Fuzzy Rank-Based Ensemble Model for Accurate Diagnosis of Osteoporosis in Knee Radiographs

Saumya Kumar[1], Puneet Goswami[1], Shivani Batra[1]
Department of Computer Science &Engineering, SRM University, Delhi NCR, Sonipat, India[1]

***Abstract*—The main factor in fractures among seniors and women post-menopausal is osteoporosis, which decreases the density of bones. Finding a low-cost diagnostic technology to identify osteoporosis in its initial stages is imperative considering the substantial expenses of diagnosis and therapy. The simplest and most widely used imaging method for detecting bone diseases is X-ray radiography, however, it is problematic to manually examine X-rays for osteoporosis as well as to identify the essential components and choose elevated classifiers. To categorize x-ray pictures of knee joints into normal, osteopenia, and osteoporosis condition categories, authors present a process in this investigation that uses three convolutional neural networks (CNN) architectures, i.e., Inception v3, Xception, and ResNet 18, to create an ensemble-based classifier model. The suggested ensemble approach employs a fuzzy rank-based unification of classifiers by taking into account two distinct parameters on the decision scores produced by the aforementioned base classifiers. Contrary to the straightforward fusion strategies that have been mentioned in the literature, the suggested ensemble methodology finalizes predictions on the test specimens by considering the confidence in the recommendations of the base learners. A 5-fold cross-validation approach has been employed to assess the developed framework using a benchmark dataset that has been made accessible to the general population. The suggested model yields an accuracy rate of 93.5% with a loss of 0.082. Further, the AUC is observed to be 98.1, 97.9 and 97.3 for normal, osteopenia and osteoporosis, respectively. The results demonstrate the model's usefulness by outperforming various state-of-the-art approaches.**

***Keywords*—Convolutional Neural Network; diagnosis; knee; osteoporosis; transfer learning models; X-rays**

I. INTRODUCTION

Osteoporosis is a serious disease that affects 200 million women globally and 9% of Americans over fifty in the US [1]. In affluent countries, one in three individuals may suffer from an OCF (osteoporotic compression fracture) [1]. After the primary injury, the risk of subsequent fractures significantly increases [2, 3]. A worse life expectancy and a higher mortality rate are both associated with even one OCF [4].

Bone mineral density (BMD) is measured using the Dual Energy X-ray Absorptiometry (DXA) procedure, which establishes the T-score and Z-score values recommended by the WHO for various phases of osteoporosis [5]. However, it has several drawbacks, such as areal estimations as well as the expensive and limited availability of the technology. The Quantitative Ultrasound System (QUS) [6], Computed Tomography (CT) [7], and Magnetic Resonance Imaging (MRI) [8] are further imaging modalities used to identify osteoporosis. Whereas CT provides 3D geometric scanning with quantitative measures but has an intense radiation exposure and does not meet the WHO's criterion of osteoporosis diagnosis, MRI is a 3 T enhanced bone microarchitecture optical technique but is highly expensive and has lesser pixel density [8]. Although QUS is easy to use, non-invasive, compact, and economical and employs acoustic signals to investigate bones, it is site-specific and lacks substantial empirical support [6]. Given these constraints, a low-cost, easily accessible, and reliable detection system is essential. To create computer-aided diagnostic (CAD) systems, the investigators used the latest developments in machine vision to analyze medical images and computer algorithms.

Numerous CAD systems, which include deep learning at multiple bone locations like the joint, vertebrae, palm, and dental, are suggested for osteoporosis assessment [9, 10, 11, 12]. However, little research has been conducted to diagnose knee osteoporosis. Being the joint that supports the body's weight and facilitates motion, the knee experiences the most strain. Women are more susceptible to tibial and fibular fractures, which raises the likelihood of osteoporotic fractures around the knee including in elderly society [13]. A significant 1-year fatality rate of 22% is recorded in senior patients who suffer femoral injuries, with poorer mobility and bad living conditions [14], and it is anticipated that almost half of the knee fractures happen in individuals who are aged over 50 years. To avoid fractures and save healthcare expenditures, an early diagnosis method is required to determine the incidence of osteoporosis in the knee bone [15].

Convolutional neural network (CNN) approaches based on deep learning are becoming increasingly widely used in recent years within CAD systems for medical image interpretation [16] owing to their state-of-the-art performance in identifying a variety of illnesses from pictures, including brain tumors [17], respiratory disease [18, 19], cancer [20] and others. In terms of classifying medical pictures, CNNs [21] have produced cutting-edge results. The fundamental issue with utilizing CNN learners is that they require a significant quantity of annotated data for training; however, it's extremely challenging to find a large-scale dataset in the health sector. Investigators have put forward the concept of transfer learning to overcome difficulties [18, 22]. In transfer learning, a CNN that has been trained on a large population and then provided with training on a lower dimension of a different issue makes use of the information learned from the larger dataset to quickly learn the characteristics of the lower dimension and so efficiently aid in image classification. Various models, however, may perform better on specific data configurations, meaning that some

categories in the dataset may have more precise categorization than others. Moreover, traditional rank-based ensemble approaches do not utilize the diversity of the forecast odds. The significant fact could thus go ignored as a consequence. This fact led authors to develop a unique technique in this study wherein authors quantified two crucial variables, i.e., prediction probability's proximity to 1 and its deviation from 1—and used all the knowledge accessible through different base classifiers. Furthermore, the proposed method combines all of these quantifiable elements to reach the final forecast, allowing it to handle the categorization issue more and produce an accurate conclusion.

