Law Texture Analysis for the Detection of Osteoporosis of Lumbar Spine (L1-L4) X-ray Images Using Convolutional Neural Networks

Kavita Avinash Patil, K V Mahendra Prashanth, A Ramalingaiah

ABSTRACT-An early detection of osteoporosis can improve life expectancy and quality of life. Vitamin deficiency, tissue loss, and hormonal changes may cause osteoporosis. Brittle and fragile bones are symptoms of this condition. It is essential to diagnose osteoporosis at the earliest possible stage to prevent bone risk factors. Dual-energy X-ray absorptiometry (DEXA) can detect osteoporosis more accurate in recent studies. DEXA can detect osteoporosis based on bone mineral density (BMD) and statistical measurements. In this work, the spine X-ray image of the inner trabecular bone microarchitecture is used to identify the normal or abnormal lumbar vertebrae L1-L4. The Law texture algorithm is used to detect osteoporosis in an image using a CNN classifier. Statistical measures of Law texture feature-based X-ray scores are developed to distinguish normal or abnormal bone. The proposed algorithm achieves 100% accuracy, sensitivity, and specificity using the correct fit model with 4-fold cross-validation. As compared to existing techniques, this algorithm is much simpler and more effective.

Index Terms **- CNN, DEXA, lumbar spine, osteoporosis, X-ray image**

I. INTRODUCTION

Osteoporosis is caused by low bone density, which results in fragile bones that are more prone to fracture. This is a silent disease, that can progress over an extended period of time. An early detection of the disease can prevent bone fracture, it often remains undetected until there is a fracture [1]. Fracture due to bone disease affect around 1.5 million people per year [2].Femur, wrist, heel, and spine are the most common osteoporosis-related fracture. Quantitative computed tomography (QCT) and Dual-energy x-ray absorption (DEXA) measurements are used to assess bone mineral density to diagnose osteoporosis [3].Using lateral view X-ray images of the lumbar spine (L1-L4), the trabecular bone micro-architecture (TBMa) features are considered as a new technique to detect osteoporosis using convolution neural network (CNN). With the advancement of deep learning techniques such as artificial - NN, CNN, and evaluation of the neural network based medical image classification has become the key technique for computer aided diagnosis [4].In this paper, the proposed work is divided into two parts: the first part involves pre-processing of X-ray images of lumbar spine L1-L4 within a region of interest (ROI). As a pre-processing step, remove the noise and enhance the texture features of L1-L4, in accordance with [5] to obtain high quality images. The second part involves classification of bones based on Law texture features of trabecular bone (TB) with CNN as a classifier. Here, TB of L1-L4 images is represented using a Law texture model (LTM) in order to get the fine texture of edges, bobs, lines, ripples. Finally, simple CNN model is used to classify the input image into normal (control cases) or abnormal (osteoporosis). In order to distinguish between normal or abnormal, the statistical values are calculated as an X-ray score for L1-L4.

Manuscript received April 11, 2022; revised November 22, 2022.This research was supported by Sri Jagadguru Balagangadhranatha Institute of Technology, affiliated to Visvesvaraya Technological University, Bangalore, India.

Kavita Avinash Patil is a Research Scholar, Department of Electronics and Communication Engineering, Sri Jagadguru Balagangadhranatha Institute of Technology, affiliated to Visvesvaraya Technological University, Bangalore, India (corresponding author, phone:+91 9731605296;(e-mail: kavitamalagatti@ gmail.com).

K. V. Mahendra Prashanth is Professor in Department of Electronics and Communication Engineering, Sri Jagadguru Balagangadhranatha Institute of Technology, Affiliated to Visvesvaraya Technological University, Bangalore, India (e-mail: kvmprashanth@sjbit.edu.in).

Dr. A Ramalingaiah is an Orthopedic surgeon, researcher in osteoporosis field and discovering drugs to treat osteoporosis at Abhilasha Orthopaedic Hospital, Bangalore, India (e-mail: armlgh.2011@gmail.com).

II. ASSOCIATED WORK

The validity of the model was examined in a study of the Osteoporosis Self-Assessment Tool (OST) as a screening criterion for osteoporosis compared with DEXA [6].The Bone mineral index (BMI) and OST score for the left femoral neck were calculated from the DEXA report. A DEXA measurement was used to determine bone mineral density [7][8]. By using Asian and Caucasian T-scores, osteopenia (OPe) and osteoporosis (OP) prevalence was calculated for Lumbar spine (LS)[9],Femoral neck (FN), and Total hip (TH).As a result, using a Caucasian rather than Asian reference range will lead to a higher diagnosis of OP but it led to overtreatment. A DEXA scan is performed on each subject [10]. Based on bone density covariate data from a single site or combined sites, likelihood ratio tests were used to evaluate improvements in fracture prediction from Cox hazards models. If non-vertebral fracture were considered along with spine fracture, the LS ceased to be useful for fracture assessment. DEXA scan images are segmented using algorithms such as k-mean clustering and mean-shifting to detect osteoporosis [11]. Furthermore, a mathematical analysis of Bone mineral density (BMD) values is proposed to detect osteoporosis accurate by using a new parameter S, which is compared with world health organization (WHO) standard. A study demonstrated that discrepancies between QCT and DEXA are common when diagnosing osteoporosis and that aortic calcification obstructs bone mineral density measurements of the LS by DEXA [12]. BMD measurements are more sensitive with QCT in Chinese elder men. This study explores whether fat fraction was related to BMD in vertebrae of healthy children[13]. Based on the revealed relationship, single-voxel proton magnetic resonance imaging can be viewed as an alternative to QCT and DEXA for osteoporosis detection without radiation exposure. Through the application of a non-local mean filter is used to eliminate the noise, machine

learning pixel label random forest and finding regions of interest, a computer-aided osteoporosis detection model uses DEXA images as input to de-noise and segment more accurate [14]. As a result, DEXA is able to diagnose more accurate and improve their performance. A CNN-identified vertebral fracture assessment image from DEXA can predict clinical fracture results with high accuracy [15].Assessing BMD in L-spinal-stenosis-surgery(LSS) patients and FN using DEXA, Axial-posterior, lateral-mid and mid-lateral BMD projections were assessed [16]. Fracture risk assessment (FRAX) tool questionnaires were filled out by LSS patients. When it comes to identifying osteoporosis patients in preoperative planning, Spinal DEXA measurements cannot be substituted by FRAX evaluation. The diagnosis of osteoporosis based on the lowest T-score from multiple BMD measurement may increase the prevalence of the condition, supporting the increased sensitivity of multiple measurement [17]. While QCT-BMD analysis (CTXA) hip areal-BMD may indicate osteoporosis more accurate, QCT spine volumetric-BMD may be more sensitive[18]. An analysis of incidental findings and the reliability of images is performed, deep learning could enhance the established DEXA method for osteoporosis diagnostics, and theoretically improve its predictive ability [19]. CNN-based method for detecting BMD qualitative analyses of CT images [20] and it can be used for osteoporosis screening in clinical settings .A gemstone-spectral-image energy spectrum CT medical image is a form of MRI that is based on deep convolutional neural network technology for osteoporosis and degenerative lesions in the lumbar spine. In terms of assessing BMD, it is more accurate than DEXA [21].The Bone Ultrasonic Scanner provides good results in assessing osteoporosis more effective in postmenopausal women [22].

III. METHODOLOGY

The proposed system work flow is carried out in the sequence of pre-processing, developing Law texture energy features of images, CNN classifier and the X-ray scores as a statistical measurement to identify normal or abnormal bone.

A. Pre-processing

The pre-processing allows to suppress undesirable distortions and enhance some desirable features. In this work, X-rays of the lumbar spine from L1 to L4 were pre-processed by combining techniques such as de-noising and enhancement [5] [23]. Using two-phase principal component analysis (PCA) noise is removed as well as sharp edges are smoothened in the image. The hybrid mean filter (HMF) is employed after PCA to get rid from the noise, which is undesirable residuals at the edges [24]. With HMF, lines and edges are protected using a linear window technique. Additionally, range filters and edge sharpening improve TB texture at the edges.

B. Law Texture Energy Features

The Law texture energy features [25] of an image can be described as a function of the spatial variation in brightness intensity. In most cases, textures are used to describe the structure of an image. The goal of this paper is to visualize the fine texture structure, i.e. trabecular bone micro-architecture of lumbar spine L1-L4 images using a one-dimensional spatial basis symmetric vectors [26] such as level $\mathbf{L}_3$ and spot $\mathbf{S}_3$.These vectors produces high center pixel level and edge vectors $\mathbf{E}_3$, which is an anti-symmetric. The basis vectors are described as 1×3 in (1):

$$\mathbf{L}_3=[1,2,1],\ \mathbf{E}_3=[-1,0,1],\ \mathbf{S}_3=[-1,2,-1] \tag{1}$$

The four, one-dimensional vectors with a size 1×5 are generated via convolution. The convolved vectors $\mathbf{L}_L\ \mathbf{L}_S, \mathbf{S}_S$ are symmetric but $\mathbf{L}_E$ is an anti-symmetric. These are represented as in (2) :

$$\begin{aligned}&\mathbf{L}_L=\mathbf{L}_3*\mathbf{L}_3,\ \mathbf{L}_E=\mathbf{L}_3*\mathbf{E}_3,\\ &\mathbf{L}_S=\mathbf{L}_3*\mathbf{S}_3,\ \mathbf{S}_S=\mathbf{S}_3*\mathbf{S}_3\end{aligned} \tag{2}$$

The above mentioned four one-dimensional vectors in (2) are essential part in this work to create 4 two dimensional 5×5 masks or filters are mentioned in (3). These masks are used to determine the Law texture energy features of X-ray spine image of L1-L4.

