deep Learning

In the present study, GoogleNET24, AlexNET25, ResNET-5026, ShuffleNET27, SqueezeNET28 models were used for osteoporosis classification through transfer learning method. Algorithms were run on a computer with Intel Core i7-7700HQ 2.80 Ghz processor, 16 GB RAM, NVIDIA GTX 1050 graphics card. Deep learning hyper-parameters are adjusted equally in order to operate the algorithms under equal conditions. The data
increase and frozen process in deep learning layers were not implemented. Deep learning algorithms were developed in MATLAB 2021a program and operated on GPU graphics card. Hyper-parameters used in deep learning models were provided in Table 1.

| Hyper Parameters | Value                                      |
|------------------|--------------------------------------------|
| Optimisation     | (SGDM) Stochastic Gradient Descent with Momentum |
| Momentum         | 0.9                                        |
| InitialLearnRate | 3.00E-04                                   |
| LearnRateDropFactor | 0.1                                   |
| LearnRateDropPeriod | 10                                   |
| L2Regularization | 1.00E-04                                   |
| GradientThresholdMethod | l2norm                          |
| GradientThreshold | Inf                                        |
| MaxEpochs        | 30(15 ~ only ShuffleNET)                  |
| MiniBatchSize    | 32                                         |
| Shuffle          | every epoch                                |
| ValidationFrequency | Number of train files /miniBatchSize       |
| ExecutionEnvironment | GPU                                          |

### 2.5. Model Training

Four different datasets were designed in order to measure the classification performance of deep learning algorithms in the present study. The original data set C1-C2-C3 was used to measure the three-class output performance of the systems. The original data set is divided into C1-C2, C1-C3, C1-(C2+C3) for measurement of the two-class output performances. The aim of this procedure was to compare the two-class output performances of the systems against the three-class output performances. The estimates of C1-labeled normal images according to C2 and C3 were analyzed in the C1-C2, C1-C3 datasets, and presence of osteoporosis-osteopenic findings was analyzed in the C1-(C2+C3) dataset.

For the test set, 20% of the data in different labels were randomly selected. The remaining data are split at each iteration through 5-fold cross validation as training by 80% and validation by 20%. The test data undetected by the system were given to the trained network and the classification performance of the system was measured. The data set used in the study was presented in Table 2.

| Data Set          | C1  | C2  | C3  | C2+C3 | Total | Test | Train+Validation |
|-------------------|-----|-----|-----|-------|-------|------|------------------|
| C1-C2-C3          | 597 | 581 | 310 | -     | 1488  | 297  | 1191             |
| C1-C2             | 597 | 581 | -   | -     | 1178  | 235  | 943              |
| C1-C3             | 597 | -   | 310 | -     | 907   | 181  | 726              |
| C1-(C2+C3)        | 597 | -   | -   | 891   | 1488  | 297  | 1191             |

Figure 3 shows the flow chart of the training, validation and testing model used in the study with 5-fold cross validation. Statistical values were obtained after each fold and the average of the performance criteria was taken after the cross-validation was completed. Four different datasets given in Table 2 were tested in 20 different scenarios through 5 different deep learning algorithms and the methodology given in Figure 3.

### 2.6. Performance Criteria
The performance evaluation of the models was conducted with the test dataset obtained by taking 20% of the entire dataset randomly. The models were thereby evaluated by previously undetected data. The complexity matrix of the system was obtained and TP, TN, FP, FN values were calculated after each fold. The precision (Equation 1), recall (Equation 2), accuracy (Equation 3), specificity (Equation 4), F1-Score (Equation 5) values of the model results were calculated by using these values.

\[
\text{precision} = \frac{TP}{TP + FP}
\]

\[
\text{recall} = \frac{TP}{TP + FN}
\]

\[
\text{accuracy} = \frac{TP + TN}{TP + FN + TN + FP}
\]

\[
\text{specificity} = \frac{TN}{TN + FP}
\]

\[
F1 - \text{Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\]

TPR (True Positive Rate) and FPR (False Positive Rate) calculations were done for each fold performance evaluation. The ROC curves of each class were drawn with these values, and AUC calculations were obtained. The average performance results of the models are presented as a table by taking the average of all performance values. Total time spent in model training was also calculated. The complexity matrix obtained in GoogleNET fold5 is given in Figure 4a, the ROC curves and AUC values of each class are given in Figure 4b. This procedure was performed at each iteration.