Ensemble learning [23] is an approach, in which the assessment ratings of many learners are combined to forecast the ultimate target class of an input data set. An ensemble method aims to capture the key characteristics of each of its component models, outperforming each base classifier individually. These systems are reliable because ensembling reduces the range or scattering of the base models' estimates. By adding substantial bias to the contending base classifiers, the ensemble model's diversity in the predictive performance of the base classifiers is mitigated. The standard ensemble approach employed in literature uses pre-calculated ratings for the classifiers and accords equal value to all constituting models' classification results. The biggest drawback of such an ensemble is the creation of fixed ratings that are challenging to change throughout the test sample classification phase. However, the suggested fuzzy rank-based ensemble technique accounts for each base classifier's forecast rank for each unique test case independently. In this approach, the ensemble technique can produce improved and more precise classification results. In the current study, authors develop a fusion technique that combines the judgment values from three basic CNN classifiers, i.e., Inception v3 [24], Xception [25], and ResNet 18 [26], to build the proposed ensemble model.

A. Key Contributions

An end-to-end classifier employing just deep learning algorithms may not function adequately on new datasets due to the dearth of data accessible in the medical sector. To create an ensemble approach that includes the forecasts from other competing systems, the authors employ three transfer learning-based CNN models. Although straightforward fusion techniques like popular vote, balanced averaging, and others have been employed in the research, they don't consider the predictor's confidence when making assertions. By taking this into account when developing the statistical framework for the suggested technique, authors can outperform basic ensemble techniques that are often employed for diagnosis. The current study's highlighting accomplishments are listed below.

- The implementation of ensemble methods employing the three base classifiers, i.e., Inception v3 [24], Xception [25], and ResNet 18 [26] improves the effectiveness of the whole system for prediction on the limited amount of accessible data.
- The fuzzy rankings of the classes in the assessment scores are determined by applying two non-linear operations of various concavities in the ensemble approach that is being presented. The lower rank serves to identify the anticipated class after computing the sum of the products of the three base classifiers' ratings.
- The employment of two non-linear operations guarantees that the confidence in the classifiers' forecasts is taken into account in the derivation of the rankings, producing more accurate recommendations.
- The method authors use to measure the difference between the forecasted and anticipated values is unique. The suggested ensemble model's improvement in accuracy is also significant.
- Regarding classification precision and sensitivity, the suggested approach surpasses various cutting-edge techniques on the benchmark knee x-ray osteoporosis image dataset [9].

B. Section Division

The manuscript is hereafter divided into sections. Section II presents the literature survey. Section III presents the details of the proposed fuzzy rank-based ensemble model. Section IV highlights the experiments done and the results achieved. Finally, Section V concludes this research.

II. Literature Survey

Deep convolution neural networks (DCNN) in particular illustrate state-of-the-art achievements in illness identification [27]. Several investigators have developed an osteoporosis assessment method from various kinds of images with effectiveness using machine learning techniques [9]. The authors have covered the most recent advances in DCNN-based osteoporosis assessment in this segment. Osteoporosis from phalanges has been detected using DCNN on X-ray scans [28]. Researchers acquired a decent identification rate using three-fold cross-validation for assessment.

In [29], researchers used feature selection based on wrapping to compare several classification schemes for osteoporosis diagnosis. To identify osteoporotic fracture risk, Naoufami et al. [30] recommended DCNN in their study (VF). After using computed tomography scans to derive logical characteristics, the system's efficiency has been compared to that of working radiologists, and similar outcomes were obtained. DCNN was utilized by Derkatch et al. [31] to precisely identify vertebral fractures in DXA pictures. Krishnaraj et al. [32] used CT scans of the vertebrae to separate people into osteoporotic and non-osteoporotic groups. They acquired high accuracy while segmenting CT images using U-net CNN. Fang et al. [33] also used vertebral CT scans to look for osteoporosis. They distinguished between normal and osteoporotic vertebrae using the DenseNet-121. The spine X-ray characteristics have been retrieved by Lee et al. [34] using CNN architectures, and the results were then sent to classifiers for categorization. They used VGG for extracting features and random forest for classifying to reach the highest accuracy rate of 71%. Yasaka et al. [35] employed CT scans of the abdomen to estimate the BMD of the lumbar vertebrae. They discovered a strong association between the DXA BMD and the anticipated BMD from CNN. Sollmann et al. [36] examined CT images of the spine and used CNN to calculate the

volumetric bone mineral density. Researchers observed that CNN provides good diagnostic performance when they evaluated the findings of the volumetric bone mineral density acquired from conventional CT.

By using a CNN and dental panoramic radiographs (DPRs), Lee et al. [37] have been able to detect osteoporosis from the tooth. This DCNN outperformed the outcomes of oral and maxillofacial radiologists. DPRs were employed by [38] to identify osteoporosis as well. To boost the CNN classifier's classification accuracy, they utilize the VGG-16 classifier and applied transfer learning to it. AlexNet Yu et al. [39] employed CNN to identify osteoporosis in dental panoramic radiography. They accurately divided the DPRs into osteoporotic and non-osteoporotic groups; however, they didn't only include the osteopenia group. Sukegawa et al. [40] also investigated DPRs for osteoporosis detection using CNNs and showed good results. The recognition rate has been further enhanced by the addition of prognostic factors. Deniz et al. [41] examined the MRIs of the proximal femur to look for osteoporosis. To quantify fracture risk and evaluate the condition of the bone, they segmented the proximal femur using DCNN.

Using X-ray scans of the pelvis, Liu et al. [42] identified osteoporosis. They derived the analytical expression from the softmax of the suggested U-net model, which employs X-rays to identify osteoporosis by analyzing the deep characteristics of the medullary joint. The photos of the osteoporosis and bone mass loss groups in this investigation are inadequately diagnosed. Utilizing CNN, Yamamoto et al. [43] identified osteoporosis in hip X-rays. Researchers discovered that adding clinical variables to the scans enhanced performance, with EfficeientNet CNN achieving the highest results. Tecle et al. employed the AlexNet Classifier to get an osteoporosis diagnosis [44]. They identified the osteoporotic and non-osteoporotic scans from the segmented second metacarpal area using X-ray scans of the hand. He et al. [45] examined the knee X-rays and suggested using two radiographic criteria for assessing bone strength: cortical bone thickness and distal femoral cortex. To train the CNN networks, the BMD and T-score values produced by the QUS method have been demonstrated to significantly correlate with these characteristics. Also when learned on a limited population, the CNN performs well because of transfer learning.