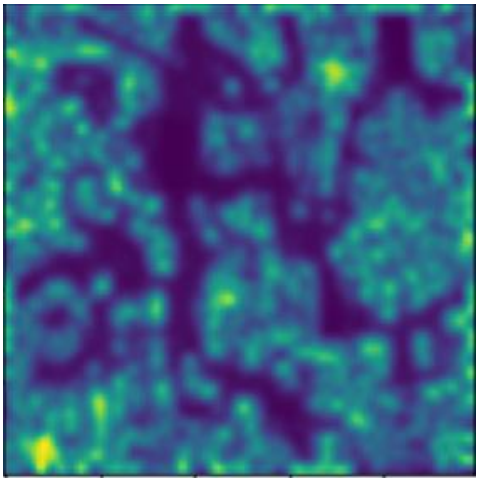
$$\begin{aligned}&\mathbf{L}_{LL}=\mathbf{L}_L^{'}\mathbf{L}_L,\mathbf{L}_{EL}=\mathbf{L}_E^{'}\mathbf{L}_L,\\ &\mathbf{L}_{SL}=\mathbf{L}_S^{'}\mathbf{L}_L,\mathbf{S}_{SL}=\mathbf{S}_S^{'}\mathbf{L}_L\end{aligned} \tag{3}$$

The mask $\mathbf{L}_{LL}$ is used only to sharp the levels , the $\mathbf{L}_{EL}$ mask is applied to sharp the edges along with levels , $\mathbf{L}_{SL}$ mask to sharp the spots along with levels , and $\mathbf{S}_{SL}$ sharp only spots within levels. These four masks are helpful to enhance the texture features in order to analyze the trabecular bone micro architecture of L1-L4 are mentioned in (4):

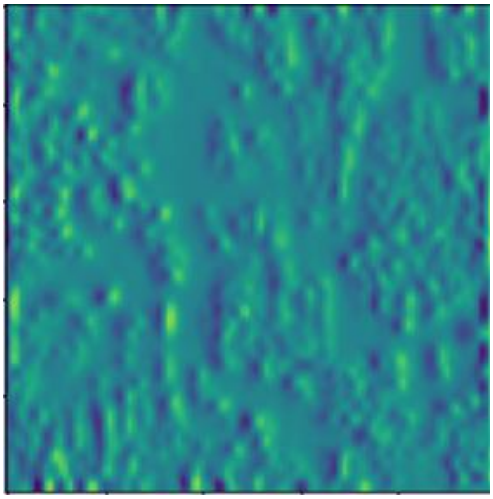
$$\begin{aligned}&\mathbf{F}_1=\mathbf{L}_{LL}*I_s,\mathbf{F}_2=\mathbf{L}_{EL}*I_s,\\ &\mathbf{F}_3=\mathbf{L}_{SL}*I_s,\quad \mathbf{F}_4=\mathbf{S}_{SL}*I_s\end{aligned} \tag{4}$$

Whereas I_s represents the input images of L1-L4 and filtered images $\mathbf{F}_1$ called as sharpen at levels $\mathbf{F}_2$ is sharpen at edges, $\mathbf{F}_3$ is sharpen at spots and $\mathbf{F}_4$ is sharpen at ridges the filtered images. With this setting, the center elements are given the most weight, while pixels near the edges of the mask are given less weight. By visualizing the image block in this way, surrounding pixels and the center are enhanced. The weight function that falls off along block edges are used to reduce the sharp transitions. In addition to this, weighted windows are also used to reduce boundary overlap, but sometimes it results even unpredictable texture measures when a mask covers more than one block texture region. In order to reduce contrast at the edges, masks are rotated at different angles. This makes the texture measures invariant to luminance shifts since the coefficients sum are equal to zero. Finally, the sum of all four filtered output images are used as an input to the CNN.

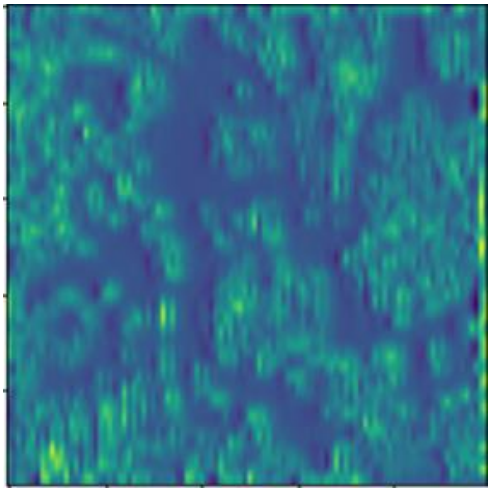
(a) Input normal image



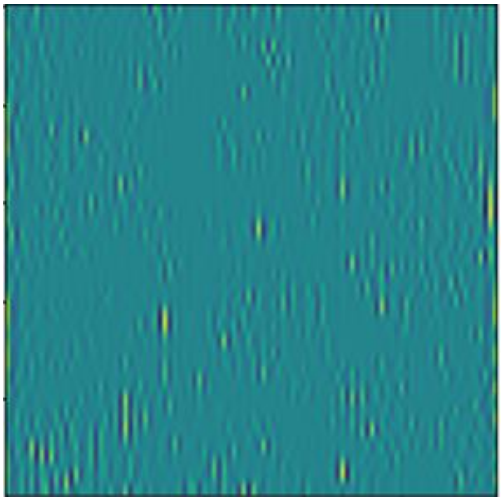
(b) Sharpen at levels



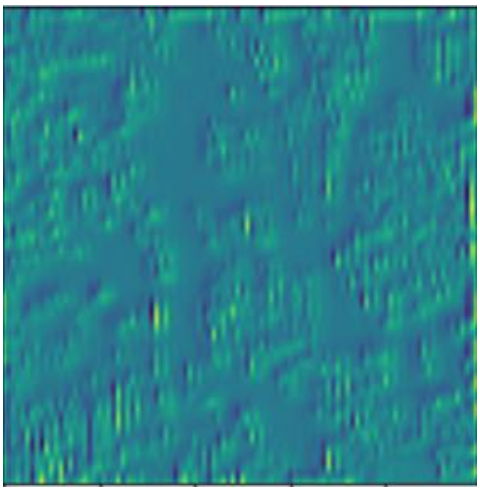
(c) Sharpen at edges



(d) Sharpen at spots



(e) Sharpen at ridges



(f) Sum image

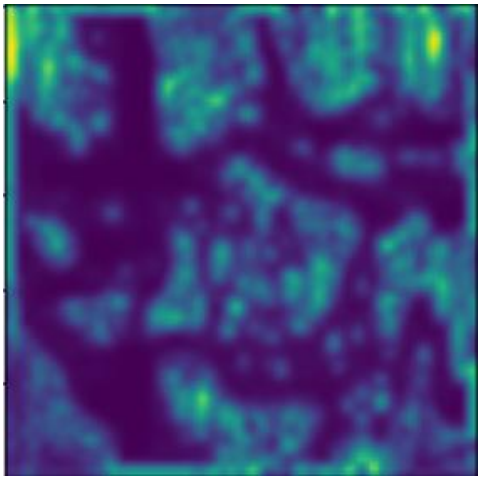
Fig. 1: (a) Input normal image and (b), (c), (d), and (e) are filtered images with respective masks, (f) Sum image.

The sum represents the all four Law texture feature image $(\mathbf{S}_{LTF})$, calculated in (5) for classification of normal or abnormal image. This sum image containing the significant texture information of the input image, which is very much helpful to improve the system performance.

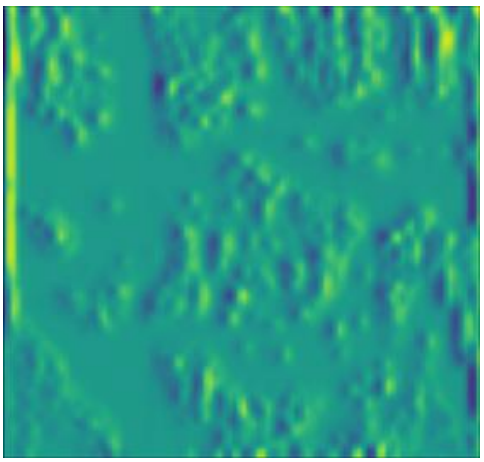
$$\mathbf{S}_{LTF} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4 \tag{5}$$

The Fig.1shows the Law texture energy features of normal image (a) input I_s (b) $\mathbf{F}_1$ (c) $\mathbf{F}_2$ (d) $\mathbf{F}_3$ (e) $\mathbf{F}_4$ (f) $\mathbf{S}_{LTF}$.

