Osteoporosis Detection in Spine using Deep Learning Methods

Dr. K R Shylaja [1] Nithin Suresh [2] Roopitha G Nayak [3] Shreyas D [4] Shrinidhi S Hegde [5]

[1] Professor, Department of Artificial Intelligence and Machine Learning, Dr. Ambedkar Institute of Technology, Bengaluru, KA
[2,3,4,5] Department of Information Science and Engineering, Dr. Ambedkar Institute of Technology, Bengaluru, KA

Abstract

The proposed model "Osteoporosis Detection in Spine using Deep Learning Methods" aims to enhance the diagnosis of osteoporosis by integrating machine learning and deep learning techniques. Traditional bone mineral density (BMD) assessments are insufficient, achieving only a 50% success rate in predicting vertebral fragility fractures (VFF). This project uses a representative dataset of spine DXA Scans and compares various models, including deep learning architectures like CNN, VGG16, and ResNet. The hybrid model obtained by ensembling ResNet with RF Classifier, CNN and VGG16 demonstrated superior performance in identifying patients at risk of VFF. The results highlight significant advancements in accuracy, laying the groundwork for future studies that can improve osteoporosis risk prediction and aid clinical decision-making with better tools for early intervention.

1. Introduction

Osteoporosis is one of the most common bone diseases, characterized by low mass of bone density and the alteration of the microarchitecture, so that the tolerance of bones is reduced and the risk of their fracture is increased. In Osteoporosis, Bone Mineral Density may be decreased; the bone microarchitecture may be disrupted, whereas the concentration and variety of proteins in bones are changed. The sites of classical Osteoporotic fractures are hip, vertebrae, and wrist. Osteoporotic fractures are defined as those occurring at a site associated with low BMD, and which at the same time increased in incidence after 50 years of age. The broken bone is typically the first visible manifestation of Osteoporosis, which has led to it being dubbed the "silent crippler," for sufferers are generally unaware they have osteoporosis until it's too late. Osteoporosis is the slow-developing, but asymptomatic condition, which further makes the fracture of bone an increasing likelihood of bone fragility fractures, in special Vertebral Fragility Fractures (VFF). It is estimated by the International Osteoporosis Foundation that 20% of men and 60% of women over fifty years of age worldwide suffer bone fractures as a result of osteoporosis. Thus, osteoporosis is, therefore, a public health problem increasing the risk of vertebral fractures, decreasing physical functioning, causing immobility, social isolation, and a sense of hopelessness, hence contributing to a low quality of life. The early treatment and prevention of VFF are, therefore, key. Osteoporosis had been among the most occurring diseases in the US, affecting over 3 million patients yearly. Since then, countless attempts to early diagnosis of osteoporosis have been made to reduce bone fracture risks. To address the challenges faced by Osteoporosis, this paper proposes a model that helps in diagnosis of Osteoporosis. The Osteoporosis detection in the spine using deep learning begins with gathering DXA spine image data, followed by data processing techniques such as augmentation, standardization, normalization, annotation, outlier rejection, and de-noising to enhance quality. The processed data is divided into training, validation, and testing datasets, where 70% is used for training, 10% for validation, and 20% for testing. A hybrid deep learning model combining RF with ResNet, CNN and VGG16 is trained to extract features and make decisions. During training, metrics like training accuracy, training loss, validation accuracy, and validation loss are monitored for optimization. The trained model is then tested to predict bone density and diagnose osteoporosis. Finally, after validating its performance, the model is deployed for practical applications where it says whether the person has osteoporosis or not.

2. Related Work

[1] Radiomics Based on Lumbar Spine Magnetic Resonance Imaging to Detect Osteoporosis: The methodology used here is Lumbar Magnetic Resonance Imaging. Here the radiomic models based on lumbar spine DXA were used to detect osteoporosis. The performance of the classification models was evaluated through the estimated area under the receiver using only the characteristic curve.

[2] Frequency of Normal Bone Measurement in Postmenopausal Women with Fracture: Dual-energy X-ray Absorptiometry and Trabecular Bone Score (TBS) is used here. Including TBS with BMD increases the identification of abnormal bone in women with fractures compared with BMD alone.

[3] Association of Bone Mineral Density with Bone Texture Attributes Extracted Using Routine Magnetic Resonance Imaging: Here they do Quantitative textural analysis using Magnetic Resonance Imaging. Texture measurements were used to predict fragility fractures. But the cervical and thoracic spinal segments were not studied.

[4] Data-Driven Diagnosis of Spinal Abnormalities Using Feature Selection and Machine Learning Algorithms: SVM, LR, Bagging SVM, and Bagging LR Models are used. Machine learning algorithms were used for predicting spinal abnormalities. But the computation time of the classifiers was not studied.