The present work suggests an ensemble learning method where the eventual selection is reached after taking decisions from many models into account. Sarwar et al. [46] employed a mean likelihood-based ensemble, while Xue et al. [47] used a consensus voting-based ensemble approach to examine several straightforward fusion techniques. Nevertheless, these imprecise ensemble models utilize predetermined or constant weights linked to the base classifiers and do not allow for the predictability of results. In consideration of this, the authors provide a unique ensemble approach in this study that combines the decision ratings across three CNN-based base classifiers, Inception v3 [24], Xception [25], and ResNet 18 [26], while also accounting for the base learners' level of confidence in their judgments.

III. Proposed Model

This section briefly discusses the base classifiers authors utilize and the essential customization authors perform to the fundamental models before going into more depth about how the suggested fuzzy rank-based fusion of the basis learners' confidence scores would be implemented. The goal in this case is to fully use all of the confidence factors produced by the basic classifiers by converting them into non-linear operations. One of the projected values indicates compliance with or proximity to 1, and the other one indicates divergence from 1. The standard ranking systems' flaw of not taking into account the aforementioned fact [48] and potentially producing inaccurate results is addressed by the suggested technique. Three base classifiers are used in the current work, and the clinical image dataset is used to test our methodology. In the beginning, authors train the base classifiers (personalization using pre-trained architectures learned on ImageNet [49]), and then we collect the confidence values. Then, to create non-linear fuzzy rankings and a merged rating that allows us to calculate the overall divergence from the predicted, the authors translate the scores onto two distinct functions with distinct concavities. A lower deviation indicates greater assurance in a given class. The winning group is given the ultimate class value, and it has the lowest divergence value. The proposed model is presented in Fig. 1.

A. Pre-Processing Input

The data fed to the proposed model is pre-processed in two steps, i.e., normalization and data augmentation.

- **Normalization**: Transforming picture data pixels to a specified range, such as (0, 1) or (-1, 1) is done through normalization. Most pictures have pixel values between 0 and 255. Large values can impede or slow the training phase in CNN. Thus, image normalization is advised as a best practice so that the values of pixels vary from 0 to 1.

- **Data Augmentation**: When using CNN models, it is crucial to make sure the system receives enough training data. Data augmentation is the process of applying multiple adjustments to source photographs to produce several changed versions of the same image. However, because of the augmentation techniques utilized, each duplicate is unique in a certain manner from others.

B. Base Classifiers

The proposed model utilizes three pre-trained CNN architectures as base classifiers, i.e., Inception v3, Xception, and ResNet 18.

- **Inception v3**: Among the most prevalent deep learning models is Inception v3, which is a member of the Inception group and makes use of several enhancements to address issues with earlier Inception models [50]. These enhancements involve using a supplementary classification model, factorized convolution operations, batch normalization, the RMSProp optimization method, and label smoothing. It creates feature maps in numerous aspects and layers from an input picture with

the proportions 299 × 299 x 3. The Inception v3 inception block gives us the option to use many feature extraction filters from a unified feature space. For more thorough feature extraction, these characteristics with various filters are combined and transmitted to the following layer.

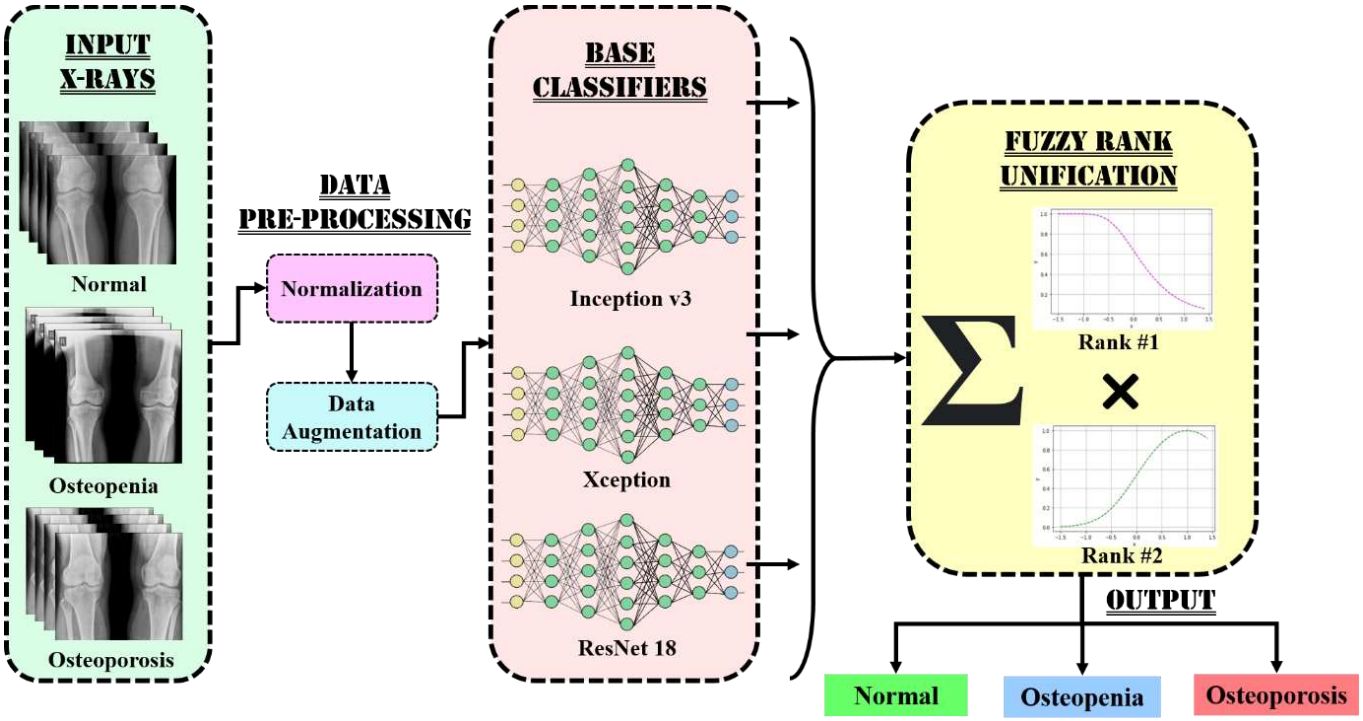

Fig. 1. Proposed model.