3. Results

3.1. Estimation Performances of Models

The average performance criteria obtained with four different datasets and the total training times of the models are given in Table 3. The AUC values of the classes of each data set and the average AUC values of the models are given in Table 4. The highest accuracy rate was 81.14% in the C1-C2-C3 dataset with AlexNET, 88.94% in the C1-C2 dataset with GoogleNET, 98.56% in the C1-C3 dataset with AlexNET in the shortest duration; and the highest accuracy rate was 92.79% in the C1-(C2+C3) dataset with GoogleNET. Since the number of labels in the datasets is not evenly distributed, the data is defined as an unbalanced dataset. Therefore, although the accuracy rate shows the classification performance, another parameter to be considered in unbalanced data sets is the AUC values. Since the AUC values reveal the area under the ROC graphs, it is frequently used in the classification performance evaluations of the models. If the model estimation classes are more than two, the AUC values give the classification performance of one class against the others. The values of both classes are equal or close to each other in cases with two outputs. When the models with the highest accuracy were compared according to their AUC values, AUC values in the C1-C2-C3 dataset with AlexNET were 0.9662(AUC-C1), 0.8826(AUC-C2), 0.9601(AUC-C3), 0.9363(AUC-Mean); AUC values in the C1-C2 data set were 0.9560(AUC-C1, AUC-C2, AUC-Mean) with GoogleNET; AUC values in the C1-C3 dataset was 0.9987 (AUC-C1, AUC-C2, AUC-Mean) with AlexNET; AUC values in the C1-(C2+C3) data set were 0.9789 (AUC-C1)0.9785 (AUC-(C2+C3)) 0.9787(AUC-Mean) with GoogleNET (Table 4). The highest estimation performances of the models trained in this study are given in Table 5.
### Table 3
The performance values of 5 deep learning models for 4 different data sets

