

Periapical Radiograph Texture Features for Osteoporosis Detection using Deep Convolutional Neural Network

Khasnur Hidjah¹

Department of Computer Science and Electronics
Universitas Gadjah Mada, Yogyakarta, Indonesia
Departement of Computer Science
Universitas Bumigora, Mataram, Indonesia

Agus Harjoko^{2*}

Department of Computer Science and Electronics
Universitas Gadjah Mada
Yogyakarta, Indonesia

Moh. Edi Wibowo³

Department of Computer Science and Electronics
Universitas Gadjah Mada
Yogyakarta, Indonesia

Rurie Ratna Shantiningsih⁴

Departement of Dentomaxillofacial Radiology
Universitas Gadjah Mada
Yogyakarta, Indonesia

Abstract—Currently, research for osteoporosis examination using dental radiographic images is increasing rapidly. Many researchers have used various methods from subject data. It indicates that osteoporosis has become a widespread disease that should be studied more deeply. This study proposes a deep Convolutional Neural Network architecture as a texture feature of dental periapical radiograph for osteoporosis detection. The subject of this study is postmenopausal Javanese women aged over 40 and data measurement result of Bone Mineral Density. The proposed model is divided into stages: 1) stage image acquisition and RoI selection, 2) stage feature extraction and classification. Various experiments with the number of convolution layers (3 layers to 6 layers) and various input block sizes and other hyper parameters were used to get the best model. The best model is obtained when the input image size is greater than 100 and less than 150 and a five of convolution layer, as well as other hyper parameters, including epochs=100, dropout=0.5, learning rate=0.0001, batch size= 16 and loss function using Adam's optimization. Validation and testing accuracy achieved by the best model is 98.10%, and 92.50. The research shows that the bigger images provide additional information about trabecular patterns in normal, osteopenia and osteoporosis classes, so that the proposed method using deep convolutional neural network as textural feature of the periapical radiograph achieves a good performance for detection osteoporosis.

Keywords—Osteoporosis; dental periapical radiograph; convolutional neural network; texture features; bone mineral density

I. INTRODUCTION

Osteoporosis is defined as a systemic skeletal disease characterized by low bone mass and micro-architectural deterioration of bone tissue [1]. Therefore, osteoporosis will increase the consequences of bone fragility and susceptibility to fracture, especially for those over 50 years of age. Once a fracture happens to someone with osteoporosis, life will be greatly affected due to disability to move and prolonged

healing process. Finally, this reduces a person's quality of life and causes various economic and social problems [2].

For example, if the injured person works as a driver or as a labor worker, he might have to retire and find some disk-related job that is not easy to obtain. In some other cases, the injured person might become severely disabled and require continuous assistance, which might burden his family. Therefore, preventative measures and early treatment of osteoporosis [3-4] are the best options to address these issues. Practical scientific and technological methods to support osteoporosis diagnosis, in this context, will provide much help to overcome the disease and reduce its negative impacts.

The most accurate BMD examination and made the gold standard by World Health Organization (WHO) is using Dual Energy X-Ray Absorptiometry (DEXA). However, access to this method is still limited in many countries. BMD examination is often available in central hospitals only, and its cost is often too expensive for many people in rural areas. Furthermore, BMD is not able to reveal the internal structure of fractured bones [4-5]. Researchers, therefore, have attempted to develop alternative methods that are more practical and more widely accessible. Several studies have found that dental data demonstrate a high correlation with BMD measurements [6-16]. The data include panoramic and periapical radiographs. Besides that, of the use widespread of periapical radiographs in dental care for the elderly with increased life expectancy and the number of studies according to BMD estimates and screening for osteoporosis using periapical radiographs. It is expected to provide benefits, namely the architecture that has been produced can then be used as an architectural model in pre-train medical images for different cases specifically using medical images with the same characteristics, which tend to be low resolution. In addition, it can also help patients who perform dental checkup at the dentist to be given a referral to chiropractors for further

*Corresponding Author

checkup, so that it can be detected early if you have osteoporosis.

Periapical radiograph is a dental radiograph technique that can image four to five teeth and their respective areas on one intraoral X-ray films [7]. At a microstructure level, images of trabecular jaw bones often show visual patterns closely related to the general condition of other bones [8,10]. Dental data have therefore become promising sources to predict Bone Mineral Density (BMD) measurements accurately. Periapical radiographs, in particular, will become the focus of this research since these radiographs are much more affordable and generally are more available. Fig. 1 shows an example of a periapical radiograph of the mandible.

The remaining part of this research is organized as follows. Section II explains some related work on osteoporosis detection using dental periapical radiograph, while Section III provides a more detailed description of the proposed method and training CNN's. Section IV contains result and discussion of the proposed method, and Section V concludes the research.

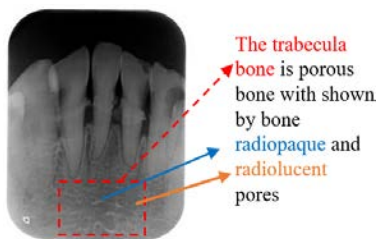


Fig. 1. Periapical Radiograph Image of the Mandibl.