- **Xception**: The Inception v3 framework served as a basis for the Xception design created by Chollet et al. [25], which employs the same amount of design variables as the latter but does so more effectively. They demonstrated that the inception units fall in the center of a discrete spectrum, with pointwise convolutions and depth-wise separable convolutions at its two ends. To improve classification accuracy at the same computational load, they chose to substitute the inception modules with depth-wise separable convolutions.
- **ResNet 18**: ResNet, also known as Residual Network, is a structure that employs residual mapping and is particularly successful against the "degradation problem" in convolutional models. It was first presented in 2015. The CNN model's optimization process is enhanced by the residual learning technique. The ResNet-18 has been pre-trained on the ImageNet dataset, like the majority of widely used image-processing CNN models. It accepts photos with a 3 × 224 × 224 size as input, which is smaller than the inception v3 model's input size. The residual network might score higher the denser it is. The ResNet model is implemented at many depths, including ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-110.

C. Fuzzy Rank Unification

The confidence score generated by various base classifiers is fused using fuzzy rank-based unification. To give rankings to the class likelihood forecasted by a base classifier, authors employ a fuzzy ranking-based technique in which the likelihood scores are exposed to two non-linear operations: the exponential operation (Eq. (1)) and the hyperbolic tangent operation (Eq. (2)). Let C_{cf}^{op} be the confidence ratings for a base classifier, 'cf' corresponding to the output class, 'op'.

$$Rank_{1\_cf} = 1 - e^{-e^{-2*\sum_{op\in 1,2,3} C_{cf}^{op}}} \quad (1)$$

$$Rank_{2\_cf} = 1 - \tanh\left(\frac{(1-\sum_{op\in 1,2,3} C_{cf}^{op})^2}{2}\right) \quad (2)$$

The two non-linear operators double the scores they provided. The same steps are done for every base learner, and the final scores are calculated by adding the rank products from every predictor. Authors employ two distinct operations with various concavities so that their outputs might be complementary. In this investigation, $1 - e^{-e^{-2*\sum C_{cf}^{op}}}$ has a concave downhill shape in the range [0, 1]. The output rank1 value will attempt to shift towards 1 due to its negative slope in the [0, 1] range. $1 - \tanh\left(\frac{(1-\sum C_{cf}^{op})^2}{2}\right)$ is concave uphill in this definition [0, 1]. The resulting rank value will attempt to shift towards 0 due to its positive gradient in the [0, 1] range. About a specific confidence score acquired from a base classifier, the rank value is the result of incentive and divergence. Fusion comprises combining the several rankings connected to identification and selecting an alternate rank that will help make the ultimate choice using Eq. (3).

$$Rank_{cf} = \sum(Rank_{1\_cf} * Rank_{2\_cf}) \quad (3)$$

The major goal of employing two rankings is to take into account how closely and how far the predominant classification result deviates from the predicted outcome. Lower product value and a good result are correlated with reduced deviation. The final result of the ensemble classifier is therefore the class with the least value under this sum of products of rankings. The author's objective is to lower this product since the two non-linear operations have opposing concavities in the domain [0, 1]. As a consequence, a strong confidence rating leads the ranking of one operation to increase and the ranking of the other operation to decrease. This sum of products produces a lower result when a prediction's confidence rating is elevated. The rank calculated using Eq. (3) can be used to determine the final grade for each class. Using Eq. (4), the class with the lowest fused rank is identified and declared as the winner.

$$class_{final} = \min_{cf \in 1,2,3} Rank_{cf} \quad (4)$$

IV. Experiments and Results

A. Environmental Setup & Dataset

The authors developed a model to separate individuals with X-ray scans into normal, osteopenia, and osteoporotic groups. This scenario has been simulated using the Python language. The systems, procedures, modules, and resources of TensorFlow 2.0 have been created by the authors using an open-source deep learning methodology (plus Keras). Python has been utilized to complete the analysis. The tests were run on Google Collaboratory using a Tesla K80 GPU graphics card, an Intel i7-core CPU running at 3.6GHz, 16GB of Memory, and the 64-bit version of Windows 11.

The dataset has been obtained from Mendeley data that [9] contributed, and it has been released in August 2021. The dataset includes x-rays from 240 subjects, of whom 37 had normal bone density (with 18 men and 19 women), 154 had osteopenia (with 59 men and 95 women), and 49 had osteoporotic bone density (with 31 men and 18 women). Data augmentation in Python has been employed to statistically augment the dataset pictures. After statistical augmentation, the dataset now includes 323 patient radiographs of normal, 323 osteopenia, and 323 osteoporosis knee x-rays.

B. Performance Evaluation Metrics

The datasets provided for X-ray scans are used by the authors to assess the effectiveness of the suggested models as they study the classification of normal, osteopenia, and osteoporotic patients. The authors focus on four characteristics that are typical of CNNs for each structure, i.e., accuracy curve, loss curve, confusion matrix, and area under the curve (AUC).

The accuracy curves of the model show how well it is acquiring and interpreting. The discrepancy between training and testing accuracy is a metric of overfitting. The training time and model orientation are shown by the loss graphs. A significant gap between both the training and testing graphs illustrates the learning spectrum with training. A confusion matrix expresses a way of how well a classifier performs in a group of testing datasets when the input variables are already known. Four essential terms are connected to every confusion matrix. [51].