|       | Precision      | Recall       | Specificity   | Accuracy       | F1-Score     | Training Time |
|-------|----------------|--------------|---------------|----------------|--------------|---------------|
| **C1-C2-C3** |                |              |               |                |              |               |
| GoogleNET | 0.7914±(0.0258) | 0.8116±(0.0279) | 0.8997±(0.0146) | 0.7926±(0.0306) | 0.8014±(0.0262) | 55 min       |
| AlexNET   | 0.8065±(0.0248) | 0.8204±(0.0149) | 0.9048±(0.0083) | 0.8114±(0.0203) | 0.8134±(0.0198) | 35 min 43 sec |
| ResNET-50 | 0.7450±(0.0183) | 0.7652±(0.0140) | 0.8737±(0.0080) | 0.7508±(0.0163) | 0.7550±(0.0159) | 114 min 19 sec |
| ShuffleNET | 0.7258±(0.0189) | 0.7502±(0.0233) | 0.8643±(0.0105) | 0.7333±(0.0193) | 0.7378±(0.0207) | 72 min 38 sec |
| SqueezeNET | 0.7722±(0.0249) | 0.7735±(0.0545) | 0.8775±(0.0325) | 0.7556±(0.0587) | 0.7725±(0.0388) | 28 min 18 sec |
| **C1-C2** |                |              |               |                |              |               |
| GoogleNET | 0.9311±(0.0251) | 0.8437±(0.0697) | 0.9362±(0.0233) | 0.8894±(0.0391) | 0.8842±(0.0459) | 43 min 32 sec |
| AlexNET   | 0.8679±(0.0265) | 0.8672±(0.0110) | 0.8638±(0.0325) | 0.8655±(0.0161) | 0.8674±(0.0140) | 25 min 30 sec |
| ResNET-50 | 0.8590±(0.0315) | 0.7899±(0.0238) | 0.8655±(0.0369) | 0.8272±(0.0137) | 0.8224±(0.0123) | 50 min 8 sec |
| ShuffleNET | 0.8228±(0.0407) | 0.7664±(0.0748) | 0.8259±(0.0626) | 0.7957±(0.0206) | 0.7904±(0.0303) | 56 min 30 sec |
| SqueezeNET | 0.9228±(0.0712) | 0.8336±(0.1069) | 0.9190±(0.0794) | 0.8757±(0.0405) | 0.8694±(0.0509) | 20 min 16 sec |
| **C1-C3** |                |              |               |                |              |               |
| GoogleNET | 0.9949±(0.0076) | 0.9832±(0.0059) | 0.9903±(0.0144) | 0.9856±(0.0074) | 0.9890±(0.0057) | 35 min 15 sec |
| AlexNET   | 0.9899±(0.0038) | 0.9882±(0.0046) | 0.9806±(0.0072) | 0.9856±(0.0049) | 0.9891±(0.0038) | 18 min 38 sec |
| ResNET-50 | 0.9810±(0.0110) | 0.9429±(0.0201) | 0.9645±(0.0210) | 0.9503±(0.0110) | 0.9614±(0.0089) | 39 min 7 sec |
| ShuffleNET | 0.9740±(0.0058) | 0.9429±(0.0162) | 0.9516±(0.0114) | 0.9459±(0.0099) | 0.9581±(0.0080) | 44 min 6 sec |
| SqueezeNET | 0.9950±(0.0075) | 0.9798±(0.0113) | 0.9903±(0.0144) | 0.9834±(0.0068) | 0.9873±(0.0052) | 18 min 28 sec |
| **C1-(C2+C3)** |                |              |               |                |              |               |
| GoogleNET | 0.9592±(0.0223) | 0.8571±(0.0380) | 0.9753±(0.0141) | 0.9279±(0.0154) | 0.9048±(0.0216) | 58 min 35 sec |
| AlexNET   | 0.9774±(0.0084) | 0.8034±(0.0666) | 0.9876±(0.0047) | 0.9138±(0.0271) | 0.8807±(0.0420) | 30 min 2 sec |
| ResNET-50 | 0.9240±(0.0192) | 0.7479±(0.0486) | 0.9584±(0.0129) | 0.8741±(0.0163) | 0.8257±(0.0276) | 65 min 49 sec |
| ShuffleNET | 0.9135±(0.0169) | 0.7429±(0.0816) | 0.9528±(0.0116) | 0.8687±(0.0298) | 0.8173±(0.0504) | 71 min 17 sec |
| SqueezeNET | 0.9587±(0.0109) | 0.7983±(0.0982) | 0.9764±(0.0083) | 0.9051±(0.0349) | 0.8679±(0.0593) | 27 min 39 sec |
Table 4
AUC value and average AUC values of each class for 5 different deep learning models in 4 different datasets