[5] Osteoporosis Detection Using Machine Learning Techniques and Feature Selection: Support Vector Machines (SVM), Random Forests, Artificial Neural Networks (ANN), Logistic Regression

(LR) methods are used. Feature selection was used to accurately identify the risk of osteoporosis in postmenopausal women. But the limitation is that structural design was not used for constructing the models. No comparative study of various analytical methods was carried out.

[6] Correlation of Bone Mineral Density with DXAT2*Values in Quantitative Analysis of Lumbar Osteoporosis: Quantitative Computed TomographyT2* Mapping is the method used here. The DXAT2*value is moderately negatively correlated with the bone mineral density assessed with quantitative computed tomography in evaluating osteoporosis in postmenopausal women and may have some potential in assessing the severity of lumbar osteoporosis for scientific research. Gender-specific study, single imaging modality, limited sample size are its limitations.

[7] Ensemble Artificial Neural Networks Applied to Predict the Key Risk Factors of Hip Bone Fracture for Elders: Clinical Data Analysis and Three-Layer Back-Propagation ANN Models are used. Three-layer back-propagation ANN models were developed separately for female and male subjects to predict hipbone fracture risk. The study's findings are based on the specific data set used, which may limit the generalizability of the results to broader populations or different demographic groups.

[8] Artificial Intelligence on the Identification of Risk Groups for Osteoporosis: Hybrid Ensembles Combining Neural Networks and Decision Trees method is used. The selected systems were evaluated based on diagnostic coverage, cost-effectiveness, and their ability to identify significant somatic factors. A comprehensive review of various AI methodologies applied in screening for osteoporosis risk. But combining various databases and clinical attributes presents complexities, particularly in developing standardized, evidence-based criteria for evaluating significant risk factors.

3. System Design and Methodology

The methodology and system design section details the approach and framework applied. The process of developing the system includes selecting appropriate models such as CNN, VGG16, and RESNET with RF, and discusses the data preprocessing, model training, and evaluation processes.

Data pre-processing: Preparing the dataset of DXA (Dual-energy X-ray Absorptiometry) spine scan images for analysis is known as data pre-processing. This starts with data collection from scan centres. Augmentation techniques create more data points to address overfitting. Standardization guarantees consistency in data formats, while normalization brings features to a uniform scale, improving the model's performance. The dataset is labelled for machine learning purposes. Outlier rejection and de-noising procedures are used to eliminate anomalies and improve image clarity, providing reliable input for model training.

Classification: This methodical division ensures dependable performance across various data scenarios, minimizing risks of overfitting or underfitting. The dataset is divided into three subsets: 70% for training, 10% for validation, and 20% for testing. Training data is used to teach the deep learning algorithms, validation data helps to fine-tune the model by evaluating it on unseen data, and testing data confirms the accuracy of the final model.

Deep learning Model: CNN, VGG16, and ResNet are combined with an RF (Random Forest) classifier in the hybrid deep learning model. VGG16's uniform architecture makes implementation easier, CNNs extract spatial features, and ResNet solves the vanishing gradient problem with residual connections. The goal of feature extraction is to accurately detect osteoporosis by extracting important metrics from DXA scans, such as bone texture and bone mineral density (BMD). An ensemble approach is used to process these features in order to produce reliable predictions.

Data partitioning: The data is divided into training (70%), validation (10%), and testing (20%) sets to make training and evaluation easier. The model learns patterns from the training set, optimizes its hyperparameters on the validation set, and is tested to make sure it generalizes well to new data. This partitioning technique is essential to producing a trustworthy and precise model.

Feature Extraction: Feature extraction is pivotal for enhancing predictive accuracy. Techniques focus on deriving BMD and analyzing bone textures like trabecular patterns. Segmentation isolates the region of interest (ROI), while dimensionality reduction methods like Principal Component Analysis (PCA) minimize redundant features. Gradient and edge detection enhance microstructural analysis, providing critical insights into bone health. Model Evaluation: Measures like accuracy, precision, recall, and F1 score are used to test the model during the evaluation phase. Hyperparameter tuning optimizes the model's performance, cross-validation ensures robustness, and real-world testing on external datasets and comparative analysis with baseline models validate the system's reliability. This thorough evaluation guarantees clinical deployment readiness while emphasizing scalability and efficiency.