- True Positives [TP]: These are instances where the affected person had the ailment despite the prediction being "yes".
- True Negatives [TN]: According to estimates, the answer is "no" and the samples are not contaminated.
- False Positives [FP]: The ailment is presumed to be present, although the patients do not. A Type I mistake may be used to describe this.
- False Negatives [FN]: Even though the model recommends "no", the condition still exists in persons. These are referred to as Type II mistakes. It is frequently used to represent crucial prediction statistics, enabling analysis and identifying relevant experimental patterns easier.

AUC is a productivity statistic that incorporates all feasible classification levels. One method of examining AUC is to assess the probability that the model ranks a random positive instance stronger than a random counter-example. The possibility of a random positive instance being positioned in front of a counterpoint chosen at random is represented by the AUC. The range of the AUC value is 0 to 1. The AUC of a framework with 100% erroneous estimates is 0.0, while the AUC of a system with 100% accurate estimates is 1.0.

C. Implementation

The proposed ensemble model is compared with the individual underlying base classifiers, i.e., Inception V3, Xception, and ResNet-18. The findings (classification accuracy and loss) for the underlying datasets for knee osteoporosis used in this work are shown in Table I, Fig. 2 (accuracy curve) and Fig. 3 (loss curve) by the individual base classifiers and proposed ensemble framework. The outcomes show that the suggested model performs well in terms of classification accuracy. The underlying dataset has a training duration per fold of 30 minutes.

Fig. 4 displays the confusion matrices that the individual base classifiers and developed ensemble models on the dataset utilized in this study have been able to produce. Fig. 5 displays the AUC of the proposed model. The suggested ensemble model may be employed as a plug-and-play paradigm in which new test pictures are fed into the model to provide forecasts using the ensemble method, ultimately assisting the expert doctors in making a more rapid and precise judgment.

TABLE I. Comparison with State-of-the-Art Methods

Model	Accuracy	Loss
Inception V3	89.8	0.217
Xception	90.9	0.208
Resnet-18	91.4	0.207
Proposed Model	93.5	0.082

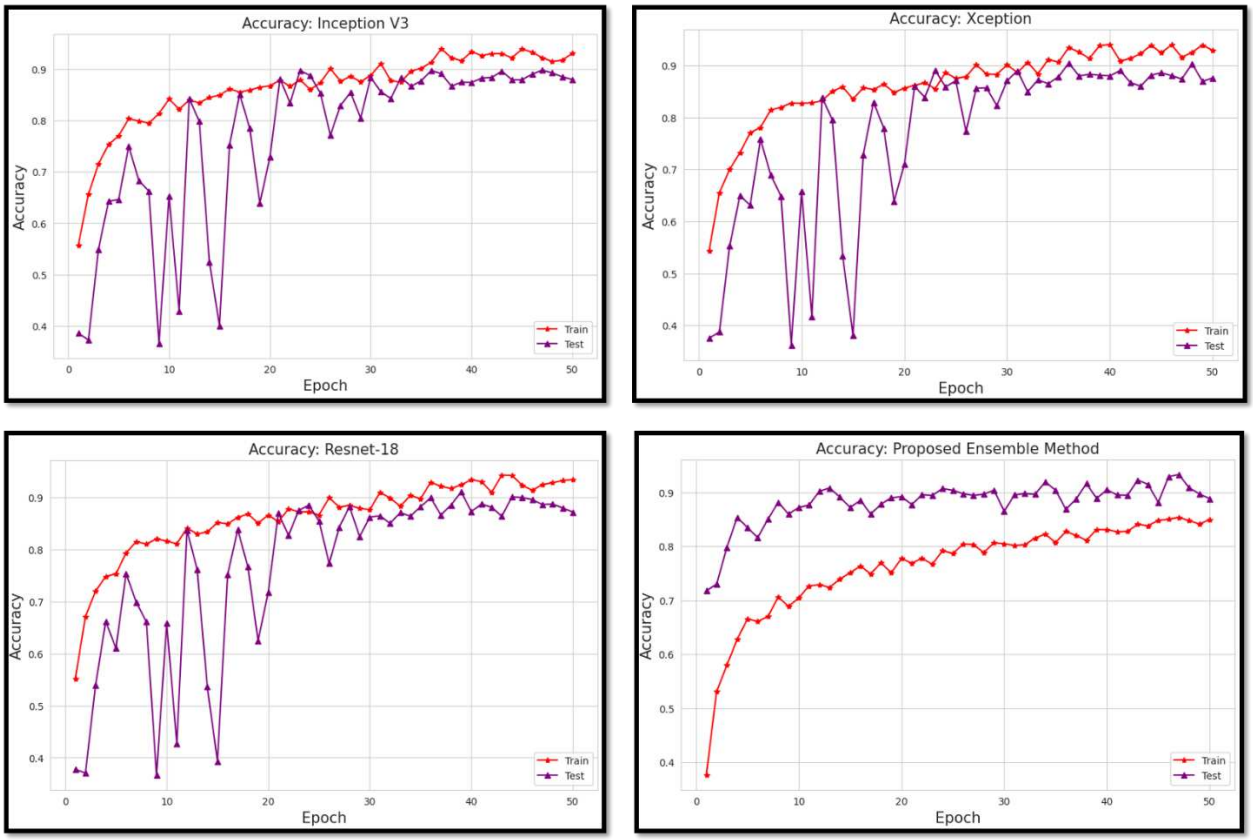

Fig. 2. Accuracy of various models.