|        | AUC-C1                     | AUC-C2                     | AUC-C3                     | AUC C2-C3             | AUC-Mean                |
|--------|----------------------------|----------------------------|----------------------------|------------------------|-------------------------|
| C1-C2-C3 | GoogleNET 0.9763±(0.0025) | 0.8725±(0.0176) | 0.9621±(0.0015) | -                      | 0.9369±(0.0072)         |
|         | AlexNET 0.9662±(0.0050)    | 0.8826±(0.0225) | 0.9601±(0.0108) | -                      | 0.9363±(0.0128)         |
|         | ResNET-50 0.9305±(0.0101)  | 0.8271±(0.0117) | 0.9458±(0.0085) | -                      | 0.9012±(0.0101)         |
|         | ShuffleNET 0.9135±(0.0060) | 0.8136±(0.0088) | 0.9458±(0.0091) | -                      | 0.8910±(0.0080)         |
|         | Squuzeenet 0.9737±(0.0065) | 0.8479±(0.0447) | 0.9495±(0.0071) | -                      | 0.9237±(0.0194)         |
| C1-C2   | GoogleNET 0.9560±(0.0141)  | 0.9560±(0.0141) | -                      | -                      | 0.9560±(0.0141)         |
|         | AlexNET 0.9366±(0.0124)    | 0.9366±(0.0124) | -                      | -                      | 0.9366±(0.0124)         |
|         | ResNET-50 0.9080±(0.0076)  | 0.9080±(0.0076) | -                      | -                      | 0.9080±(0.0076)         |
|         | ShuffleNET 0.8806±(0.0065) | 0.8806±(0.0065) | -                      | -                      | 0.8806±(0.0065)         |
|         | Squuzeenet 0.9550±(0.0186) | 0.9550±(0.0186) | -                      | -                      | 0.9550±(0.0186)         |
| C1-C3   | GoogleNET 0.9985±(0.0008)  | -                      | 0.9985±(0.0008) | -                      | 0.9985±(0.0008)         |
|         | AlexNET 0.9987±(0.0012)    | -                      | 0.9987±(0.0012) | -                      | 0.9987±(0.0012)         |
|         | ResNET-50 0.9883±(0.0060)  | -                      | 0.9883±(0.0060) | -                      | 0.9883±(0.0060)         |
|         | ShuffleNET 0.9915±(0.0012) | -                      | 0.9915±(0.0012) | -                      | 0.9915±(0.0012)         |
|         | Squuzeenet 0.9975±(0.0022) | -                      | 0.9975±(0.0023) | -                      | 0.9975±(0.0022)         |
| C1-(C2+C3) | GoogleNET 0.9789±(0.0050) | -                      | -                      | 0.9785±(0.0057) | 0.9787±(0.0053)         |
|         | AlexNET 0.9810±(0.0051)    | -                      | -                      | 0.9810±(0.0051) | 0.9810±(0.0051)         |
|         | ResNET-50 0.9574±(0.0071)  | -                      | -                      | 0.9574±(0.0070) | 0.9574±(0.0070)         |
|         | ShuffleNET 0.9438±(0.0110) | -                      | -                      | 0.9438±(0.0110) | 0.9438±(0.0110)         |
|         | Squuzeenet 0.9775±(0.0096) | -                      | -                      | 0.9774±(0.0099) | 0.9775±(0.0098)         |

Table 5
The highest estimation performances obtained from the models

|        | Precision | Recall/ Sensitivity | Specificity | Accuracy | F1-Score | AUC | Training Time |
|--------|-----------|---------------------|-------------|----------|----------|-----|---------------|
| C1-C2-C3 | AlexNET   | 0.8065±(0.0248) | 0.8204±(0.0149) | 0.9048±(0.0083) | 0.8114±(0.0203) | 0.8134±(0.0198) | 0.9363±(0.0128) | 35 min 43 sec |
| C1-C2   | GoogleNET | 0.9311±(0.0251) | 0.8437±(0.0697) | 0.9362±(0.0233) | 0.8894±(0.0391) | 0.8842±(0.0459) | 0.9560±(0.0141) | 43 min 32 sec |
| C1-C3   | AlexNET   | 0.9899±(0.0038) | 0.9882±(0.0046) | 0.9806±(0.0072) | 0.9856±(0.0049) | 0.9891±(0.0038) | 0.9987±(0.0012) | 18 min 38 sec |
| C1-(C2+C3) | GoogleNET | 0.9592±(0.0223) | 0.8571±(0.0380) | 0.9753±(0.0141) | 0.9279±(0.0154) | 0.9048±(0.0216) | 0.9787±(0.0053) | 58 min 35 sec |