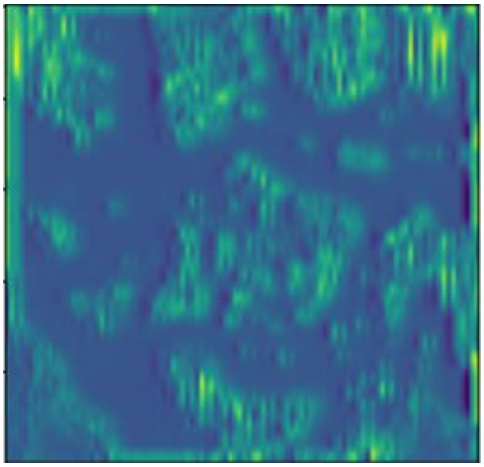
(a) Input abnormal image



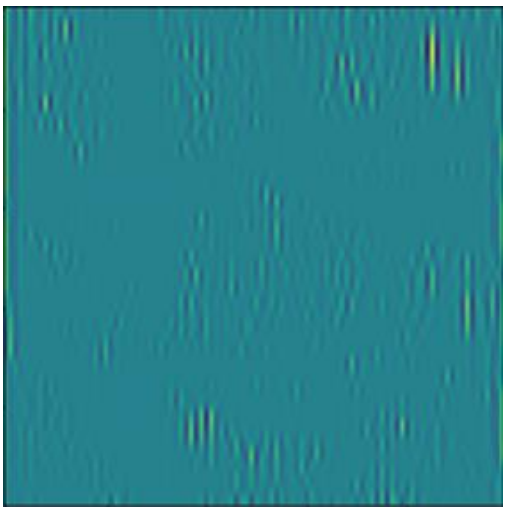
(b) Sharpen at levels



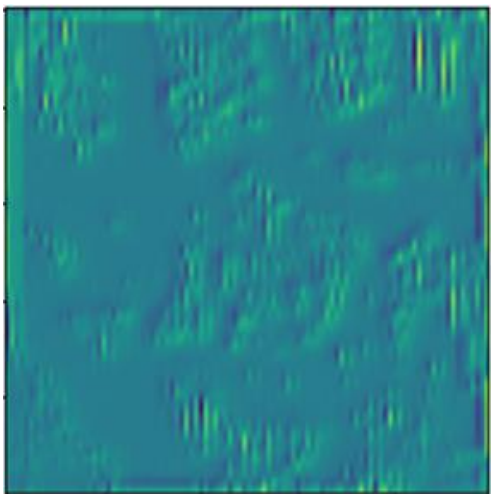
(c) Sharpen at edges



(d) Sharpen at spots



(e) Sharpen at ridges



(f) Sum image

Fig. 2: (a) Input abnormal image and (b),(c),(d), and (e) are filtered images with respective masks, (f) Sum image

Similarly, Fig. 2 shows the Law texture energy features for abnormal image. Analyzing micro-architecture of L1-L4 image texture can help to distinguish normal or abnormal bone.The image has a larger gap between the microstructure in abnormal than in the normal .The trabecular bone does not have a uniform density or microstructure in abnormal and throughout the bone, the density is heterogeneous.

C. Architecture of CNN

Using complex algorithms and artificial neural networks, deep learning has emerged as a useful tool for analyzing images in the current scenario and the systems are trained to classify images as like human brains learn from its experience. Numerous types of artificial neural networks used for image classification [27]. CNN is a one type of deep learning network. CNNs are very successful in deep learning as they have a variety of functions that make them a significant part of emerging and rapidly growing fields. Most neural networks consist of an input layer, a hidden layer, and an output layer. The model visualization is as shown in Fig.3, the network is a very simple and requires less computation time to execute it. The input pixels are provided as arrays to the input layer. Multiple hidden layers in CNNs could be used to compute the features from images by performing different operations. This CNN model is a feed-forward network as data flows from its input to its output in a one-directional manner. Many different types of CNNs have been developed; however, in general, they are made up of layers of convolution and pooling or sub-sampling, which are organized into modules. As in a normal feed-forward neural network [28], either a single or multiple layers are fully connected.

Since CNNs can detect patterns and make sense of them, and the best way to classify the images using these architectures. They are well suited for real-world scenarios, since they produce good results and can accurately locate and identify the patterns in images. The capability of CNNs to achieve spatial invariance is a critical characteristic. This means that they can recognize and extract image features regardless of where they are in the image. As a result, CNN is a powerful tool when it comes to deep learning. In a recent paper published in neurological computation [29], feature maps have been pooled so that feature map resolution in spatial domain can be reduced and improved spatial invariance due to input distortions. In addition to reducing the number of parameters required for processing an image, the pooling layer also reduces the amount of memory used and the cost involved as well. Although CNNs are frequently [30] used in image analysis, they can also be used to analyze and classify image. Fig. 3 shows that, there are total 11 layers and 39,257 parameters which are the sum of the trainable parameters (39,193) and the non-trainable parameters (64).

IV. PERFORMANCE EVALUATION MEASURES

As machine learning and statistics progress, the people still get confused about true positives and false negatives, which is why one of the measures used to evaluate system performance is called the confusion matrix. It is simply a matter of categorizing things as positive or negative, for instance abnormal or normal. First class considers positives, while second class considers negatives. In general, it depends on the objective of the study whether it take one as positive or negative. Depending on the perspective, it can appear abnormal (positive) or normal (negative). When sample data is obtained from the dataset and modeling is used to predict its classification or label. When a model identified a class, it would be called true, whereas a model could not identify the class would be called false.

Confusion matrices represent predictions on classification problems in a [31] [32]. Below is a confusion matrix discussed in experimental results. In the matrix, False Negative (FN) and False Positive (FP) change position based on the predicted and actual values, but True Positive (TP) and True Negative (TN) remain in the same position. In the dataset, a positive P represents the number of real abnormal , while a negative N defined as the number of normal images in dataset as mentioned in (6):

$$N = FP + TN\,, \quad P = TP + FN \tag{6}$$

A model's accuracy, precision, recall, and F1 score can be calculated from the confusion matrix. The performance of the model can be understood through these measures. A model's accuracy (A) can be expressed as the ratio between the total number of subjects and the number that the model classified correctly and is calculated in (7):

$$A = \frac{TP + TN}{P + N} \tag{7}$$

In terms of precision (Pr), refers to the proportion of positives identified correctly by the model compared to the total number of positives subjects as mentioned in (8):

$$Pr = \frac{TP}{TP + FP} \tag{8}$$

An accurate bone condition test is described by its sensitivity and specificity mathematically. In the case of individuals who are satisfied with the condition, it is considered a positive , while in the case where it is not satisfied, it is considered a negative. An indicator of sensitivity of the system is the probability of true positives (TP) rate of a positive test provided that it is abnormal as in (9). Specificity is the true negative (TN) rate measures probability of negative test given that results are truly normal as in (10):

S and Sp are the sensitivity and specificity respectively.

$$S = \frac{TP}{TP + FN} \tag{9}$$

$$Sp = \frac{TN}{FP + TN} \tag{10}$$

Recall (R) in (11) refers to the system predicted positives out of total positives.

$$R = \frac{TP}{TP + FN} \tag{11}$$

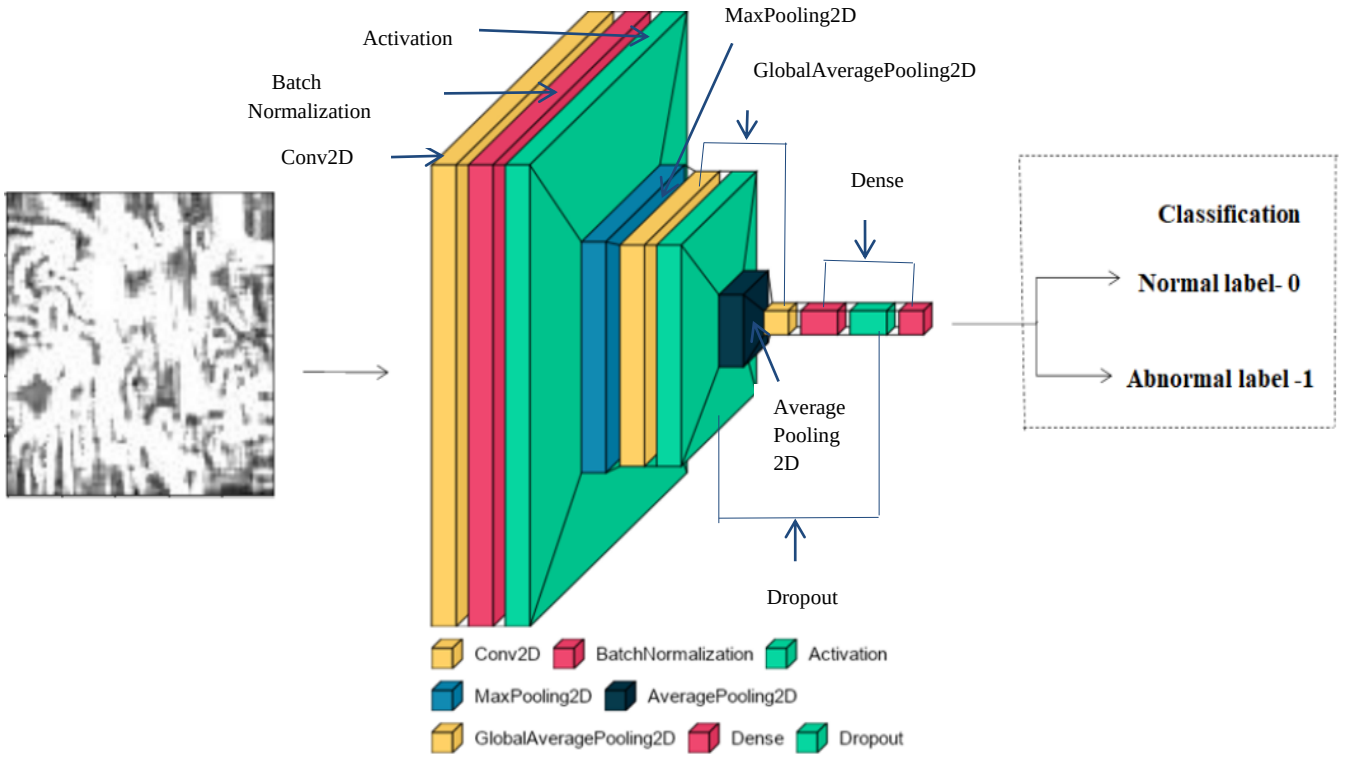


Fig. 3: Model visualization

An accurate test is measured by the F-score in statistical analysis of binary classification. Input image precision and recall are used to calculate this value. In F1-score , precision and recall are summed to get the harmonic mean as illustrated in (12). The F1-score lowering overall scores if

$$F1 = \frac{2TP}{2TP + FN + FP} \qquad (12)$$

precision or recall are very small. In this way, it contributes to balancing the two metrics in order to maintain the system.