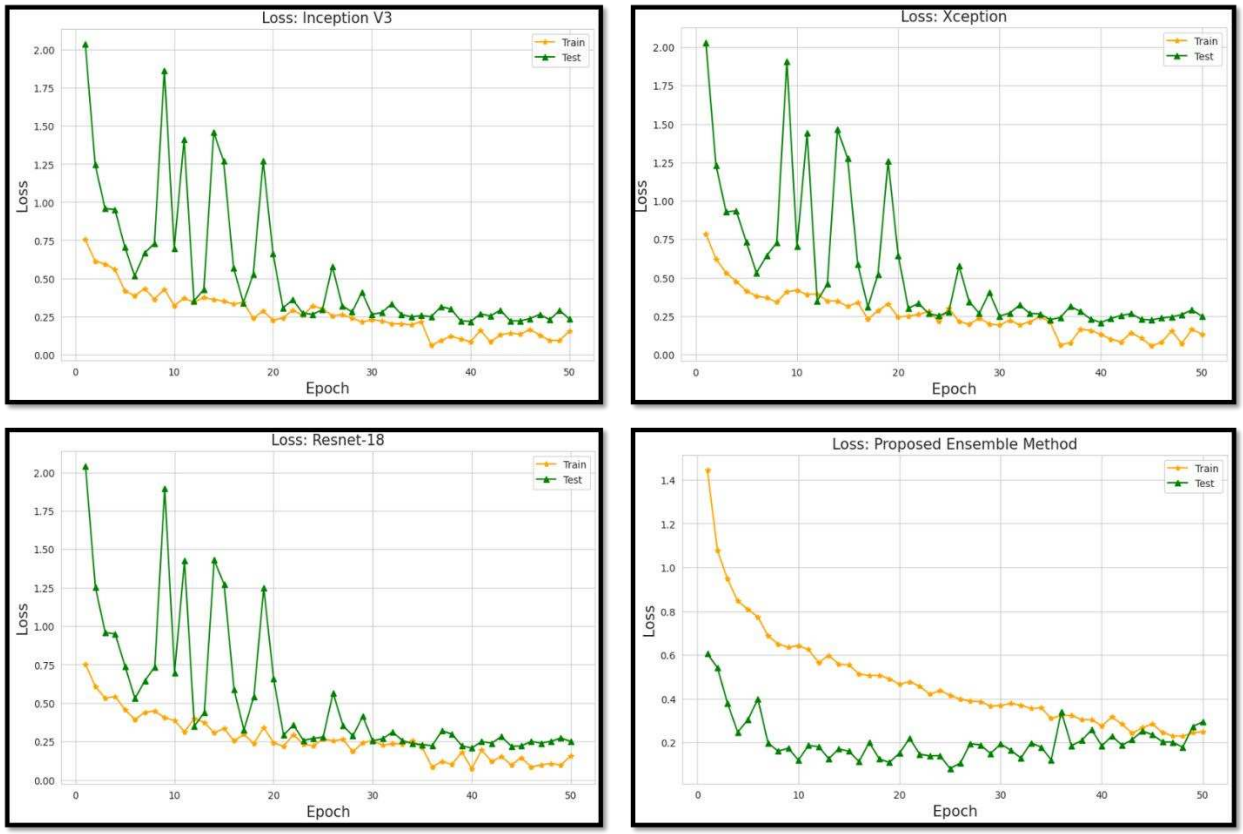

Fig. 3. Loss of various models.

Fig. 4. Confusion Matrices of various models.

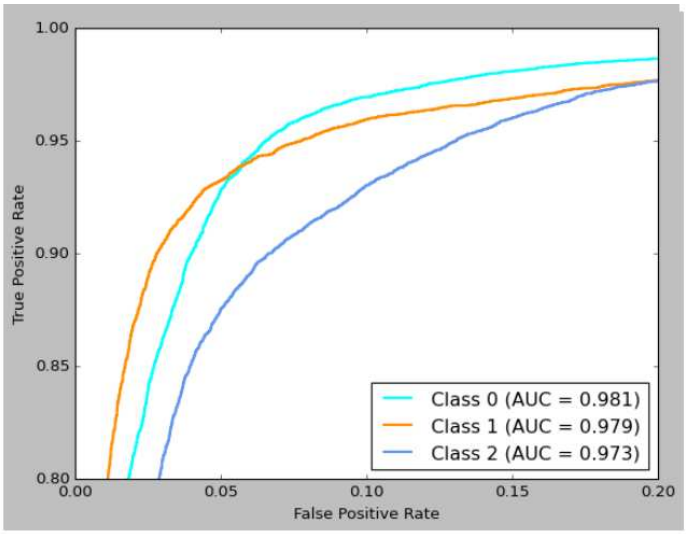

Fig. 5. AUC for the proposed model; class 0 represents normal, class 1 represents osteopenia, and class 2 represents osteoporosis.

V. Conclusion and Future Work

The suggested model provides the best results of 93.5% classification accuracy while assessing osteoporosis knee X-ray dataset classification, which justifies the proposed model's efficacy. Further, the AUC is observed to be 98.1, 97.9 and 97.3 for normal, osteopenia and osteoporosis, respectively. Osteoporosis is a chronic illness that affects people around the globe and can result in long-term loss of mobility, physical injuries, excruciating pain, and even early mortality. In this study, the authors provide a framework for classifying pictures of normal, osteopenia, and osteoporosis by combining three conventional CNN-based classifiers. Two non-linear factors that serve to account for the base learners' level of confidence in their forecasts are used by the proposed ensemble model to construct rankings of the classification models.

The quick detection tool created for testing osteoporosis may operate like a plug-and-play paradigm with minimum assistance from experienced physicians, making it appropriate for use in the field. Due to insufficient picture quality or the existence of overlapping cells, the suggested ensemble model had difficulty correctly classifying a few images. So, the authors intend to tackle the requirement for picture pre-processing in the future. For separating overlapping cells,

authors may use image enhancement methods or cell slicing. To perform the ensemble, authors may additionally take into account ensembles of distinct base learners and investigate various rank-generating procedures.

REFERENCES

[1] N. Wright, A. Looker, K. Saag, J. Curtis, E. Delzell, S. Randall and B. Dawson-Hughes, "The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine," *Journal of bone and mineral research,* vol. 19, no. 11, pp. 2520-2526, 2014.