3.2. Visualization of Model Estimations

Feature maps were determined by Grad-CAM method on the images labeled with C1, C2, C3 given in Figure 1 through the model weights with the highest accuracy rates. Heat maps were obtained according to the weights obtained after the dataset and transfer learning by overlaying the original image with the Grad-CAM map. The performance of tissue characteristics determined by the oral radiologist may be thereby compared with the maps acquired by deep learning. GoogleNET Grad-CAM results for dataset C1C2 were provided in Figure 5; AlexNET Grad-
Osteoporosis is a silent disease and individuals are not aware of this condition until they experience bone fractures. Therefore, dentists will be able to detect patients at an early stage and direct them for early treatment with an effective CAD system for detection of osteoporosis on panoramic radiographs. Accordingly, deep learning models of AlexNET, GoogleNET, ResNET-50, SqueezeNET and ShuffleNET were trained through the transfer learning method in this study in order to predict osteoporosis by using MCI over panoramic radiographs of female patients over 50 years old. The results were evaluated by performance criteria including accuracy, sensitivity, specificity, F1-score, AUC, and training duration.

There are several studies in which the presence of osteoporosis was investigated on panoramic radiographs with some classical image processing methods, machine learning algorithms and deep learning algorithms in the literature. Jae-Seo Lee et al. estimated the presence and absence of osteoporosis through the single-column deep convolutional neural network (SC-DCNN), multi-column deep convolutional neural network (MC-DCNN) methods by cutting the right and left mandibular regions of 1268 panoramic radiographs labeled with MCI. A trial conducted with 200 test data, the MC-DCNN method provided an accuracy rate of 98.5% and an AUC value of 0.9987. It is obvious in the study where the cross-validation method was not used that a different result would be obtained in each trial. Different from the study stating that healthy and osteoporotic individuals were differentiated only, estimation of osteopenic findings scored as C2 was performed in this research. The most striking result was in the C1-C3 dataset where the radiological change related to osteoporosis is most evident among the model performances given in Table 5 where the highest prediction performances are shown.

Classification of this dataset was performed through the AlexNET model with a total training duration of 18 minutes and 38 seconds with an accuracy rate of 98.56%, and an AUC value of 0.9987. The feature extraction map of the model is seen when the Grad-CAM images given in Figure 6 are reviewed for differentiation of C1-C3 where the cortical bone change was most evident. As the score progresses from C1 to C3, the cortical change differentiates at length to include the entire area of interest, the areas shown in red are the regions on which the model bases on for distinction between C1 and C3. The total training duration of 43 minutes and 32 seconds with GoogleNET in differentiation of C1-C2 scores has realized with an accuracy rate of 88.94% and a value of 0.9560 AUC. The C2 score shows radiological features close to the C1 score; this may be explained by the lower AUC values in differentiation of C1 and C2. Since the data is unevenly distributed (imbalanced because the number of labels in the datasets is not evenly distributed), the classification estimation performance of the model may be evaluated over the AUC. The three-class C1-C2-C3 dataset has realized with AlexNET with 35 minutes and 43 seconds of total training duration with an accuracy rate of 81.14% and AUC value of 0.9363. The review of GradCAM images given in Figure 7 revealed that the strong features in C1 and C3 scores were acquired from the mandibular cortex at length, and C2 score was acquired from certain regions where the mandibular cortex show porosity in a patchy pattern. In the C1-(C2+C3) dataset where the C2-C3 dataset was labeled as osteoporotic-osteopenic changes, the C1 dataset was labeled as no osteoporosis, presence and absence of osteoporosis and osteopenia were estimated; an accuracy rate of 92.79% and an AUC value of 0.9787 were acquired with a total training duration of 58 minutes and 35 seconds. Values above 0.9 were obtained in the review of AUC results of four datasets. The heat maps from which the model weights are obtained in the Grad-CAM images and the regions where the oral radiologist interprets the textural differences between the disease scores are consistent (Figures 5-8). These results show that the models may be used as a CAD system in order to assist the oral radiologist to estimate osteoporosis and mapping the localization sites of tissue change through MCI.