V. IMAGE DATASETS AND DEXA-REPORTS

A direct digital X-ray prototype was used to obtain images of Lumbar Vertebrae 0.812 m was the focal distance. All patients received X-rays at 75-80 kV and 80 mAs. Dr. A Ramalingaiah (ortho) Orthopedic Doctor, Bangalore, supported for datasets and the corresponding DEXA reports. The Abhilasha Orthopedic Hospital in Banashankari 3rd Stage Bangalore is located at No.271, 5th Block 100ft Road. There are 162 JPEG-format, 2D radiographic X-ray images of 76 (normal) and 86 (osteoporosis or abnormal). Additionally, DEXA reports will be provided. of spine were collected for each person. Every DEXA report describes the statistical analysis status of the lumbar spine L1-L4 of each person. Table I offers X-ray image databases for various ROIs on the lateral view spine (L1-L4) i.e., for every person four ROI images which are compared with the DEXA results in this study.

Total number of ROI samples L1-L4 are 304 + 344 =648 including normal and abnormal. The size of the L1-L4 samples is mention in the Table I and each vertebra of image samples are 648/4 = 162, i.e., The total number is divided into four groups since the samples considered in this study are four lumbar spines L1-L4. The ROI sample of L1-L4 image datasets is prepared by considering each sample group separate, in order to achieve reliable results experimentally.

TABLE I
DATA BASE DESCRIPTION

Contents	Normal	Abnormal
LS-Spine (162=76+86)	76	86
DEXA report	76	86
Age-Range	35-80	35-60
ROI image size	100×100 (L1) 98×98 (L2) 104×104 (L3) 104×104 (L4)	100×100 (L1) 98×98 (L2) 104×104 (L3) 104×104 (L4)
Total number of samples (L1-L4)	76×4= 304	86×4=344

VI. EXPERIMENTAL PERFORMANCE ANALYSIS

A. Training process

In CNNs, certain patterns of input data are used to train the system. When a number of samples have been processed,

the model gets updated. This is referred to as the sample batch size. In the training dataset, the number of epochs in machine learning is the number of passes. Batches are typically equal to one and can be greater than or less than the sample number of the training dataset. Epochs [33] are typically integer values between one and a finite value in neural networks. This enables the algorithm to run for any duration. The batch size and epochs in the learning algorithm are hyper parameters of a training model associated with integers. As these values are not internal parameters of the model, they cannot be learned from a learning process; they must be specified when a new algorithm is trained on a training dataset. Similarly, these numbers are not fixed values and need to be tried in various integer values before finding the best one. The ROI images of L1 data, consider the following example. The 4-fold partitioning of the 162 images are considered for system process (3-partition of data for training and 1-partition for testing i.e., $162/4 \approx 40$). The 162-40=122 images are utilized for training process and remaining 40 images for testing. The epochs in the system are considered as 500, assume a dataset of 122 samples (i.e., the row-wise) and a batch size of 32.With 32 samples in each of the 4 batches, the weights of the model are updated as each batch of 32 samples is passed. In addition to this, one epoch requires the model to be updated four times. In this case, one epoch in machine learning requires four batches. Furthermore, since the epoch number is 500, the model itself runs 500 times. A model with four batches or updates means that (4×500) 2000 batches are used in the training process. An experimental study of the training process evaluated two steps of processing: curves in the learning process and analysis of model behavior.

1) Curves in learning process

In the training process, it is common for dual learning curves to be generated both on the training and validation datasets of a machine learning model. The model can also be optimized according to the cross-entropy loss and its performance can also be determined with the classification accuracy, as in the case of classification predictive modeling problems. The learning curves for each metric are plotted, with each plot displaying one for training data set and one for validation dataset.

- Learning curves optimization: Model parameters can be optimized by the learning curves are calculated on the loss metrics and accuracy metrics.
- Learning curve for training dataset: Measures the model learning rate based on its training dataset.
- Learning curve for validation: The learning curve is derived from a hold-out validation dataset and it shows that, how the model can be made effective.

2) Analyzing the behavior of a model

System models can be investigated in order to determine their nature and thereby improve the learning performance, the dynamics of a learning curve can be used in three scenarios: a) Under-fitting(UF), b) Correct-fitting(CF) c) Over-fitting(OF), which are used to examine the learning curves. This three-step assessment is necessary during the training process to determine the overall system performance, while testing the model with unseen input images.

a) The model enters under-fitting if the training loss is less than the validation loss.

In this work the model enters into this mode of action when the epoch number (e_no) is varying from $100 \le$ e_no ≤ 500. The Fig. 4 shows training system loss and its accuracy when e_no=100.The training data =122 is again divided into training set and validation set. During training, an unbiased evaluation of the model is provided by the validation data set, as well as ability to update the model hyper parameters. Here, a Table III records the corresponding loss and accuracy for various validation split. Taking an example, 0.2 % of validation split for L1 (training data 122, $122\times0.2\approx24$) set is 98 (122-24) for the training dataset, and 24 for the validation set.

In this Fig. 4, it is evident that the training and validation losses are more for a smaller number of epochs as in (a), while epochs increase, the losses decrease. However, training accuracy of the model increases with an increase in epochs as in (b). In spite of the epochs being 100, both curves do not coincide, i.e., there is an unequal distribution of losses and system is poor validation. As a result of this, the system indicates that trained model is still under fitting for testing.

b) Correct-fitting occurs when the validation loss equals the training loss.

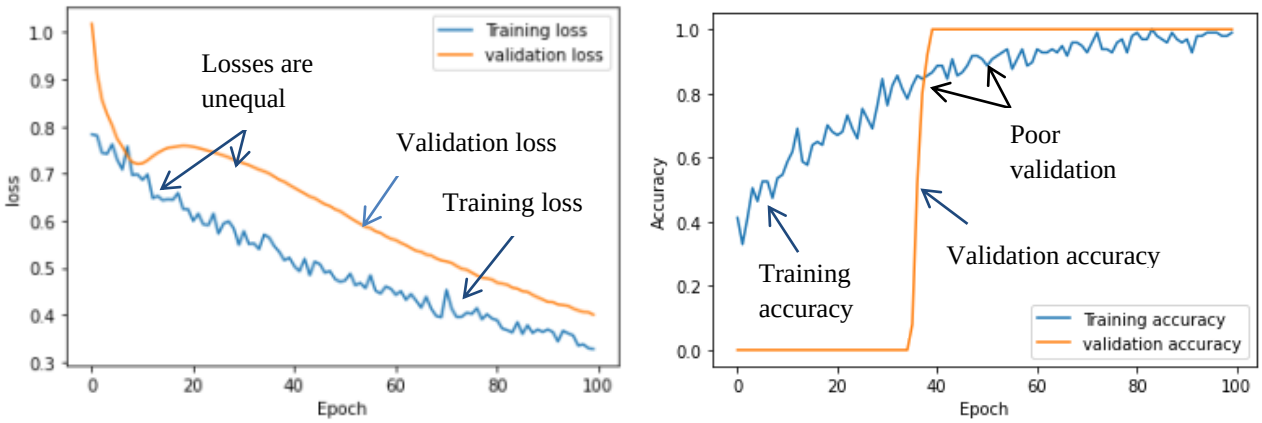
This mode of action is utilized by the model whenever epoch is more than 100. The training loss as in (a) and accuracy as in (b) of the system are shown in the Fig. 5 when e_no=500. In this observation both the curve loss decreases when epoch number increases, once the number is 500 ,both the losses coincide each other that means losses are equal, as a result this leads to 100% training validation accuracy. However, if the system tested under this condition, the model gives good results.

c) In over-fitting, the validation and training losses curves go parallel even for a greater number of epochs.

Despite of an increase in epoch more than 500, the two curves never meet each other as show in Fig. 6. as result of this, unequal losses as in (a) and 50% of the training validation accuracy as in (b). Testing the system under these conditions leads to less accurate results. The model is trained separately for the L2-L4 image dataset, which yields similar results.