II. RELATED WORK

Several studies have been carried out to examine osteoporosis using periapical radiographs. In [8] has made one of attempts to predict bone mass from such radiographs. Data and the corresponding labels (BMD measurements) are collected from 60 postmenopausal women aged over 40. This research extracts and combines visual features such as areas, lengths, and peripheries of “bright” blobs, numbers of terminal/branch points, and clinical data such as ages, body sizes, and calcium intakes. The overall number of pixels in the RoI divides that of black pixels results in the visual features of the trabecular area. The total amount of pixels in the external of the trabeculae in the binary images displays the periphery. In other words, it represents a proportion of the entire area of the trabeculae or the entire area of RoI. The skeletonized image computes the complete area of the skeletonized trabeculae (total amount of black pixels), the number of terminal points (free ends, that is, black pixels with only one neighboring black pixel), and the number of branch points (crossing points, that is, black pixels with three or more neighboring black pixels). These are used to represent the percentage of trabecular length, area and perimeter. Classification and regression tree analysis (CART) uses patients in groups of normal or low bone mass categories. It is evident from the CART analysis of clinical and radiographic features that the main element to categorize patients as having normal or low bone mass were age (± 42.5 years) and the number of terminal points as a function of the periphery (± 0.09). This algorithm conscientiously

distinguished 22 normal patients by BMD (specificity = 100%) and 31 patients with low bone mass (sensitivity = 81%). The total accuracy was 88.33%. A denomination of the corresponding predicted and actual bone category, the weighted kappa index, was 0.76. To identify women with low BMD, trabecular morphology analysis was an alternative. Another research [9] was also conducted displaying the combination of upper and lower jaw radiographs from 505 postmenopausal women aged 45–70 years. Dense, sparse or mixed trabecular patterns were identified by five observers. The gradings were integrated into a single averaged observer score per jaw in which the RoI can be identified on each by scanned radiographs. The RoIs compounded with image analysis software measured 25 photographs' characteristics. Pearson correlation and multiple linear regressions, which were used in identifying the averaged observer score, showed that 14 image features were significantly correlated with the observer judgment for the two jaws. Other features, which give details of osteoporotic patients with fewer but bigger marrow spaces than controls, are less compatible with the sparseness of the trabecular pattern than a rather crude measure for a structure such as the average grey value. To sum up, the human concept of sparseness is emanated more from average grey values of the RoI than from geometric details within the RoI. In [11], the bone mass prediction on porosities, connectivity, and orientations of porous was shown in trabecular images and a combination of the anthropometric features (weight, height, age, body mass index). While a decision tree was used to select the feature, a backpropagation artificial neural network was used for classification. By combining age, weight, height, body mass index and features of trabecular morphology interdental bone, identifying postmenopausal women with low bone mass are much easier. In this study, however, age is considered one of the biggest contributors to loss of bone mass. Porosity, the oblique porous, and the vertical porous are crucial porous features. This study distinguishes anthropometric and radiographic features, which then is analyzed individually. Both anthropometric features and the radiographic features have high accuracy with 80.33% and 87.04%, respectively. This work has been extended further [12] that combine data from periapical and panoramic radiographs. Furthermore in [13], a method for osteoporosis identification based on the validated trabecular area was presented on digital dental radiographic images. The image RoI of the validated trabecular area on the images should be obtained through a sequence of morphological operations, which is then evaluated using the Dice similarity method. In analyzing osteoporosis, a mineral density is estimated using dual X-ray absorptiometry in two areas and by extracting RoI through statistical features (deviation, entropy, homogeneity, and correlation). Feature extraction and feature selection are used to analyze the four features. The selection process applies the C4.5 feature selection method. Subsequently, to estimate the existence of osteoporosis, a multilayer perceptron of statistical texture analysis is employed. 0.8924. is obtained as the result of the average dice similarity coefficient for all of RoIs. The most suitable method in this proposed study, achieving an accuracy of 87.87%, is a multilayer perceptron classifier.

A study on the analysis of the mandibular trabecular structure in postmenopausal women using periapical

radiographs was conducted by [14]. The mandibular trabecular structure parameter used was the thickness of the trabeculae compared with the results of BMD DXA measurements in the lumbar (spine) and femoral (hip) areas. Determine the RoI manually with a size of 100pixels x 100pixels with a position 2mm from the apical edge of the left and right posterior parts of the lower jaw (mandible). Measurement of trabecular thickness using slice geometry features, namely bone area fraction size by dividing the number of pixels classified as bone by the total pixel area in RoI and trabecular thickness 2D is the average trabecular width in RoI. These two parameters were correlated with femoral and lumbar BMD values. Then the measurement results with statistical analysis showed a statistically significant difference between the normal group and the osteoporosis group compared to the normal group and the osteopenia group, so it can be said that thinning of the trabecular structure is more clearly seen in postmenopausal women with osteoporosis, with bone quality that can be detected earlier using the trabecular thickness parameter.

Although several methods have been proposed to examine osteoporosis using dental periapical radiographs, the methods generally rely on morphological analysis and geometric features of the images [8–14]. Only a little work has been conducted to investigate the effectiveness of texture features such as [15–16] that employed features derived from the Gray Level Co-occurrence Matrix (GLCM). However, the employed features are considered to be handcrafted, which might be suboptimal for the given problem. Therefore, the facts mentioned above have suggested further investigation on the use of texture features, particularly those that are directly learned from data. This research proposes deep learning in the analysis of texture features for the prediction of osteoporosis. Deep learning has worked effectively in many areas, including computer vision, hyperspectral image processing, medical image analysis [17], and natural language processing include in tuning for hyperparameters online [18]. Compared to conventional methods such as support vector regressors and multi-layer perceptron, based on feature, deep learning has some advantages, such as working on two-dimensional data directly, less susceptibility to local optimal, and the ability to learn texture features from data [19]. The other advantages of DCNN are transferability connections and sparse connections. The transferability connections are certain layers of network architecture that can reproduce weights for different tasks. Sparse connections are infrequent connections that can reduce redundant connectivity, thereby reducing computing costs [21].

One particular model of a deep learning is convolutional neural network or also called deep convolutional neural network (DCNN). Since 2012, DCNN [17–20] has been led to a series of breakthroughs for image classification [22]. Deep learning-based computer-aided diagnosis for breast cancer [23] and lung cancer [24] has been applied in radiology. In addition, there are many other studies that use DCNN to detect or classify diseases, such as [25] comparing the performance of three CNN models (models VGG19, Resnet50v2 and Densenet201) with X-rays data sets of patients with COVID-19, pneumonia, and tuberculosis with a large number of data sets.