Ki-Sun Lee et al. classified panoramic images of 680 patients as osteoporosis and no-osteoporosis according to BMD values; 20% of the data was used for random test set, the remaining data was used for training and validation, data were given to deep learning inputs with 5-fold cross validation. The system performances of deep learning models of CNN, VGG-16, VGG-16_TF, VGG-16_TF_FT trained with simple three convolution layers are discussed. The highest values were obtained by VGG-16_TF_FT model with an accuracy rate of 84.00% and AUC of 0.858. Grad-CAM images were created through the model weights acquired with VGG-16_TF_FT. According to these images, the weak lower border of the mandibular cortical bone and the region presenting with less intense, spongey bone tissue in the periphery were evaluated as the regions where strong features were obtained in accurate estimation of osteoporosis. In the normal labeled image, the region
presenting with the strong lower border of the mandibular cortical bone and the surrounding dense tissue were determined as the region where strong features were obtained for estimation.

ROI cutting of panoramic radiographs was performed with ImageJ program in this study. The subsampling was then performed according to the row and column sizes consistent for the input of deep learning models. A homogeneous dataset was thereby created in which osteoporosis may be examined in the most adequate anatomical region in each panoramic radiograph. Subsampling panoramic radiographs first followed by cutting through a designated region would create a heterogeneous dataset for the area of interest. Therefore, deep learning models will result in biased training by searching for features in different mandible points and image spaces. Retrieving the entire mandible instead of just taking the ROI would cause search for more features in unnecessary regions by the model and increase the training duration. Although BMD indexing is defined as the gold standard by the World Health Organization, the MCI index suggested by Klemetti has been used in many studies. A random test dataset of 20% was determined in the study; the remaining data were given to the model inputs with 5-fold cross validation as training and validation. Undifferentiation of the random test data from the data and not providing to system inputs by 5-folc cross validation both causes failure to interpret the reaction of models against data which were not seen before, and average output of the model obtained through cross validation. Models would produce different results in each run without cross validation.

The limitation of the present study was a lack of a gold standard for diagnosis of osteoporosis in each of the cases. However, it is known that the mandibular cortical shape is significantly compatible with the skeletal BMD obtained with DXA. The images were acquired manually in order to include the mandibular inferior margin in the center of the mandibular body as a preoperative preparation for classification of osteoporosis. The construction of a network for scanning of osteoporosis from dental panoramic radiographs through automated detection of the ROI from untrimmed dental panoramic radiographs is required.

5. Conclusion

The regions where the textural features change in the radiological examination of the oral radiologist and regions where the deep learning models extract features are consistent. The deep learning-based CAD system implemented in this study may accelerate the processes that could help the diagnosis for the radiologist such as presence detection of osteoporosis through MCI and the evaluation of the textural changes between the scores. Early intervention and diagnostic losses that may occur due to the fact that dentists deal with general dental problems and the number of oral radiologists may be resolved quickly by the AI-based systems developed.

Declarations

Funding: None

Informed consent: Not required. The study protocol was approved by Necmettin Erbakan University Faculty of Dentistry Research Ethics Committee (No: 10-84). In case of retrospective study, Necmettin Erbakan University Faculty of Dentistry waived the need for informed consent.