4. System and Hardware Description

The research was conducted on a system with an Intel Core i5-4460 processor, 4GB of DDR4 RAM, and a 256GB solid-state drive. A standard USB keyboard and mouse were used for input, while a 24-inch monitor with 1920x1080 resolution served as the primary output device. The software environment for this research was built upon Anaconda Distribution version 2.3.1, providing a robust platform for managing dependencies and environments. Key software components included TensorFlow Python version 3.10 for deep learning operations and PyTorch Library version 1.13. This configuration provided an adequate platform for the computational demands of the research, including the analysis of DXA (Dual-energy X-ray Absorptiometry) scans.

5. System Implementation

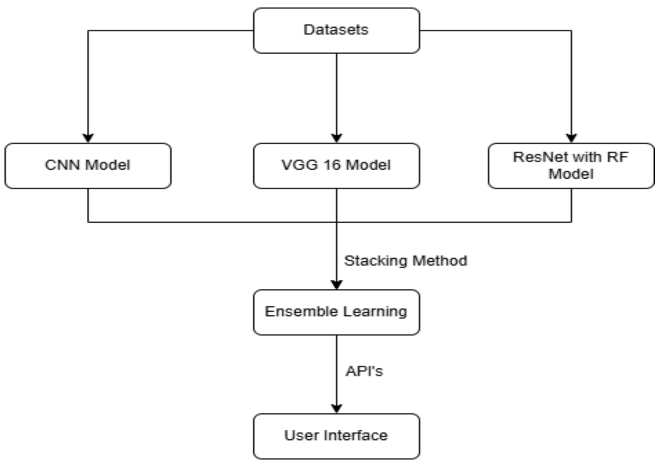

Figure 1: *This diagram illustrates the machine learning pipeline of our model combined with ensemble learning model and APIs.*

The models are trained through iteration to achieve optimum performance in detecting the osteoporotic condition using our own dataset. To create a hybrid model, the outputs of VGG16, CNN, and ResNet are combined. The integration step utilizes FastAPI as a backend framework to facilitate easy interaction between the users and the model. The web application is crafted using HTML and CSS, allowing for a seamless experience for the medical staff. The validation results include accuracy, precision, recall, and F1 score as well as the performance of different models through cross-validation techniques. Then the system is instantiated in a clinical environment for testing purposes. Active monitoring of system performance and retuning gear updates permits the system in providing a trustworthy aid for early diagnosis and treatment of Osteoporosis.

VGG16, CNN, ResNet deep learning models along with Random Forest regression models have been combined in a single system by applying knowledge of Ensemble Learning in order to develop a model that could detect Osteoporosis of the spine. The development cycle begins with gathering data and cleaning spine images to remove any distortion or noise.

Ensemble in simple form means 'a collection of things in a more special sense within the context of Machine Learning terminology Ensemble learning refers to the approach of combining several ML models in order to yield a better and stronger prediction than any of the single model could come up with. It uses an ensemble of fast algorithms (classifiers) of which decision trees are for learning and lets them cast votes.

Thus, now all these models developed such as CNN, ResNet with RF Classifier and VGG 16 are combined together with the help of ensemble learning technique, known as stacking.

Stacking: It employs the first train of some set of models to predict a target and, uses the predictions from the first set as input to the second set of models for generating more accurate predictions. It also allows greater flexibility in combining heterogeneous base models with the commutator being any machine learning algorithm.

But now with this model created, depending on the votes cast, the number of votes cast across all models is taken and a new, more accurate final output is generated that is better than the outputs from the individual models.

Base Models (Level-0 Models):

CNN: The model is programmed to learn patterns using spine images by attempting to reason based on previously engrained spatial and local details.

ResNet with RF Classifier: Again, images are deepened through Resnet and then passed on to the Random Forest Classifier for making a Decision.

VGG16: Used its pre-trained weights for the dataset in further training.

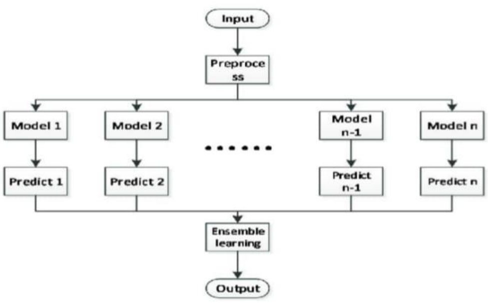

Figure 2: *This diagram illustrates the general workflow of an ensemble learning approach. The input data is preprocessed and then fed into multiple individual models (Model 1 to Model n).*

6. Performance Evolution

The performance evolution of our osteoporosis detection model began with the selection of three key architectures: VGG16, ResNet, and a custom CNN. Each model was trained independently on a curated dataset of spine images obtained from DXA scans. The initial phase involved preprocessing of the data by applying augmentation, normalization, and de-noising techniques to enhance model performance and improve generalization. Various hyperparameters such as learning rate, batch size, and number of epochs were tuned during this phase to identify optimal configurations for each base model.