By considering several capabilities and advantages of DCNN, our contributions are:

- We designed and determined the best DCNN configuration model to extract in-depth features from dental periapical radiograph images from multiple image block sizes and multiple convolution layers with varying hyper parameter.
- The result of DCNN architecture or configuration can be used to detect osteoporosis disease.

Measured the effectiveness of the best model by conducting trials using data from previous researchers to achieve state-of-the-art performance for the detection or classification of osteoporosis using dental periapical radiograph.

III. METHOD

This study proposes a deep Convolutional Neural Network architecture as a texture feature of dental periapical radiographs that can be used for osteoporosis detection. An extensive examination of the network is conducted to obtain the optimal network configuration and hyperparameters, which include input image size, number of kernels, filter kernel size, dropout value, and learning rate value. The system results will be compared with results of BMD measurements in the femur and lumbar areas using DXA.

The proposed model is broadly divided into two stages, namely the training stage and the testing stage. Each stage consists of several processes, namely, image acquisition and ROI selection, feature extraction and classification (see Fig. 2).

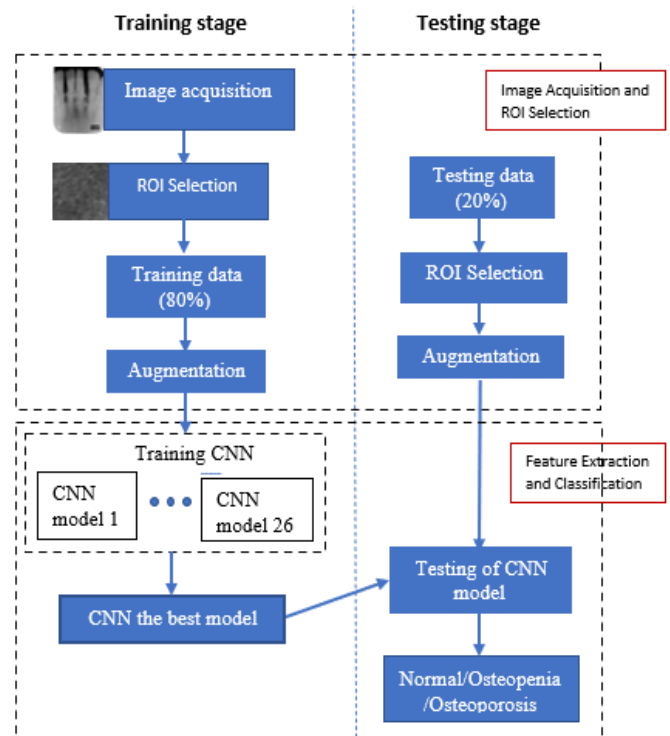


Fig. 2. Proposed Method for Osteoporosis Detection.

A. Image Acquisition and RoIs Selection

We used a dental X-ray device to obtain digital periapical radiographs of mandibular anterior teeth of 31 postmenopausal women aged over 40 years old. This research makes use of the Villa Sistemi device with an electrical specification of 70 kVp/8 mA that uses photostimulable phosphorus plates (PSP) as image receptors. All periapical radiographs were processed from the Radiology Department, Prof. Soedomo Dental Hospital of Universitas Gadjah Mada using the DBSWin4.5 (Dürr Dental, Bietigheim-Bissingen) to produce digital grayscale images in the JPG format. All images are assessed for quality assurance by a dentist (Fig.3).

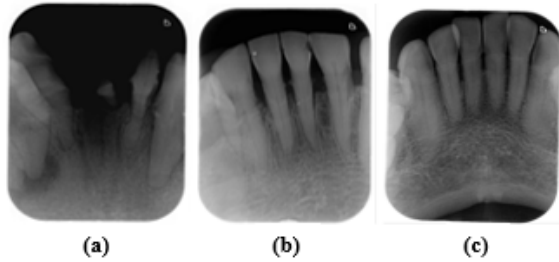


Fig. 3. Periapical Radiograph of the Mandible (a) Normal, (b) Osteopenia, and (c) Osteoporosis.

Regions of interests (RoIs) are then selected semi-manually from the images to obtain the most appropriate parts for further processing. The selection procedure marks the upper left corner of the trabecular area then moves to the lower right to form the maximal square that can be extracted from the images. so that RoIs is obtained with various sizes, at least 250x250 pixels Assuming that trabecular areas' sizes do not vary significantly across people, all the extracted RoIs are resized to a standard size of 250 x 250 pixels. It can be considered as a normalization step favoured by subsequent processes. Fig. 4 shows the RoIs selection process as well as the resized images. The resized RoIs are divided further into overlapping blocks (with 10 pixels overlap), each of which will become an input to a convolutional neural network. This process is called augmentation.

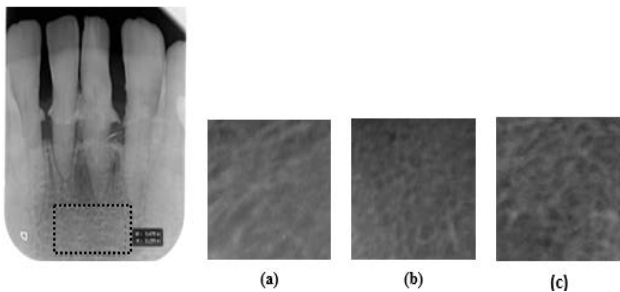


Fig. 4. RoIs Selection and Resizing (a) Normal, (b) Osteopenia and (c) Osteoporosis.

B. Features Extraction and Classification

Feature extraction and classification is performed using a deep convolutional neural network (CNN) which takes image blocks as input and produces a prediction class as output. The prediction class in this study consisted of normal (N), osteopenia (Oa), and osteoporosis (Os) which was a further

decrease in the bone mass of the examined subjects. The deep CNN configuration used is shown in Fig. 5 or details can be seen in Table I. The deep CNN configuration consists of five building blocks, namely convolution layer, activation layer, pooling layer, fully connected layer, and soft-max layer. The first convolution layer uses a kernel size of 5x5, while the second to fifth convolution layers use a 3x3 kernel size. While in the pooling layer, all layers use a 2x2 kernel.