References

1. Taguchi, A.; Tanaka, R.; Kakimoto, N.; Morimoto, Y.; Arai, Y.; Hayashi, T.; Kurabayashi, T.; Katsumata, A.; Asaumi, J. J. O. R., Clinical guidelines for the application of panoramic radiographs in screening for osteoporosis. 2021, 1–20.
2. Akyol, Y.; Ulus, Y.; Bilgici, A.; Kuru, Ö., Premenopozal ve Postmenopozal Kadınlarda Osteoporoz Bilgi ve Farkındalık Düzeyinin Karşılaştırılması. Turkish Journal of Osteoporosis/Turk Osteoporoz Dergisi 2020, 26 (1).
3. YILDIRIM, M. A., Osteoporoz Tanısı ve İzlem Parametreleri. Klinik Tip Aile Hekimliği 11 (3), 119–121.
4. Taguchi, A., Triage screening for osteoporosis in dental clinics using panoramic radiographs. Oral diseases 2010, 16 (4), 316–327.
5. Hastar, E.; Yilmaz, H. H.; Orhan, H., Evaluation of mental index, mandibular cortical index and panoramic mandibular index on dental panoramic radiographs in the elderly. European journal of dentistry 2011, 5 (01), 060–067.
6. Calciolari, E.; Donos, N.; Park, J.; Petrie, A.; Mardas, N., Panoramic measures for oral bone mass in detecting osteoporosis: a systematic review and meta-analysis. Journal of dental research 2015, 94 (3_suppl), 17S-27S.
7. Chan, H. P.; Hadijiski, L. M.; Samala, R. K., Computer-aided diagnosis in the era of deep learning. Medical physics 2020, 47 (5), e218-e227.
8. Fujita, H.; technology, AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. Radiological physics 2020, 13 (1), 6–19.
9. Tuzoff, D. V.; Tuzova, L. N.; Bornstein, M. M.; Krasnov, A. S.; Kharchenko, M. A.; Nikolenko, S. I.; Sveshnikov, M. M.; Bednenko, G. B., Tooth detection and numbering in panoramic radiographs using convolutional neural networks. *Dentomaxillofacial Radiology* 2019, 48 (4), 20180051.

10. Lee, J.-H.; Kim, D.-H.; Jeong, S.-N.; Choi, S.-H., Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm. *Journal of dentistry* 2018, 77, 106–111.

11. Orhan, K.; Bayrakdar, I.; Ezioh, M.; Kravtsov, A.; Özyürek, T., Evaluation of artificial intelligence for detecting periapical pathosis on cone-beam computed tomography scans. *International endodontic journal* 2020, 53 (5), 680–689.

12. Krois, J.; Ecket, T.; Meinhold, L.; Golla, T.; Kharbot, B.; Wittemeier, A.; Dörfer, C.; Schwendicke, F., Deep Learning for the Radiographic Detection of Periodontal Bone Loss. *Scientific Reports* 2019, 9 (1), 8495.

13. Roberts, M. G.; Graham, J.; Devlin, H., Image texture in dental panoramic radiographs as a potential biomarker of osteoporosis. *IEEE Transactions on biomedical engineering* 2013, 60 (9), 2384–2392.

14. Alzubaidi, M. A.; Otoom, M.; biomedicine, p. i., A comprehensive study on feature types for osteoporosis classification in dental panoramic radiographs. *Computer methods* 2020, 188, 105301.

15. Khojastepour, L.; Hasani, M.; Ghasemi, M.; Mehdizadeh, A.; Tajeripour, F., Mandibular trabecular bone analysis using local binary pattern for osteoporosis diagnosis. *Journal of biomedical physics engineering* 2019, 9 (1), 81.

16. Kavitha, M. S.; Ganesh Kumar, P.; Park, S.-Y.; Huh, K.-H.; Heo, M.-S.; Kurita, T.; Asano, A.; An, S.-Y.; Chien, S.-I., Automatic detection of osteoporosis based on hybrid genetic swarm fuzzy classifier approaches. *Dentomaxillofacial Radiology* 2016, 45 (7), 20160076.

17. Kavitha, M. S.; Asano, A.; Taguchi, A.; Kurita, T.; Sanada, M., Diagnosis of osteoporosis from dental panoramic radiographs using the support vector machine method in a computer-aided system. *BMC medical imaging* 2012, 12 (1), 1–11.

18. Arifin, A. Z.; Asano, A.; Taguchi, A.; Nakamoto, T.; Ohtsuka, M.; Tsuda, M.; Kudo, Y.; Tanimoto, K., Use of fuzzy neural network in diagnosing postmenopausal women with osteoporosis based on dental panoramic radiographs. *Journal of Advanced Computational Intelligence Vol* 2007, 11 (8).