Once the individual models were trained and validated, their outputs were combined using a stacking ensemble approach. In this approach, the predictions of VGG16, ResNet, and CNN served as inputs to a meta-model, which was designed to learn and improve the final prediction accuracy. Logistic regression was chosen as the initial meta-model due to its simplicity and effectiveness in combining diverse predictions. Several variations of the meta-model, including shallow neural networks, were experimented with to determine the most effective stacking strategy.

To prevent overfitting and enhance model robustness, techniques such as dropout, L2 regularization, and early stopping were applied during training. Furthermore, k-fold cross-validation was employed to ensure that the model performed consistently across different subsets of the data. After obtaining promising results, the final model was subjected to testing on an independent dataset, and its performance was evaluated using metrics such as accuracy, sensitivity, specificity, and F1-score.

CNN model outperforms every other model with accuracy score of 97% whereas accuracy of VGG16 came out to be about 90% and ResNet with RF classifier gave the least accuracy score among other models with the accuracy of 87%. The hybrid model using Ensemble learning techniques particularly Stacking obtained from combing all these trained models gave the accuracy of 93% and is used to predict the risk of Osteoporosis in Spine.

The optimized stacked model demonstrated superior performance compared to individual models, indicating that the ensemble approach effectively leveraged the strengths of each architecture. The final model was then prepared for deployment by converting it into an efficient format suitable for real-time inference.

7. Results

Figure 3: *Home page of the SpineCare web application, featuring an image upload interface for osteoporosis risk prediction.*

Figure 4: *FAQs page of the SpineCare web application.*

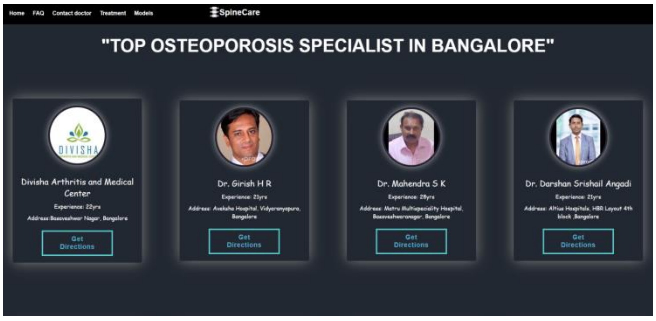

Figure 5: *Section of the SpineCare website showcasing a directory of top osteoporosis specialists in Bangalore with contact information and location details.*

Figure 6: *Section of the SpineCare website dedicated to discussing treatment options for osteoporosis, including medication and lifestyle recommendations.*

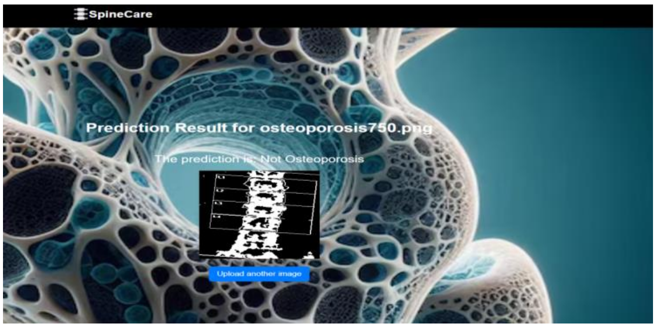

Figure 7: *This is the result screen of the SpineCare web app. It shows that the uploaded image of a spine was analyzed, and the prediction is that the person does not have osteoporosis*

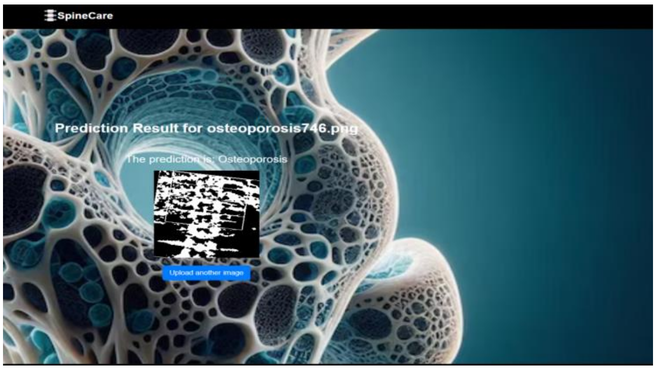

Figure 8: *This is the result screen of the SpineCare web app. It shows that the uploaded image of a spine was analyzed, and the prediction is that the person does have osteoporosis.*