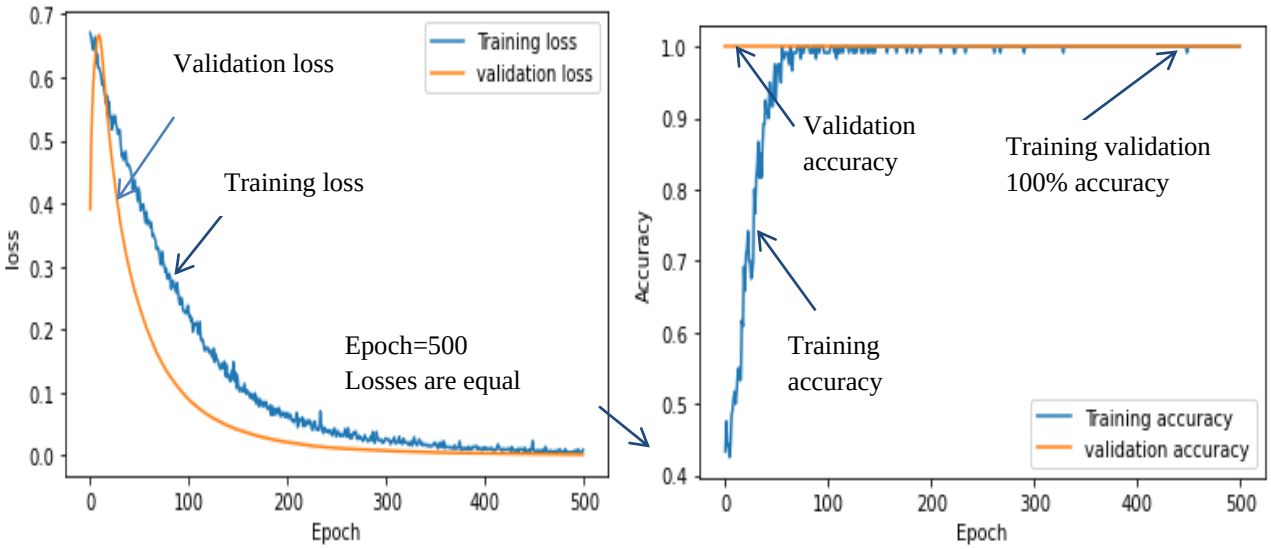
B. Test the model with k-Fold Cross Validation

Machine learning models are evaluated by K-fold cross-validation, so the model is tested for different folding. By estimating the performance of a model using unseen data, it provides a robust estimate. A model is trained on each subset (4-fold) except one that is held out from the training and its performance is evaluated on the subset that was held out (test dataset). Each subset will be included in the held-out test set until all subsets are included. The test-set =40 (Normal (TN) =19 and Abnormal (TP) =21) for L1 dataset and validation split= 0.2%, batch size=32. The under fitting model confusion matrix as illustrated in Fig.7 with 1 abnormal i.e., (TP) out of 21 abnormal but it then shows 20 as normal i.e., (FN) , total ($P=1+20$).



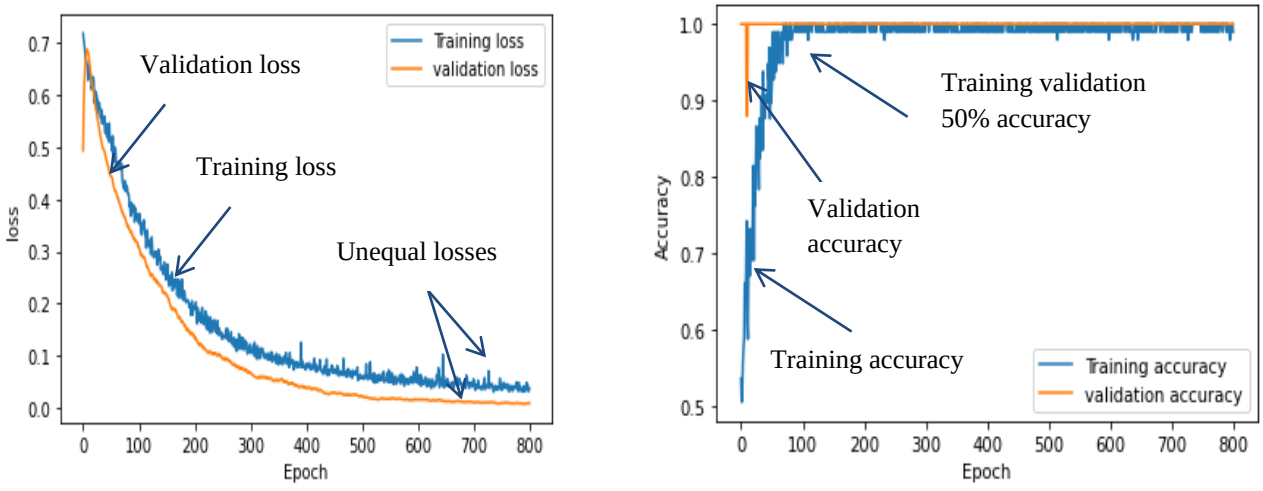
(a) Loss (b) Accuracy

Fig. 4: Under-fitting model for epoch=100 for dataset L1(a)Loss and (b)Accuracy curves.



(a) Loss (b) Accuracy

Fig. 5: Correct-fitting model for epoch=500 for set L1(a) Loss and (b) Accuracy curves.



(a)Loss (b) Accuracy

Fig. 6: Over-fitting model for epoch=800 for set L1(a) Loss and (b) Accuracy curves