19. Park, C.; Took, C. C.; Seong, J.-K., Machine learning in biomedical engineering. Springer: 2018; Vol. 8, pp 1-3.

20. Pouyanfar, S.; Sadik, S.; Yan, Y.; Tian, H.; Tao, Y.; Reyes, M. P.; Shyu, M.-L.; Chen, S.-C.; Iyengar, S. S., A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)* 2018, 51 (5), 1–36.

21. Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. In *Grad-cam: Visual explanations from deep networks via gradient-based localization*, Proceedings of the IEEE international conference on computer vision, 2017; pp 618-626.

22. Klemetti, E.; Kolmakov, S.; Kröger, H., Pantomography in assessment of the osteoporosis risk group. *European Journal of Oral Sciences* 1994, 102 (1), 68–72.

23. Ferreira, T.; Rasband, W., ImageJ user guide. *ImageJ/Fiji* 2012, 1, 155–161.

24. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. In *Going deeper with convolutions*, Proceedings of the IEEE conference on computer vision and pattern recognition, 2015; pp 1-9.

25. Krizhevsky, A.; Sutskever, I.; Hinton, G. E., Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 2012, 25, 1097–1105.

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

27. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. In *Shufflenet: An extremely efficient convolutional neural network for mobile devices*, Proceedings of the IEEE conference on computer vision and pattern recognition, 2018; pp 6848-6856.

28. Iandola, F. N.; Han, S.; Moskewicz, M. W.; Ashraf, K.; Dally, W. J.; Keutzer, K., SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv preprint arXiv:1602.07360* 2016.

29. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P, Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998, 86 (11), 2278-2324.

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

31. Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. In *A survey on deep transfer learning*, International conference on artificial neural networks, Springer: 2018; pp 270-279.

32. Lee, J.-S.; Adhikari, S.; Liu, L.; Jeong, H.-G.; Kim, H.; Yoon, S.-J., Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study. *Dentomaxillofacial Radiology* 2019, 48 (1), 20170344.
33. Devlin, H.; Karayianni, K.; Mitsea, A.; Jacobs, R.; Lindh, C.; van der Stelt, P.; Marjanovic, E.; Adams, J.; Pavitt, S.; Horner, K., Diagnosing osteoporosis by using dental panoramic radiographs: the OSTEODENT project. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology* 2007, 104 (6), 821–828.

34. Kavitha, M. S.; An, S.-Y.; An, C.-H.; Huh, K.-H.; Yi, W.-J.; Heo, M.-S.; Lee, S.-S.; Choi, S.-C., Texture analysis of mandibular cortical bone on digital dental panoramic radiographs for the diagnosis of osteoporosis in Korean women. *Oral surgery, oral medicine, oral pathology and oral radiology* 2015, 119 (3), 346-356.

35. Sharma, S.; Agarwal, A.; Singh, C., Osteoporosis: Detection Using Dental Radiography. *J Dent Oral Sci* 2020, 2 (3), 1–7.

36. Lee, K.-S.; Jung, S.-K.; Ryu, J.-J.; Shin, S.-W.; Choi, J., Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs. *Journal of clinical medicine* 2020, 9 (2), 392.

**Figures**

![Images](image.jpg)

**Figure 1**

(a) C1; (b) C2; (c) C3 according to MCI

**Figure 2**

Cutting of the images as right and left, the standardization process, and conversion into three channels
Figure 3

The flow chart of data splitting, 5-fold cross validation and performance evaluation approaches used in the study
Figure 4

(a) The complexity matrix obtained in GoogleNET fold5 (b) ROC curves and AUC values of each class obtained in GoogleNET Fold 5.

Figure 5

GoogleNET Grad-CAM results for dataset C1C2 (a) C1 (b) C2

Figure 6

GoogleNET Grad-CAM results for dataset C1C3 (a) C1 (b) C3

Figure 7

AlexNET Grad-CAM results for dataset C1C2C3 (a) C1 (b) C2 (c) C3
Figure 8

GoogleNET Grad-CAM results for dataset C1-(C2+C3) (a) C1 (b) C2 (c) C3