[2] S. Roux, F. Cabana, N. Carrier, M. Beaulieu, P. April, M. Beaulieu and G. Boire, "The World Health Organization Fracture Risk Assessment Tool (FRAX) underestimates incident and recurrent fractures in consecutive patients with fragility fractures," *The Journal of Clinical Endocrinology & Metabolism,* vol. 99, no. 7, pp. 2400-2408, 2014.

[3] A. Hodsman, W. Leslie, J. Tsang, and G. Gamble, "10-year probability of recurrent fractures following wrist and other osteoporotic fractures in a large clinical cohort: an analysis from the Manitoba Bone Density Program," *Archives of internal medicine,* vol. 168, no. 20, pp. 2261-2267, 2008.

[4] J. Center, T. Nguyen, D. Schneider, P. Sambrook and J. Eisman, "Mortality after all major types of osteoporotic fracture in men and women: an observational study," *The Lancet,* vol. 353, no. 9156, pp. 878-882, 1999.

[5] H. Dimai, "Use of dual-energy X-ray absorptiometry (DXA) for diagnosis and fracture risk assessment; WHO-criteria, T-and Z-score, and reference databases," *Bone,* vol. 104, pp. 39-43, 2017.

[6] D. Hans and S. Baim, "Quantitative ultrasound (QUS) in the management of osteoporosis and assessment of fracture risk," *Journal of Clinical Densitometry,* vol. 3, no. 322-333, p. 20, 2017.

[7] H. Jiang, C. Yates, A. Gorelik, A. Kale, Q. Song and J. Wark, "Peripheral Quantitative Computed Tomography (pQCT) measures contribute to the understanding of bone fragility in older patients with low-trauma fracture," *Journal of Clinical Densitometry,* vol. 21, no. 1, pp. 140-147, 2018.

[8] U. Ferizi, H. Besser, P. Hysi, J. Jacobs, C. Rajapakse, C. Chen, P. Saha, S. Honig and G. Chang, " Artificial intelligence applied to osteoporosis: a performance comparison of machine learning algorithms in predicting fragility fractures from MRI data," *Journal of Magnetic Resonance Imaging,* vol. 49, no. 4, pp. 1029-1038, 2019.

[9] I. Wani and S. Arora, "Computer-aided diagnosis systems for osteoporosis detection: A comprehensive survey," *Medical & biological engineering & computing,* vol. 58, pp. 1873-1917, 2020.

[10] S. Batra and S. Sachdeva, "Organizing standardized electronic healthcare records data for mining," *Health Policy and Technology,* vol. 5, no. 3, pp. 226-242, 2016.

[11] S. Batra and S. Sachdeva, "Pre-processing highly sparse and frequently evolving standardized electronic health records for mining," in *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning*, IGI Global, 2021, pp. 8-21.

[12] S. Sachdeva, D. Batra and S. Batra, "Storage Efficient Implementation of Standardized Electronic Health Records Data," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2020.

[13] C. Court-Brown and B. Caesar, "Epidemiology of adult fractures: a review.," *Injury,* vol. 37, no. 8, pp. 691-697, 2006.

[14] R. Stange and M. Raschke, "Osteoporotic distal femoral fractures: When to replace and how," in *Surgical and Medical Treatment of Osteoporosis*, 2020.

[15] S. Wang, P. Wu, C. Lee, C. Shih, Y. Chiu and C. Hsu, "Association of osteoporosis and varus inclination of the tibial plateau in postmenopausal women with advanced osteoarthritis of the knee," *BMC Musculoskeletal Disorders,* vol. 22, pp. 1-8, 2021.

[16] S. Yadav and S. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big data,* vol. 6, no. 1, pp. 1-18, 2019.

[17] S. Lu, S. Wang and Y. Zhang, "Detecting pathological brain via ResNet and randomized neural networks," *Heliyon,* vol. 6, no. 12, p. e05625, 2020.

[18] S. Batra, H. Sharma, W. Boulila, V. Arya, P. Srivastava, M. Z. Khan and M. Krichen, "An Intelligent Sensor Based Decision Support System for Diagnosing Pulmonary Ailment through Standardized Chest X-ray Scans," *Sensors,* vol. 22, no. 19, p. Sensors, 2022.

[19] S. Batra, R. Khurana, M. Z. Khan, W. Boulila, A. Koubaa and P. Srivastava, "A Pragmatic Ensemble Strategy for Missing Values Imputation in Health Records," *Entropy,* vol. 24, no. 4, p. 533, 2022.

[20] A. Salau and S. Jain, "Adaptive diagnostic machine learning technique for classification of cell decisions for AKT protein," *Informatics in Medicine Unlocked,* vol. 23, p. 100511, 2021.

[21] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[22] A. Pathak, S. Batra and V. Sharma, "An Assessment of the Missing Data Imputation Techniques for COVID-19 Data," in *Proceedings of 3rd International Conference on Machine Learning, Advances in Computing, Renewable Energy and Communication: MARC 2021*, Singapore, 2022.

[23] A. Pathak, S. Batra and H. Chaudhary, "Imputing Missing Data in Electronic Health Records," in *Proceedings of 3rd International Conference on Machine Learning, Advances in Computing, Renewable Energy and Communication: MARC 2021*, Singapore, 2022.

[24] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[25] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017.

[26] S. Ayyachamy, V. Alex, M. Khened and G. Krishnamurthi, "Medical image retrieval using Resnet-18," in *Medical imaging 2019: imaging informatics for healthcare, research, and applications*, 2019.

[27] O. El-Gayar, L. Ambati and N. Nawar, "Wearables, artificial intelligence, and the future of healthcare," in *AI and Big Data's Potential for Disruptive Innovation*, IGI Global, 2020, pp. 104-129.