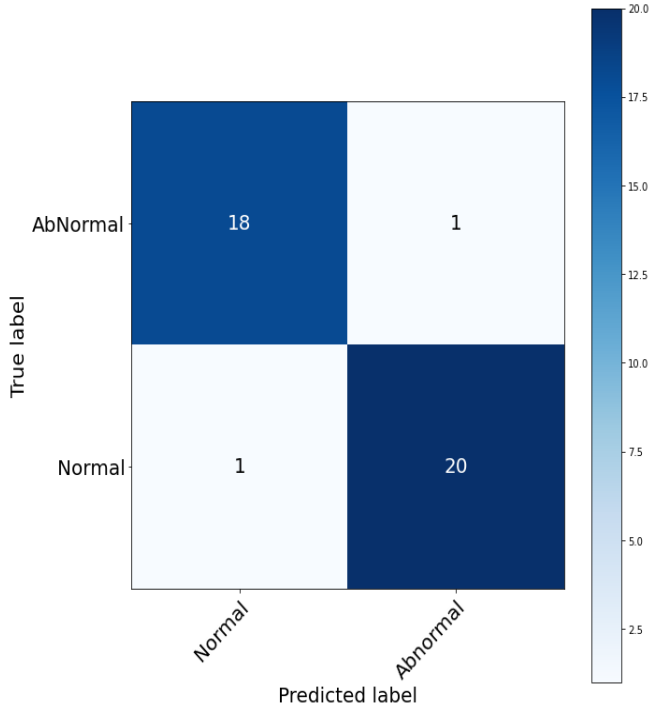


Fig. 7: Confusion matrix diagram test accuracy for UF

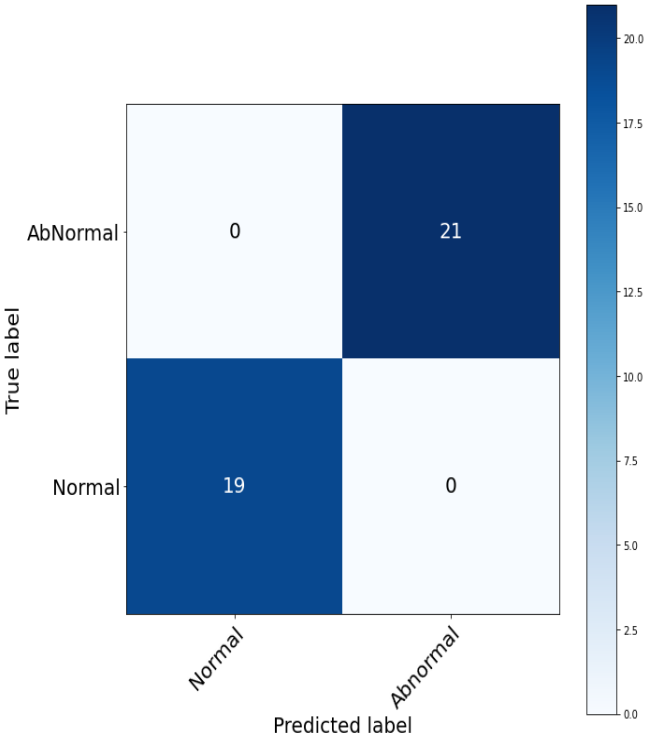


Fig. 8: Confusion matrix diagram test accuracy for CF

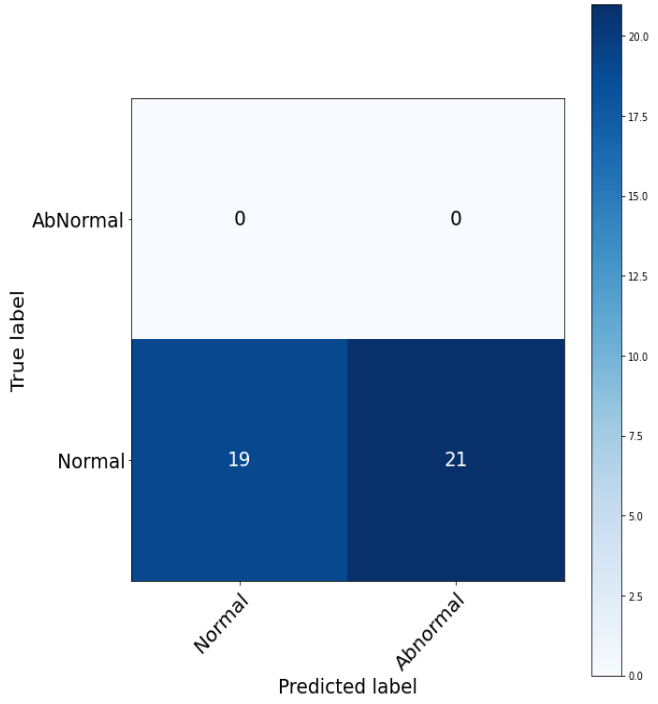


Fig. 9: Confusion matrix diagram test accuracy for OF

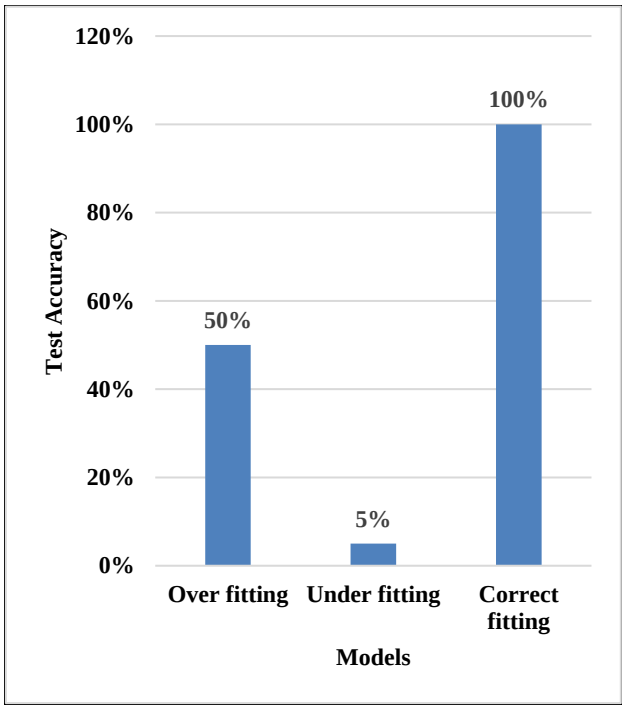


Fig. 10: Model test accuracy

And the (TN) true normal 19 shows as $N = 18 + 1$ i.e.,18 as abnormal and 1 normal, which results accuracy of 5 % was achieved for the 4-fold test. According to Fig. 8, the correct-fitted model has a true positive of 21 and a true negative of 19, gives it as 100% test accuracy i.e., $P = 21 + 0$ and $N = 0 + 19$. Here the Fig. 9 is the confusion matrix for the over fitted model, $P = 0 + 21$ and $N = 0 + 19$,thus giving 50% test accuracy. Moreover, the model is also tested one by one by considering remaining group of the L2-L4 image dataset ,it results similar output. The Fig. 10 illustrates graphical representation of model test accuracy for OF ,UF and CF. For different scenarios, the confusion matrix shows a testing sample of 40, training validation split of 0.2% and batch size of 32 as in Table II and with system accuracy, true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are tabulated in Table II for 4- fold . As a result, the correct fitting model results in 100% accuracy over the other two scenario.

TABLE II
CONFUSION MATRIX FOR DIFFERENT SCENARIO

Model	No. of Epoch	True Positive	True Negative	False Positive	False Negative	Test Accuracy
Over fitting	800	0	19	0	21	50%
Under fitting	100	1	1	18	20	5 %
Correct fitting	500	21	19	0	0	100%

TABLE III
VARIOUS VALIDATION SPLITS FOR THE CORRECT FITTING MODEL (4-FOLD, EPOCH-500)

Validation split (%)	Total data =122		Training Loss	Training Accuracy	Validation Loss	Validation Accuracy	Test Accuracy
	No. of training data	No. of validation data					
0.2	98	24	0.0075	1.0000	0.0065	1.0000	1.0
0.3	86	36	0.0133	1.0000	0.0110	1.0000	1.0
0.4	73	49	0.0013	1.0000	7.5511	0.1111	0.5
0.5	61	61	0.0725	1.0000	8.8114	0.3388	0.4

TABLE IV
VARIOUS CROSS-FOLD MODEL FOR TRAINING AND TESTING

Cross-fold	Total L1 samples -162			
	Training		Testing	
	Normal	Abnormal	Normal	Abnormal
2	38	43	38	43
3	51	58	25	28
4	57	65	19	21

For the correct fitting model, the Table III illustrates various validation splits, epoch=500, batch size=32. If the validation split in training process increases, real training process data reduces since more data will be used for validation but the model will not receive enough data to train. This results in less accuracy of training, but for good accuracy, it is imperative to have a sufficient amount of data. With this model, an accuracy of 100% is achieved if the validation split is 0.2% and 0.3%. The Table IV illustrates different ways of folding data sets. A training dataset and a testing dataset is different for different folding. For 2-fold, training set=81 (normal (N) =38, abnormal (Abn) =43), testing set = 81 normal (N) =38, abnormal (Abn) = 43. For 3-fold, training set=109 (N=51,Abn=58), testing set= 53 (N=25, Abn=28).For the 4-fold, training process dataset =122 (N=57,Abn=65),testing set= 40 (N=19, Abn=21).As the training set becomes larger than the testing set, the system becomes more accurate. Based on Table II, experimental results show that a 4-fold with 500 epochs provides the best fitted model with 100% test accuracy. Using this as the reference, the dataset is varied for different folding to get the confusion matrix as illustrated in Table V with epoch=500, batch size=32, validation split=0.2. The TP and TN values are accurate with the 4-fold since the training dataset is larger than the test dataset, so the system will learn the maximum texture features.

TABLE V
CONFUSION MATRIX FOR DIFFERENT CROSS-FOLD

Cross-fold	True Positive	True Negative	False Positive	False Negative
2	14	11	08	07
3	16	13	06	05
4	21	19	0	0

TABLE VI
PERFORMANCE EVALUATION MEASURES (PERCENTAGE) EPOCH=500

Cross-fold	Precision	Sensitivity	Specificity	Recall	F1-Score	Accuracy
2	64	67	58	67	65	63
3	60	76	68	76	74	70
4	100	100	100	100	100	100

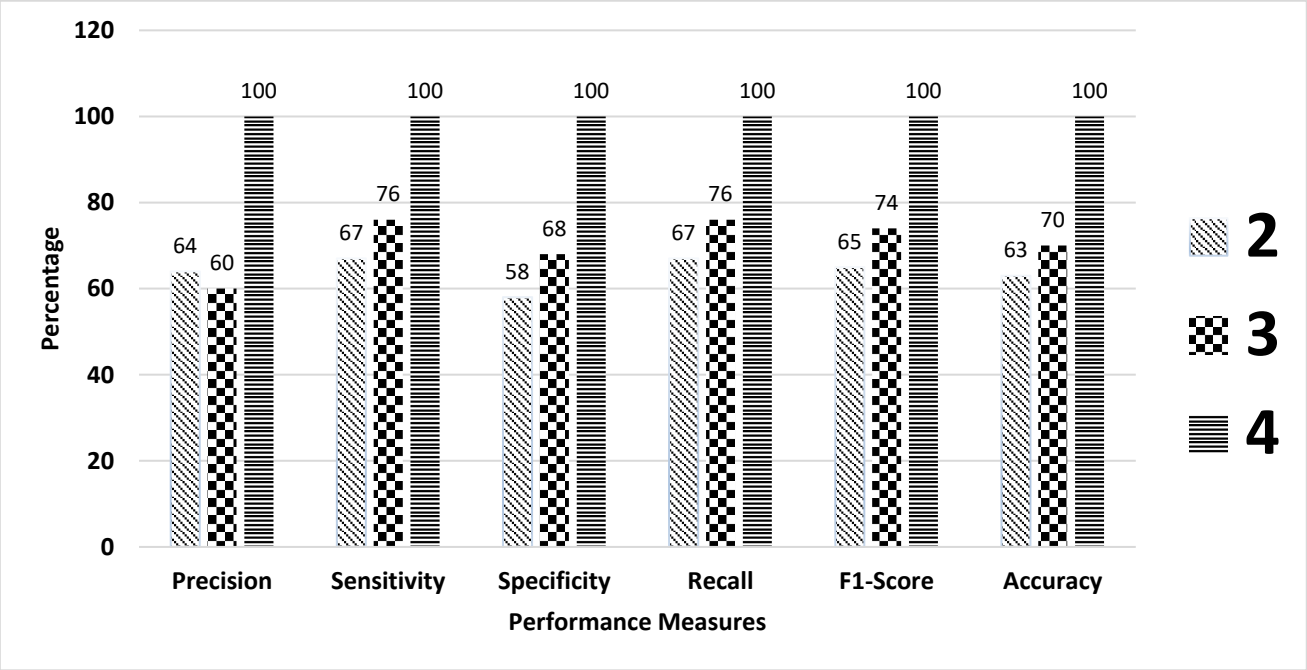


Fig. 11: System performance analysis

Table VI illustrates that : analysis of the system performance is measured in percentages with epochs fixed at 500 is calculated based on confusion matrix Table V, various measures are calculated for different data folding. There is a greater level of accuracy for all measures with a training validation split of 0.2 for 4-folding. As shown in Fig.11, the 4-fold cross-validation demonstrates 100% accuracy using the correct fitting model. With 4-folding, the model is able to learn all the texture patterns since enough training is provided by considering good number of epochs. The receiver operating characteristic curve (ROC) curves are used in binary (Abnormal-1, Normal-0) classification problems to assess how effective this algorithm. TP rate versus FP rate is plotted as a probability curve. The AUC [34] measures a classifier ability to differentiate between classes using the ROC curve; it serves as a summary indicator of its performance. In general, a high area under the curve (AUC) indicates a model is more accurate in separating abnormal from normal classes .Abnormal and normal class images cannot be distinguished by AUC=0.5 as shown in Fig. 12 in under fitting model, the baseline and ROC curves are coinciding each other i.e., poor performance. Abnormal class images are appropriately distinguished from normal class images when AUC = 1 as shown in Fig. 13 with correct fitting model and ROC curve occupies entire area, so the system gives 100 % accuracy i.e., excellent performance.