The Deep CNN configuration showed at Fig. 5 and Table I is the best model from the results of experiments that have been carried out on each block size and convolution layer size, activation layer, pooling layer including number of the kernel used (6, 16, 32, 64, 128) and the filter kernel size for each convolution layer as well as several parameters such as learning rate value, dropout value and number of epochs (as shown in Fig. 6-9 and Table II-III).

CNN's themselves are inspired by a neuro-biological process in which connectivity patterns between neurons resemble the visual cortex model [26] and [27]. CNN's work on two-dimensional data of multiple depths and operate in a layer-by-layer order [28].

Convolution layers serve to extract features from an input image (edges, corners, or crosses) using responses to some special character presenting in the input. Activation layers determine "relevant" convolutional kernels. The layers produce stacks of feature maps, each of which parts within the produced feature maps to be used in subsequent processing. Relevant parts, in this case, will be "active" after passing through the activation function, which is the rectified linear units (ReLU).

The ReLU layer is an activation function obtained through the equation.

$$ReLU(x) = \max(0, x) \tag{1}$$

Where x is the input to the neuron and the transfer function is finely approximated to the rectifier into an analytic function.

Pooling layers help the network avoid overfitting by reducing some network parameters and the respective computations. The pooling layers work as a non-linear down-sampling process that divides outputs of activation layers into subregions and collects maximum values from the subregions. From an $n \times n$ input, with n represented of size of image and k represented from size of kernel, a pooling layer will produce an $\left(\frac{n}{k}\right) \times \left(\frac{n}{k}\right)$ output.

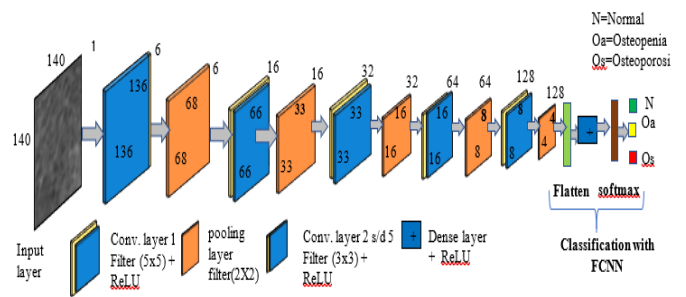


Fig. 5. Architectural Model of CNN for Osteoporosis Detection.

TABLE I. CONFIGURATION MODEL OF CNN FOR OSTEOPOROSIS DETECTION

LAYER TYPE		INPUT SIZE	NUMBER OF KERNEL	FILTER KERNEL SIZE	OUTPUT SHAPE
Input layer	input	140x140x1	-		140x140x1
Conv. layer 1	convolution	140x140x1	6	5x5	136x136x6
	activation	136x136x6	6		136x136x6
	max pooling	136x136x6	6	2x2	68x68x6
Conv. layer 2	convolution	68x68x6	16	3x3	66x66x16
	activation	66x66x16	16		66x66x16
	max pooling	66x66x16	16	2x2	33x33x16
Conv. layer 3	convolution	33x33x16	32	3x3	33x33x32
	activation	33x33x32	32		33x33x32
	max pooling	33x33x32	32	2x2	16x16x32
Conv. layer 4	convolution	16x16x32	64	3x3	16x16x64
	activation	16x16x64	64		16x16x64
	max pooling	16x16x64	64	2x2	8x8x64
Conv. layer 5	convolution	8x8x64	128	3x3	8x8x128
	activation	8x8x128	128		8x8x128
	max pooling	8x8x128	128	2x2	4x4x128
	Dropout 0.5	4x4x128			4x4x128
FCNN	flatten	4x4x128			2048
	dense128	128			128
	activation (ReLU)	128			128
	Dropout 0.5	128			128
	dense3	3			3
output softmax	activation (softmax)	3			3

In CNN's, a convolution layer is normally tied with an activation layer and a pooling layer (showed Fig. 5). This bundle is repeated several times to produce a "thick" stack of down-sampled feature maps at the end of the sequence. Fully connected layers take the vectorized (flattened) form of this

stack and are also tied with some activation layers to produce output vectors. The lengths of these vectors are normally equal to the number of prediction classes. Values within these vectors are converted into probability values by SoftMax layers at the end of CNN's. The output layer present in the last layer of CNN to the normalized exponential function or softmax is a generalization of the logical function of a k-dimensional z vector into a k-dimensional $\sigma(z)$ vector with a real number value between [0, 1]. The SoftMax function is written in the following equation (2):

$$\sigma: R^k \rightarrow [0,1]^K \quad (2)$$

$$\sigma(Z) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \text{ for } j=1, \dots, K$$

where σ is softmax notation symbol, z is a vector of the inputs to the output layer, K is dimensions of vector z, and j is the index of the output unit. Table I shows the specifications of the model configuration.

C. Training of CNNs

To build the proposed system, data are collected from postmenopausal Javanese women aged over 40 years. A total of 31 subjects have agreed to participate in the research and have signed informed consent. Ethical clearance has also been obtained from the ethics and advocacy unit of the Faculty of Dentistry of Universitas Gadjah Mada (UGM) with the number: 0061/KKEP/FGK-UGM/EC/2019. Some criteria are used to exclude subjects from the research. These include suffering from cancer with bone metastases, kidney failure, metabolic diseases (hyperparathyroidism, hypoparathyroidism, osteomalacia, renal osteodystrophy, and osteogenesis imperfecta), and taking drugs that affect bone metabolism.