8. Conclusion

This project used sophisticated deep learning models and ensemble techniques to detect osteoporosis. Because of its powerful feature extraction capabilities, CNN showed the highest accuracy; however, it also required a significant amount of data and computational power. VGG16 is easier to use and requires fewer resources, it can be used in settings where interpretability is crucial or where resources are scarce. The combination of a Random Forest (RF) Classifier and ResNet produced a balance between interpretability and depth, with RF enhancing robustness through decision tree ensembles and ResNet's residual learning capturing complex patterns. By training a final model on intermediate predictions from various models, a stacking ensemble model improved performance even more. Individual models were outperformed by this method. The study highlights how to improve medical image analysis and diagnose osteoporosis more accurately in order to improve patient outcomes by customizing model selection based on accuracy, computational requirements, and interpretability.

9. Future Work

The Future Enhancements for the osteoporosis detection system aim to improve its integration, functionality and adaptability. Key areas include:

Clinical Integration: Connecting the model with Clinical Decision Support Systems (CDSS) for automated risk assessment and treatment recommendations.

Combining multimodal imaging (CT, CCT, DXA, ultrasound) to enhance diagnostic accuracy Data Fusion:

AI: Implementing techniques like Grad-CAM or SHAP to improve the transparency of models and improvements trust with clinicians.

Real-Time Processing: Optimizing for faster, edge device deployment and immediate results.

Longitudinal Analysis: Using RNNs or LSTMs to track disease progression over time with sequential images.

Personalized risk: By integrating non-imaging factors like age, BMI and lifestyle for customized risk prediction. Expand the system to detect other skeletal disorders such as scoliosis and fractures.

References

[1] Marko Rak (B) and Klaus-Dietz Tönnies, *A Learning-Free Approach to Whole Spine Vertebra Localization in DXA*, Department of Simulation and Graphics, Otto von Guericke University, Magdeburg, Germany, 2016.

[2] Jonathan S. Ramos, Jamilly G. Maciel†, Mirela T. Cazzolato, Caetano Traina Jr., *BEAUT: A Radiomic Approach to Identify Potential Lumbar Fractures in Magnetic Resonance Imaging*, Institute of Mathematics and Computer Science (ICMC), University of Sao Paulo (USP), Sao Carlos, SP, BR 13566-590 †Ribeirao Preto Medical School (FMRP), University of Sao Paulo (USP), Ribeirao Preto, SP, BR 14048-900, 2021.

[3] Jamilly Gomes Maciel I, Iana Mizumukai de Araújo II, Lucio C. Trazzi I, Paulo Mazzoncini de Azevedo-Marques, *Association of Bone Mineral Density with Bone Texture Attributes Extracted Using Routine Magnetic Resonance Imaging*, Financial Support: JGM: CAPES CODE 1; FJAP: FAPESP 2018/14060-9; MHN-B: FAPESP 2018/04266-9, CNPq 307138/2017-8 and 404795/2018-8, 2020.

[4] Ahmed A. Ewees, Mohamed Abd Elaziz, Rasha M. Arafa, Rania M. Ghoniem, *Improved Approach Based on Fuzzy Rough Set and Sine-Cosine Algorithm: A Case Study on Prediction of Osteoporosis*, supported by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University through the Fasttrack Research Funding Program, 2020.

[5] Jin Liu, Jian-Wei Liao, Wei Li, Xiao-Jun Chen, Jia-Xin Feng, Lin Yao, Pan-Hui Huang, Zhi-Hai Su, Hai Lu, Yu-Ting Liao, Shao-Lin Li, Ya-Jun Ma, *Correlation of Bone Mineral Density with DXA T2 Values in Quantitative Analysis of Lumbar Osteoporosis*, 2020.

[6] Q. Liu, X. Cui, Y.-C. Chou, M. F. Abbod, J. Lin, J.-S. Shieh, *Ensemble Artificial Neural Networks Applied to Predict the Key Risk Factors of Hip Bone Fracture for Elders*, Biomed. Signal Process. Control, vol. 21, pp. 146–156, Aug. 2015.

[7] Agnaldo S. Cruz, Hertz C. Lins, Ricardo V. A. Medeiros, José M. F. Filho, Sandro G. da Silva, *Artificial Intelligence on the Identification of Risk Groups for Osteoporosis: A General Review*, 2021.

[8] Zhiwei Chen, Wenqiang Luo, Qi Zhang, Baiying Lei, Tianfu Wang, Zhong Chen, *Osteoporosis Diagnosis Based on Ultrasound Radio Frequency Signal via Multi-Channel Convolutional Neural Network*, Sun Yat-sen Memorial Hospital Ethics Board (No: SYSEC-KY-KS-2019-159), 2021.