[28] K. Hatano, S. Murakami, H. Lu, J. Tan, H. Kim and T. Aoki, "Classification of osteoporosis from phalanges CR images based on DCNN," in *2017 17th International Conference on Control, Automation and Systems (ICCAS)*, 2017.

[29] H. Chang, Y. Chiu, H. Kao, C. Yang and W. Ho, "Comparison of classification algorithms with wrapper-based feature selection for predicting osteoporosis outcome based on genetic factors in a Taiwanese women population," *International journal of endocrinology,* vol. 2013, 2013.

[30] N. Tomita, Y. Cheung and S. Hassanpour, "Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans," *Computers in biology and medicine,* vol. 98, pp. 8-15, 2018.

[31] S. Derkatch, C. Kirby, D. Kimelman, M. Jozani, J. Davidson and W. 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, pp. 405-411, 2019.

[32] A. Krishnaraj, S. Barrett, O. Bregman-Amitai, M. Cohen-Sfady, A. Bar, D. Chettrit, M. Orlovsky and E. Elnekave, "Simulating dual-energy X-ray absorptiometry in CT using deep-learning

segmentation cascade," *Journal of the American College of Radiology,* vol. 16, no. 10, pp. 1473-1479, 2019.

[33] Y. Fang, W. Li, X. Chen, K. Chen, H. Kang, P. Yu, R. Zhang, J. Liao, G. Hong and S. Li, "Opportunistic osteoporosis screening in multi-detector CT images using deep convolutional neural networks," *European Radiology,* vol. 31, pp. 1831-1842, 2021.

[34] S. Lee, E. Choe, H. Kang, J. Yoon and H. Kim, "The exploration of feature extraction and machine learning for predicting bone density from simple spine X-ray images in a Korean population," *Skeletal radiology,* vol. 49, pp. 613-618, 2020.

[35] K. Yasaka, H. Akai, A. Kunimatsu, S. Kiryu and O. Abe, "Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network," *European radiology,* vol. 30, pp. 3549-3557, 2020.

[36] N. Sollmann, M. Löffler, M. El Husseini, A. Sekuboyina, M. Dieckmeyer, S. Rühling, C. Zimmer, B. Menze, G. Joseph, T. Baum and J. Kirschke, "Automated opportunistic osteoporosis screening in routine computed tomography of the spine: comparison with dedicated quantitative CT," *Journal of Bone and Mineral Research,* vol. 37, no. 7, pp. 1287-1296, 2022.

[37] J. Lee, S. Adhikari, L. Liu, H. Jeong, H. Kim and S. Yoon, "Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study," *Dentomaxillofacial Radiology,* vol. 48, no. 1, p. 20170344, 2019.

[38] K. Lee, S. Jung, J. Ryu, S. Shin and J. Choi, "Evaluation of transfer learning with deep convolutional neural networks for screening osteoporosis in dental panoramic radiographs," *Journal of clinical medicine,* vol. 9, no. 2, p. 392, 2020.

[39] S. Yu, P. Chu, J. Yang, B. Huang, F. Yang, V. Megalooikonomou and H. Ling, "Multitask osteoporosis prescreening using dental panoramic radiographs with feature learning," *J Smart Health,* 2019.

[40] S. Sukegawa, A. Fujimura, A. Taguchi, N. Yamamoto, A. Kitamura, R. Goto, K. Nakano, K. Takabatake, H. Kawai, H. Nagatsuka and Y. Furuki, "Identification of osteoporosis using ensemble deep learning model with panoramic radiographs and clinical covariates," *Scientific reports,* vol. 12, no. 1, pp. 1-10, 2022.

[41] C. Deniz, S. Xiang, R. Hallyburton, A. Welbeck, J. Babb, S. Honig, K. Cho and G. Chang, "Segmentation of the proximal femur from MR images using deep convolutional neural networks," *Scientific reports,* vol. 8, no. 1, p. 16485, 2018.

[42] J. Liu, J. Wang, W. Ruan, C. Lin and D. Chen, "Diagnostic and gradation model of osteoporosis based on improved deep U-Net network," *Journal of medical systems,* vol. 44, pp. 1-7, 2020.

[43] N. Yamamoto, S. Sukegawa, A. Kitamura, R. Goto, T. Noda, K. Nakano, K. Takabatake, H. Kawai, H. Nagatsuka, K. Kawasaki and Y. Furuki, "Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates," *Biomolecules,* vol. 10, no. 11, p. 1534, 2020.

[44] N. Tecle, J. Teitel, M. Morris, N. Sani, D. Mitten and W. Hammert, "Convolutional neural network for second metacarpal radiographic osteoporosis screening," *The Journal of Hand Surgery,* vol. 45, no. 3, pp. 175-181, 2020.

[45] Q. He, H. Sun, L. Shu, Y. Zhu, X. Xie, Y. Zhan and C. Luo, "Radiographic predictors for bone mineral loss: Cortical thickness and index of the distal femur," *Bone & joint research,* vol. 7, no. 7, pp. 468-475, 2018.

[46] A. Sarwar, V. Sharma and R. Gupta, "Hybrid ensemble learning technique for screening of cervical cancer using Papanicolaou smear image analysis," *Personalized Medicine Universe,* vol. 4, pp. 54-62, 2015.

[47] D. Xue, X. Zhou, C. Li, Y. Yao, M. Rahaman, J. Zhang, H. Chen, J. Zhang, S. Qi and H. Sun, "An application of transfer learning and ensemble learning techniques for cervical histopathology image classification," *IEEE Access,* vol. 8, pp. 104603-104618, 2020.

[48] M. Monwar and M. Gavrilova, "Multimodal biometric system using rank-level fusion approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* vol. 39, no. 4, pp. 867-878, 2009.

[49] J. Deng, W. Dong, R. Socher, L. Li, K. Li and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009.

[50] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015.

[51] D. Kermany, M. Goldbaum, W. Cai, C. Valentim, H. Liang, S. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan and J. Dong, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell,* vol. 172, no. 5, pp. 1122-1131, 2018.