After dental periapical radiographs are acquired from the Radiology Department, Prof. Soedomo Dental Hospital of Universitas Gadjah Mada, the women went for bone density examination at the Radiology Department of Dr Sardjito Hospital Indonesia. Periapical radiographs used in this research have gone through an assessment process to ensure their quality. As for bone density estimation, a lunar prodigy primo DEXA densitometer (GE Lunar Corporation, Madison, WI, USA) is used to scan the subjects' spine and femur regions at an exposure of 42 μ Gy for 1.27 minutes. Bone density values were converted into T-scores to determine osteoporosis, osteopenia, or normal. These categories were then used as the labels for the collected periapical radiographs. The conversion was conducted using the standard procedure specified by the World Health Organization (WHO). Based on BMD measurements of 13 subjects, three subjects were classified as normal, six subjects were classified as osteopenia, and four subjects were classified as osteoporosis.

Training is carried out by varying sizes of image blocks (as input), numbers of convolution layers, the use of dropout layers, and sizes of kernels. The max function is used in the pooling layers with sizes of kernels of 2x2 and strides of 2. Overlapping blocks are extracted from the collected images and are augmented by applying small random rotation, scaling, and vertical flip. This process produces thousands of image blocks that are further divided into a training set and

test set. The training set contains 80% of the overall data, while the test set contains the remaining 20% of the data. Table II, shows a summary (minimum, average, and maximum) of training and validation accuracy on 26 CNN models from 3 to 6 layers for each image size.

IV. EXPERIMENTS AND RESULT

A. Experiments Models

This section presents the experiments of the proposed model for osteoporosis detection. The first part of the experiment investigates the optimal configurations of CNNs. For this purpose, we evaluate different sizes of image blocks, namely 40x40, 50x50, 60x60, ..., 150x150 pixel. CNN's are built with 3 and 4 convolutional layers, and when the sizes of image blocks are greater than 100x100 pixel, the networks are also built with 5 and 6 convolutional layers. Two sizes of convolution kernels are applied during the experiments, i.e. 3x3 and 5x5. The strides of the convolution kernels are fixed to 2, and padding is used to maintain the inputs' original sizes during the convolution. Besides, in the 5th convolution layer block, a dropout layer of 0.5 is added. To facilitate classification by providing rules for removing or keeping neurons with probability values between 0 and 1, and the value of the learning rate used = 0.0001. See Tables II (A)-(D), Table III and Fig. 6 to Fig. 8 summarizes performance of CNNs with the different configurations.

TABLE II. TRAINING ACCURACY AND VALIDATION ON 26 CNN MODELS FOR EACH BLOCK SIZE

(B) FOUR LAYERS

block size (pixel)	size of CNN layer					
	four layers (consists six models)					
	Minimum		Average		Maximum	
	Train	Valid	Train	Valid	Train	Valid
40x40	0.704	0.633	0.791	0.670	0.916	0.772
50x50	0.886	0.674	0.912	0.517	0.945	0.742
60x60	0.907	0.684	0.941	0.710	0.960	0.735
70x70	0.907	0.684	0.941	0.710	0.960	0.735
80x80	0.896	0.700	0.948	0.736	0.979	0.757
90x90	0.932	0.707	0.957	0.743	0.979	0.782
100x100	0.945	0.697	0.961	0.738	0.979	0.818
110x110	0.934	0.705	0.957	0.738	0.975	0.759
120x120	0.931	0.710	0.954	0.736	0.976	0.763
130x130	0.944	0.751	0.959	0.778	0.981	0.823
140x140	0.942	0.785	0.962	0.885	0.987	0.882
150x150	0.935	0.834	0.957	0.863	0.985	0.892

(A) THREE LAYERS

block size (pixel)	size of CNN layer					
	three layers (consists six models)					
	Minimum		Average		Maximum	
	Train	Valid	Train	Valid	Train	Valid
40x40	0.713	0.647	0.812	0.683	0.895	0.711
50x50	0.794	0.653	0.876	0.698	0.908	0.717
60x60	0.860	0.660	0.897	0.699	0.928	0.730
70x70	0.867	0.663	0.9025	0.701	0.931	0.740
80x80	0.880	0.664	0.913	0.687	0.939	0.713
90x90	0.903	0.668	0.935	0.705	0.961	0.738
100x100	0.911	0.656	0.858	0.679	0.954	0.697
110x110	0.881	0.624	0.858	0.529	0.974	0.723
120x120	0.929	0.712	0.950	0.731	0.962	0.742
130x130	0.914	0.730	0.942	0.746	0.973	0.771
140x140	0.909	0.764	0.951	0.800	0.969	0.816
150x150	0.919	0.798	0.958	0.814	0.977	0.833

(C) FIVE LAYERS

block size (pixel)	Size of CNN Layer					
	five layers (consists seven models)					
	Minimum		Average		Maximum	
	Train	Valid	Train	Valid	Train	Valid
40x40	0.772	0.658	0.863	0.703	0.915	0.725
50x50	0.880	0.668	0.907	0.691	0.930	0.749
60x60	0.923	0.693	0.942	0.723	0.967	0.772
70x70	0.923	0.693	0.946	0.728	0.967	0.798
80x80	0.961	0.744	0.970	0.777	0.981	0.837
90x90	0.962	0.758	0.970	0.782	0.981	0.846
100x100	0.956	0.771	0.973	0.812	0.981	0.866
110x110	0.968	0.742	0.976	0.808	0.986	0.918
120x120	0.959	0.754	0.973	0.813	0.985	0.903
130x130	0.966	0.824	0.976	0.856	0.985	0.89
140x140	0.968	0.856	0.978	0.885	0.998	0.956
150x150	0.925	0.841	0.967	0.889	0.984	0.916

(D) SIX LAYERS

block size (pixel)	Size of CNN Layer					
	six layers (consists seven models)					
	Minimum		Average		Maximum	
	Train	Valid	Train	Valid	Train	Valid
40x40	not available					
50x50						
60x60						
70x70						
80x80						
90x90						
100x100						
110x110						
120x120						
130x130						
140x140	0.769	0.698	0.944	0.816	0.985	0.869
150x150	0.461	0.462	0.902	0.815	0.989	0.918

Based on 26 CNN models that have been trained (see Tables II(A)–(D), it is known that in the experiment with five layers and an image block size of 140x140, the minimum, average, and maximum values are higher than the number of convolution layers and other image blocks (see red text in Table IIC). This means that the configuration of five layers, the image block size of 140x140, and some hyper parameters as in table I are the best models of the 26 existing models. The best CNN model means a model that can differentiate trabecular patterns in the normal class, osteopenia, and osteoporosis.