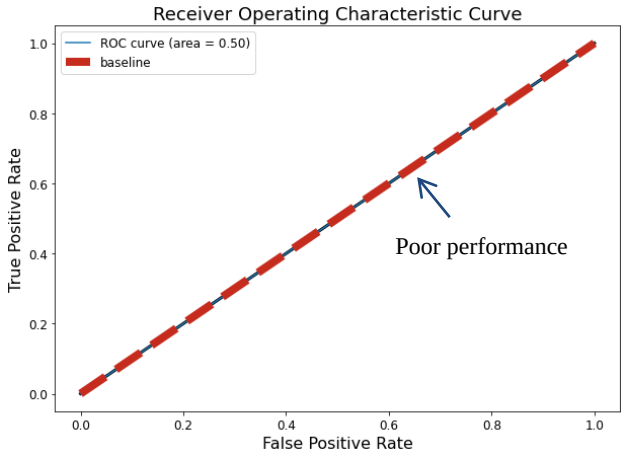


Fig. 12: AUC-ROC curve for under fitting

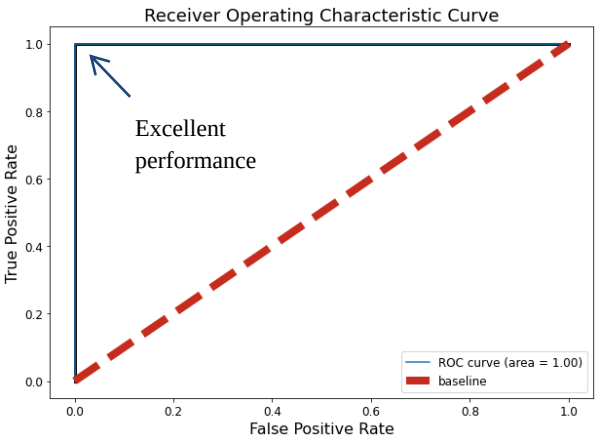


Fig. 13: AUC-ROC curve for correct fitting

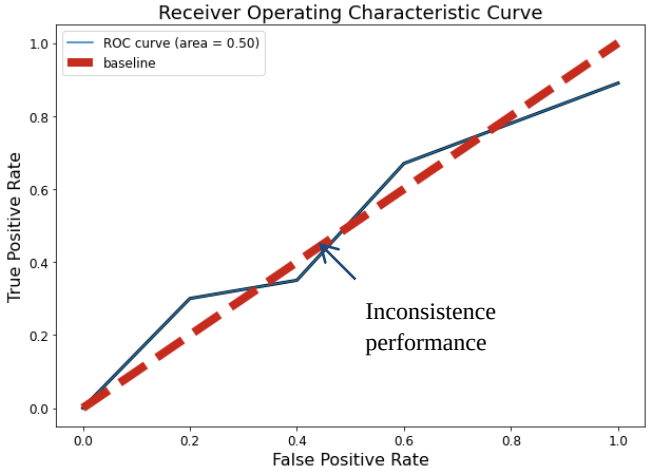


Fig. 14: AUC-ROC curve for over fitting

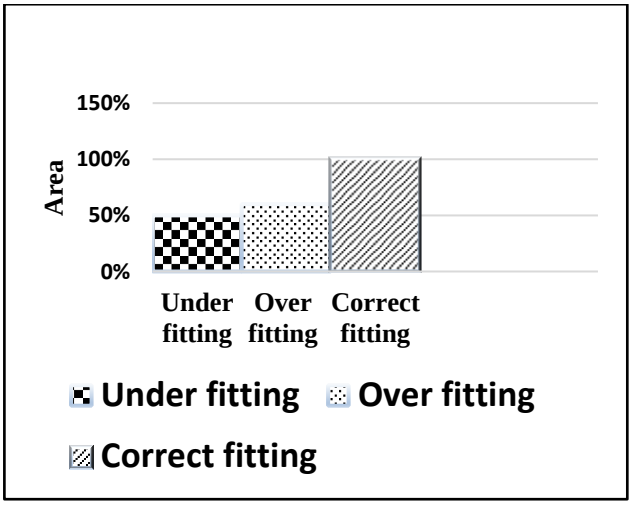


Fig. 15: Model-specific area

Fig.14 over fitting model, this implies that the classifier predicts either a random or constant class for all the images .In this case the system accuracy is varying i.e., inconsistence performance. Overall AUC values indicate how well a classifier separates normal and abnormal classes. Therefore, higher AUC means more accuracy. In different scenarios, the area occupied by the model is as shown in Fig.15. Correct fitting model occupy 100% area, i.e., the system is very sensitive to changes in input according to whether the test data is true or false.

C) X-ray score lumbar spine(L1-L4)

The Z-Score as well as T-Score are reported in DEXA to identify the osteoporosis in bone[35]. T-score reflects a person's bone density relative to that of a healthy 30-year-old of the same gender. A person's Z-score is calculated by comparing their bone density to the average bone density of the same age and gender. When scores are lower (more negative), bone density is lower: If the T-score is -2.5 or less, osteoporosis is diagnosed. Osteoporosis cannot be diagnosed with only Z-scores. Osteoporosis can be caused by a variety of causes, including a low Z-score. Scientific research requires the use of statistical methods to analyze variety of data sets [36]. In addition to that, it assists in making appropriate decisions based on the results of the research. In this study, Law texture images (normal-76,abnormal-86) are used to determine the normal and abnormal bone of L1-L4 based on four statistical measures called X-scores in order to correlate the reports obtained from the DEXA . Average pixel values (data points) of normal and abnormal pixels are analyzed using the mean. Data points are assessed in terms of their standard deviation in order to calculate a normal distribution. According to this distribution, data points deviate from their mean positions in different ways. In order to identify normal and abnormal, the correct fitting model is used to calculate the statistical values by considering Law texture feature image as mentioned in (5). As the intensity values vary in L1-L4, the mean value is calculated based on the average. Normal distributions have the following useful properties: Z-scores are used to assess it, which simply show the standard deviation of each point from the mean. Data points are converted to Z-scores by subtracting the mean value and divide by the standard deviation of the distribution. Kurtosis measures how weighted or unweighted the data are in relation to a normal distribution. A high kurtosis means there are often high outliers or heavy weighted in data sets. Lack of symmetry, or skewness, is a measure of symmetry of the data points about the mean. For a normal image, L1-L4 must have mean, skewness, and kurtosis values greater than or equal to the values in Table VII. Images that do not meet these criteria are abnormal.

TABLE VII
THE AVERAGE X-RAY SCORE

Lumbar Vertebrae	Mean	Standard deviation	Skewness	Kurtosis
L1	10.70	0.99	-15.19	13.50
L2	11.7	1.33	-12.42	18.49
L3	11.0	0.94	-11.51	13.31
L4	11.2	1.15	-10.93	18.02

TABLE VIII
EVALUATION OF THE PROPOSED SYSTEM AGAINST THE EXISTING SYSTEM

Reference	Imaging Modalities	Classifiers	Accuracy (%)	Sensitivity (%)	Specificity (%)	Cross validation
[5]	X-ray and DEXA report	DT	99.30	99.11	99.50	5-fold
		RF	99.07	99.10	99.03	5-fold
[6]	DEXA	-	65.00	71.00	59.00	-
[13]	DEXA	CNN	94.00	87.40	88.40	
[17]	DEXA	CNN	96.00	87.40	88.40	10-fold
[18]	DEXA/CT	CNN	76.65	-	-	-
[19]	CT	CNN		86.67	92.31	-
[20]	DEXA	Linear and Non-linear	87.7	87.70	63.20	-
Proposed	**X-ray with DEXA**	**CNN**	**100**	**100**	**100**	**4-fold**

Standard deviations must be less than or equal to those listed in the Table VII to be considered normal, if not, the image is abnormal. Table VIII illustrates a comparison of osteoporosis detection using different methods with system performance. The proposed system is nevertheless more accurate than other existing systems.

VII SUMMARY

To avoid fracture risk, early detection of osteoporosis is important in recent years due to its significant impact on society. Using a cost-effective method, abnormal lumbar spine (L1-L4) is identified more efficient in proposed work. In order to make the system perform better, Law texture analysis with CNN plays an important role. With the correct fit model, the work implemented here is 100% accurate, with no complications. In recent years, the DEXA is the measurements for bone and soft tissue are not standardized, which is a major drawback. In addition to this, measurements can vary depending on the manufacturer and type of instrument. The system is further made general for which universal data sets are required. In order to improve the system performance for the detection of osteoporosis, especially in the lumbar spine, the dataset to be prepared in a way that combination of image data as well as clinical information of osteoporosis. An automatic osteoporosis diagnosis system will be more sophisticated as a result.

ACKNOWLEDGMENT

Dr. A Ramalingaiah provided all necessary dataset and validated the experimental outcomes.

REFERENCES

[1] T. Sözen, L. Özışık, and N. Ç. Başaran, "An overview and management of osteoporosis," *Eur. J. Rheumatol.*, vol. 4, no. 1, pp46-56, 2017.

[2] M. Gass and B. Dawson-Hughes, "Preventing osteoporosis-related fractures: an overview," *Am. J. Med.*, vol. 119, no. 4, pp S3–S11, 2006.