Fig. 6-8 also shows experimental results in finding the best model.

Fig. 6 is a graph of the image blocks size against the accuracy of each convolutional layers. From Fig. 6, it is known that the larger the image blocks size (input blocks size greater than 100), the accuracy tends to increase, especially on the five-layer and four-layer CNN. Indicates that the image blocks size greater than 100 provides additional information on osteoporosis examination.

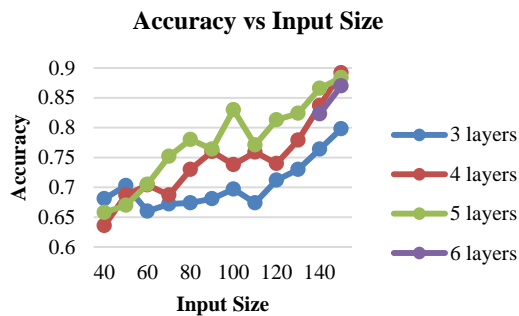


Fig. 6. Graph Accuracy Vs Input Size Graph with Kernel Size 3x3.

In Fig. 7, it is known that a CNN with five layers and an image size greater than and equal to 100 indicates increased accuracy.

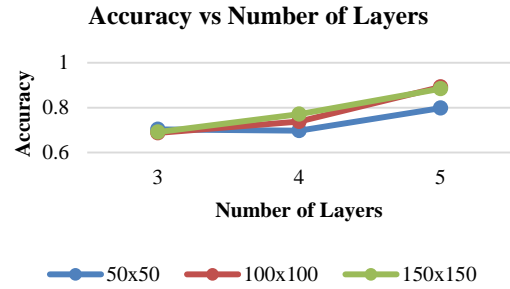


Fig. 7. Graph Accuracy Vs Number of Layers.

Fig. 8 shows the effect of the Dropout (DO) value on accuracy, the DO value = 0.5 has a better training and validation accuracy.

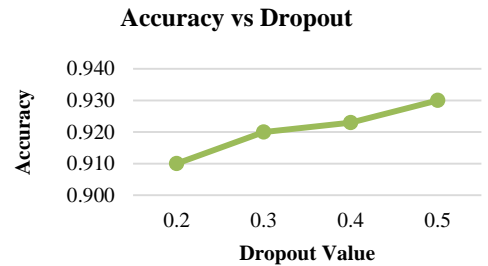


Fig. 8. Graph Accuracy Vs Dropout.

The following Table III shows the experimental results of the best model from the input image size of 40x40 pixels to 150x150 pixels with epoch = 20. And look (highlight) the best of accuracy is at 140x140 pixels.

TABLE III. TRAINING ACCURACY AND VALIDATION ON THE BEST MODELS FOR EACH BLOCK SIZE WITH EPOCH = 20

Block Size	Accuracy		Loss	
	training	validasi	Training	Validation
40x40	0.865	0.713	0.293	0.746
50x50	0.908	0.749	0.207	0.690
60x60	0.929	0.772	0.163	0.722
70x70	0.949	0.798	0.120	0.653
80x80	0.970	0.837	0.072	0.526
90x90	0.970	0.846	0.071	0.462
100x100	0.956	0.866	0.106	0.373
110x110	0.976	0.918	0.057	0.180
120x120	0.974	0.903	0.066	0.224
130x130	0.966	0.890	0.079	0.290
140x140	0.977	0.956	0.056	0.091
150x150	0.973	0.916	0.064	0.227

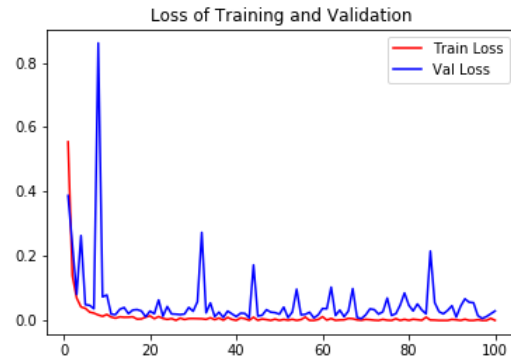
B. Training Models

The total number of training data is 40,755 images, 32,617 images for training, and 4069 images for validations. The dimension of the image input on this model is 140×140 pixels. The batch size is 16, the learning rate value is 0.0001, DO=0.5 and number of epoch = 100. Then the loss function uses Adam optimization.

Training is executed on a computer with specifications processor Intel Core i7-7500U processor specifications, 8 GB RAM, GPU: NVIDIA GeForce GTX 840, Windows 10 operating system, Python 3.7 Programming Language with an editor spyder (python3.7).

Fig. 9(a)-(b) shows the trend accuracy and loss of the training process and the validation of the best models (using architecture and configuration in Fig. 5 or Table I)

Research on osteoporosis examination using dental periapical radiograph images has been carried out with a satisfactory level of accuracy and can represent a computer-aided diagnosis system. Besides, it can enrich the extraction of textural features, which is currently known for the examination of osteoporosis with dental periapical radiograph images, which previously mostly using morphological features.



(b)

Fig. 9. (a) Graph the Accuracy of the Training and Validation Process of the Best Model, (b) Graph the Loss of the Training and Validation Process of the Best Model.

C. Testing Models

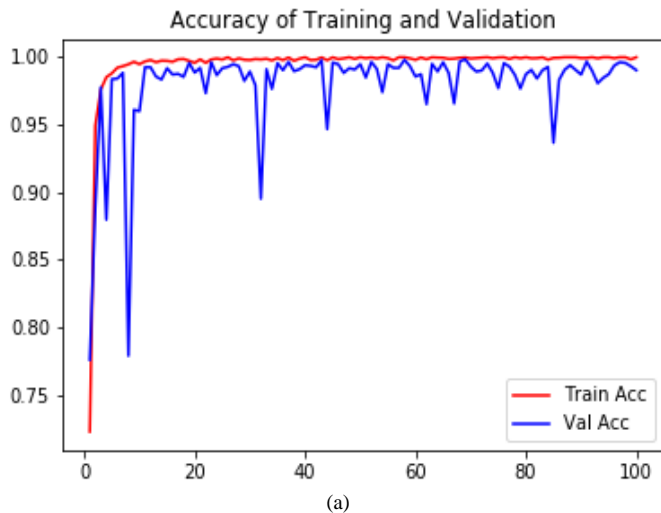
We use 4,069 images as sample test data. Model performance measured using four performance measures parameters, namely Precision, Recall, F1 score and Accuracy. The Precision, Recall, F1 score, and Accuracy values of the best model testing result see Table IV.

TABLE IV. OSTEOPOROSIS – TESTING REPORT

	Precision	Recall	F1 Score
Normal	0.88	0.88	0.88
Osteopenia	0.84	0.98	0.93
Osteoporosis	0.83	1.00	0.91
Accuracy	0.92		
Macro avg	0.97	0.97	0.97
Weighted avg	0.85	0.92	0.90

Table V shows a comparison of the performance of the osteoporosis examination between the proposed methods and those of other previous researchers. In the testing process using datasets from previous researchers [11], the dataset Augmented first so that the amount of data used in the testing process is proportional to the amount of data used in this research.

When compared to previous related work, our method has the highest validation accuracy and testing accuracy, with a validation accuracy of 98.10% and a testing accuracy of 92.50%.



(a)

TABLE V. PERFORMANCE COMPARISON OF EXAMINATION OSTEOPOROSIS

Parameters	Authors		
	<i>Licks etc. (2010) [8]</i>	<i>Sela etc. (2015) [11]</i>	<i>This Research</i>
Subject	Women over 40 years age	Women over 40 years age	Women over 40 years age
Number of training data	60 (22 normal and Osteopenia, 38 osteoporosis)	54 (11 normal, 22 Osteopenia, 21 osteoporosis)	13(3 normal, 6 osteopenia, 4 osteoporosis) augmented: 32,617 (7,527 normal, 15,054 osteopenia, 10,036 Osteoporosis)
Number of testing data	60 (22 normal and Osteopenia, 38 osteoporosis)	54 (11 normal, 22 Osteopenia, 21 osteoporosis) Augmented: 4,069 (939 normal, 1,878 osteopenia 1,252 osteoporosis)	4,069 (939 normal, 1,878 osteopenia 1,252 osteoporosis)
Gold- standard (Classtarget)	Lumbar/Fe-moral BMD DXA (Normal, osteopenia/osteoporosis)	Lumbar/Fe-moral BMD DXA (Normal, osteopenia, osteoporosis)	Lumbar/Fe-moral BMD DXA (Normal, osteopenia, osteoporosis)
Type of Feature extraction	Analysis of morphological (The trabecular area, the periphery, a proportion trabecular length, area, and perimeter (M1-M14)	Analysis morphological (porosity, the size of porosity and orientation of porous)	Texture feature (feature map CNN)
Method of Classification	Classification and Regression Tree Analysis (CART)	C45 (3,6,9-fold validation)	Fully Connected Neural Network (FCNN)
Accuracy	88.33%	86.67%	98.10%
Accuracy from joint of number of testing data [11] and this research	-	-	96.10%
Accuracy from of number testing data [11]	-	-	92.50%

V. CONCLUSION

As shown in Table I, the highest validation accuracy is achieved when the block size is 140x140, the number of convolution layers is 5, and the size of the convolution kernel is 5x5 for the first layer and 3x3 for the other layers. It can then be concluded that the bigger the image block, the higher the validation accuracy. This tells us that bigger images provide additional information that helps discriminate trabecular patterns in normal, osteopenia, and osteoporosis classes. The improvement in accuracy, though, does not change much when the block size is increased from 140x140 to 150x150. This indicates that 140x140 has provided most of the information required by CNNs to distinguish osteoporosis. The training and validation accuracy achieved by the best model is 99.50% and 98.10%, respectively, while the loss of training and validation is 1.30% and 5.40%. Then the testing accuracy is 92.50%.

VI. FUTURE WORK

For this reason, research on osteoporosis examination using dental periapical radiograph images can continue to be carried out, considering that research for osteoporosis examination using dental periapical radiograph images is still rarely used compared to dental panoramic images. This study can be developed by adding a process to increase the

resolution of dental periapical radiographs that tend to be a low resolution at the pre-processing stage and applying the automatic ROI selection method [29].

Further, it can also increase the number of data collections for normal and osteoporosis classes and can use variations in the image of the trabecular bone area of the left and right posterior mandibles.

ACKNOWLEDGMENT

The authors would like to thank the Universitas Gadjah Mada and the Universitas Bumigora Mataram for funding this research. The author also would thank the Radiology Department, Prof. Soedomo Dental Hospital of Universitas Gadjah Mada and the Radiology Department of Dr Sardjito Hospital Indonesia. for supporting this research in acquisition image.

REFERENCES

- [1] M. L. Brandi, "Microarchitecture, the key to bone quality," *Rheumatology*, vol. 48, no. suppl 4, pp. iv3-iv8, Oct. 2009.
- [2] J. E. Adams, "Dual-energy X-ray absorptiometry," *Med Radiol Diagnostic imaging Radiat Oncol.*, vol. 2, pp. 105-124, 2008.
- [3] P. Watanabe et al., "Morphodigital study of the mandibular trabecular bone in panoramic radiographs," *Int. J. Morphol.*, vol. 25(4), no. 4, pp. 875-880, 2007.