[3] T. M. Link, "Radiology of osteoporosis," *Can. Assoc. Radiol. J.*, vol. 67, no. 1, pp28–40, 2016.

[4] M. Fradi, M. Afif, and M. Machhout, "Deep learning-based approach for bone diagnosis classification in ultrasonic computed tomographic images," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 12, pp80-87, 2020.

[5] K. A. Patil, K. M. Prashanth, and A. Ramalingaiah, "Osteoporosis Detection in Lumbar spine L1-L4 based on Trabecular Bone Texture Features," *Int. J. Intell. Eng. Syst.*, vol. 14, no. 6, pp80–94, 2021.

[6] N. E. Hassan, S. A. El-Masry, R. A. El-Banna, S. Y. El Sherity, and others, "Early detection of osteoporosis in premenopausal Egyptian women," *Egypt. Pharm. J.*, vol. 19, no. 2, pp149-154, 2020.

[7] S. S. Yeap *et al.*, "Different reference ranges affect the prevalence of osteoporosis and osteopenia in an urban adult Malaysian population," *Osteoporos. Sarcopenia*, vol. 6, no. 4, pp168–172, 2020.

[8] K. A. Patil, K. M. Prashanth, and A. Ramalingaiah, "Texture Feature Extraction of Lumbar Spine Trabecular Bone Radiograph Image using Laplacian of Gaussian Filter with KNN Classification to Diagnose Osteoporosis," in *Journal of Physics: Conference Series*, vol. 2070, no. 1, pp1-11, 2021.

[9] K. A. Patil and K. M. Prashant, "Segmentation of Lumbar [L1-L4] AP Spine X-ray images using various Level Set methods to detect Osteoporosis," in *2021 IEEE Bombay Section Signature Conference (IBSSC)*, pp1–6, 2021.

[10] W. D. Leslie, L. M. Lix, J. F. Tsang, P. A. Caetano, M. B. D. Program, and others, "Single-site vs multisite bone density measurement for fracture prediction," *Arch. Intern. Med.*, vol. 167, no. 15, pp1641–1647, 2007.

[11] S. N. Fathima, R. T. Selvi, and M. P. Beham, "Assessment of BMD and Statistical Analysis for Osteoporosis Detection," *Biomed. Pharmacol. J.*, vol. 12, no. 4, pp1907–1914, 2019.

[12] X. Xu *et al.*, "Discordance in diagnosis of osteoporosis by quantitative computed tomography and dual-energy X-ray absorptiometry in Chinese elderly men," *J. Orthop. Transl.*, vol. 18, pp59–64, 2019.

[13] P. Menshchikov, A. Ivantsova, T. Akhadov, M. Ublinskiy, A. Manzhurtsev, and N. Semenova, "1H MRS as a novel quantitative method for osteoporosis detection," in *Journal of Physics: Conference Series*, vol. 1461, no. 1, pp1-4, 2020.

[14] D. Hussain and S.-M. Han, "Computer-aided osteoporosis detection from DXA imaging," *Comput. Methods Programs Biomed.*, vol. 173, pp87–107, 2019.

[15] S. Derkatch, C. Kirby, D. Kimelman, M. J. Jozani, J. M. Davidson, and W. D. Leslie, "Identification of vertebral fractures by convolutional neural networks to predict nonvertebral and hip fractures: a registry-based cohort study of dual X-ray absorptiometry," *Radiology*, vol. 293, no. 2, pp405–411, 2019.

[16] C. Bergh, A.-C. Söderpalm, and H. Brisby, "Preoperative dual-energy X-ray absorptiometry and FRAX in patients with lumbar spinal stenosis," *J. Orthop. Surg.*, vol. 13, no. 1, pp1–7, 2018.

[17] Y.-C. Lu *et al.*, "Prevalence of osteoporosis and low bone mass in older Chinese population based on bone mineral density at multiple skeletal sites," *Sci. Rep.*, vol. 6, no. 1, pp1–9, 2016.

[18] X.-H. Ma, W. Zhang, Y. Wang, P. Xue, and Y.-K. Li, "Comparison of the spine and hip BMD assessments derived from quantitative computed tomography," *Int. J. Endocrinol.*, vol. 2015, pp1-5, 2015.

[19] T. Nissinen *et al.*, "Detecting pathological features and predicting fracture risk from dual-energy X-ray absorptiometry images using deep learning," *Bone Rep.*, vol. 14, pp1-11, 2021.

[20] C. Tang *et al.*, "CNN-based qualitative detection of bone mineral density via diagnostic CT slices for osteoporosis screening," *Osteoporos. Int.*, vol. 32, no. 5, pp971–979, 2021.

[21] S. Mu, J. Wang, and S. Gong, "Application of Medical Imaging Based on Deep Learning in the Treatment of Lumbar Degenerative Diseases and Osteoporosis with Bone Cement Screws," *Comput. Math. Methods Med.*, vol. 2021, pp1-10, 2021.

[22] V. Egorov *et al.*, "Osteoporosis detection in postmenopausal women using axial transmission multi-frequency bone ultrasonometer: Clinical findings," *Ultrasonics*, vol. 54, no. 5, pp1170–1177, 2014.

[23] Y. Li, B. Wang, K. Zhang, Z. Jiang, W. Shi, and W. Liu. "A Novel Automatic Method Based on U-Net for Lung Fields Segmentation," *Engineering Letters*, vol. 30, no. 2, pp636-643, 2022.

[24] W. T. Chan and K. S. Sim, "Termination factor for iterative noise reduction in MRI images using histograms of second-order derivatives," *Int. J. Comput. Sci.*, vol. 48, no. 1, pp174–180, 2021.

[25] K. I. Laws, "Rapid texture identification," in *Image processing for missile guidance*, vol. 238, pp376–381,1980.

[26] S. Dash and U. Jena, "Texture Classification Using Laws' Filter in Various Color Spaces," *Int. J. Eng. Technol. IJET*, vol. 9, no. 2, pp 454–460, 2017.

[27] Y. Hu, X. Zhang, J. Yang, and S. Fu. "A Hybrid Convolutional Neural Network Model Based on Different Evolution for Medical Image Classification," *Engineering Letters*, vol. 30, no. 1, pp168-177, 2022.

[28] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp2352–2449, 2017.

[29] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 international conference on engineering and technology (ICET)*, pp. 1–6, 2017.

[30] J.-S. Lee, S. Adhikari, L. Liu, H.-G. Jeong, H. Kim, and S.-J. Yoon, "Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study," *Dentomaxillofacial Radiol.*, vol. 48, no. 1, pp1-8, 2019.

[31] Z. Vujović, "Classification model evaluation metrics," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp599–606, 2021.

[32] K. A. Patil, K. M. Prashanth, and A. Ramalingaiah, "Classification of Osteoporosis in the Lumbar Vertebrae using L2 Regularized Neural Network based on PHOG Features," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 4, pp413-423, 2022.

[33] S. Pothuganti, "Review on over-fitting and under-fitting problems in Machine Learning and solutions," *Int J Adv Res Electr Electron Instrum. Eng*, vol. 7, no. 9, pp3692–3695, 2018.

[34] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in *Proceedings of the 23rd international conference on Machine learning*, pp233–240 ,2016.

[35] J. Wendlová, "Differences in distribution of T-scores and Z-scores among bone densitometry tests in postmenopausal women (a comparative study).," *Wien. Med. Wochenschr. 1946*, vol. 152, no. 23–24, pp591–595, 2002.

[36] P. Kaur, J. Stoltzfus, V. Yellapu, and others, "Descriptive statistics," *Int. J. Acad. Med.*, vol. 4, no. 1, pp60-63, 2018.

BIBLIOGRAPHY

Kavita Avinash Patil* received post-graduation in Very Large-Scale Integration Design and Embedded system from Karnataka Lingayat Education College of Engineering Belgaum in 2012. Faculty at New Horizon College of Engineering Bangalore, pursuing Ph.D. in the Visvesvaraya Technological University. Area of interest includes Biomedical Engineering, machine learning, and artificial intelligence. Total 10 years of teaching experience.

Dr. K. V. Mahendra Prashanth is received Bachelor of Engineering degree in Electronics & Communication Engineering at National Institute of Engineering, Mysore; post-graduate degree in Power Electronics from University of Visvesvaraya College of Engineering Bangalore. Ph.D. degree in Electronics & Communication Engineering from Visvesvaraya Technological University, working as Professor in Department of Electronics & Communication Engineering, Chief Coordinator for Research and

Development and Post-Graduation programmes at Sri Jagadguru Balagangadhranatha Institute of Technology, Bangalore. A survey report issued by BIOMED publications identifies two of his technical papers as the top two. His memberships include IEEE, Acoustical society of India, ISTE, IEI, and IIAV.

Dr. Ramalingaiah A received his Bachelor of Medicine and a Bachelor of Surgery from the Bangalore University, Bangalore Medical College in 1984 and Master in Surgery Orthopedics from Gulbarga University. Government Medical College Bellary in 1993(KMC Registration number-22719).Received his fellowship in Arthroscopy, Spine Surgery, A.O. Trauma Basic and Advanced A.O. Pediatric Trauma, Postgraduate Diploma in Clinical and Research and leadership in healthcare. 32 years of experience as a surgeon, teaching and research. A total of 12 papers have been published in national and international journals in the field of orthopedics. Inspecting the facilities at various Medical Colleges across the India on behalf of Medical Council of India.