- [4] G. D. Elia, G. Caracchini, and G. D. Elia, "P22461252.pdf," Bone Fragility imaging Tech. Clin. Cases Miner. Bone Metab., vol. 6, no. 3, pp. 234–246, 2009.
- [5] K. Horner and H. Devlin, "The relationship between mandibular bone mineral density and panoramic radiographic measurements.," J. Dent., vol. 26, no. 4, pp. 337–343, 1998.
- [6] A. F. Leite, P. T. de S. Figueiredo, C. M. Guia, N. S. Melo, and A. P. de Paula, "Correlations between seven panoramic radiomorphometric indices and bone mineral density in postmenopausal women," Oral Surgery, Oral Med. Oral Pathol. Oral Radiol. Endodontology, vol. 109, no. 3, pp. 449–456, 2010.
- [7] E. Whaites, Essential of dental radiography and radiology, 4th ed. London: Churchill Livingstone Elsevier, 2007.
- [8] R. Licks, V. Licks, F. Ourique, H. R. Bittencourt, and V. Fontanella, "Development of a prediction tool for low bone mass based on clinical data and periapical radiography," Dentomaxillofacial Radiol., vol. 39, no. 4, pp. 224–230, 2010.
- [9] W. G. M. Geraets et al., "Selecting regions of interest on intraoral radiographs for the prediction of bone mineral density," Dentomaxillofacial Radiol., vol. 37, no. 7, pp. 375–379, Oct. 2008.
- [10] W. G. M. Geraets, C. Lindh, and H. Verheij, "Sparseness of the trabecular pattern on dental radiographs: Visual assessment compared with semi-automated measurements," Br. J. Radiol., vol. 85, no. 1016, pp. 455–460, 2012.
- [11] E. I. Sela, S. Hartati, A. Harjoko, R. Wardoyo, and M. Mudjosemedi, "Feature Selection of the Combination of Porous Trabecular with Anthropometric Features for Osteoporosis Screening," Int. J. Electr. Comput. Eng., vol. 5, no. 1, pp. 78–83, 2015.
- [12] E. I. Sela and R. Widyaningrum, "Osteoporosis Detection using Important Shape-Based Features of the Porous Trabecular Bone on the Dental X-Ray Images," Int. J. Adv. Comput. Sci. Appl., vol. 6, no. 9, pp. 247–250, 2015.
- [13] E. I. Sela and R. Pulungan, "Osteoporosis identification based on the validated trabecular area on digital dental radiographic images," Procedia Comput. Sci., vol. 157, pp. 282–289, 2019.
- [14] S. F. Diba, R. S. Gracea, D. Rurie Ratna Shantiningsih, and Khasnur Hidjah, "Analysis of mandible trabecular structure using digital periapical radiographs to assess low bone quality in postmenopausal women," Saudi Dent. J., 2021.
- [15] M. M. S. Enny Itje Sela, Sri Hartati, Agus Harjoko Retantyo Wardoyo, "Segmentation on the Dental Periapical X-Ray Images for Osteoporosis Screening," IJACSA, vol. 4, No. 7, p. 158, 2013.
- [16] P. L. Lin, H. C. Hsu, P. Y. Huang, P. W. Huang, and P. Chen, "Alveolar bone-loss area localization in periapical radiographs by texture analysis based on fbm model and GLC matrix," 2014 IEEE Int. Symp. Bioelectron. Bioinformatics, IEEE ISBB 2014, 2014.
- [17] P. Huang, P. Huang, P. Lin, and H. Hsu, "Alveolar bone-loss area detection in periodontitis radiographs using hybrid of intensity and texture analyzed based on fbm model," pp. 13–16, 2014.
- [18] N. Tajbakhsh et al., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1299–1312, May 2016.
- [19] Bakhshwain and A. Sagheer, "Online Tuning of Hyperparameters in Deep LSTM for Time Series Applications," Int. J. Intell. Eng. Syst., vol. 14, no. 1, pp. 212–220, 2020.
- [20] [M. E. Wibowo, "Rancangan Pelatihan Paralel Jaringan Saraf Deep Learning Berbasis Map-Reduce," in Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9 ISSN (Printed) : 2579-7271 Fakultas Sains dan Teknologi, UIN Sultan Syarif Kasim.
- [21] D. D. Himabindu, "A Survey on Computer Vision Architectures for Large Scale Image Classification using Deep Learning," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 10, pp. 105–120, 2021.
- [22] K. He, X. Zhang, S. Ren, and Jian Sun, "Deep residual learning for image recognition," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 770–778, 2016.
- [23] J. Z. Cheng et al., "Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans," Sci. Rep., vol. 6, no. October 2015, pp. 1–13, 2016.
- [24] K. L. Hua, C. H. Hsu, S. C. Hidayati, W. H. Cheng, and Y. J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," Onco. Targets. Ther., vol. 8, pp. 2015–2022, 2015.
- [25] L. Mangeri, "Chest Diseases Prediction from X-ray Images using CNN Models: A Study," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 10, pp. 236–243, 2021.
- [26] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
- [27] M. A. Nielsen, "Neural_Networks_and_Deep_Learning_Pdf.Pdf," Determ. Press, 2015.
- [28] D. Stathakis, "How many hidden layers and nodes?," Int. J. Remote Sens., vol. 30, no. 8, pp. 2133–2147, 2009.
- [29] E. I. Sela, R. Pulungan, R. Widyaningrum, and R. R. Shantiningsih, "Method for automated selection of the trabecular area in digital periapical radiographic images using morphological operations," Healthc. Inform. Res., vol. 25, no. 3, pp. 193–200, 2019.