Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/AN16.20260349
Published: June 23, 2026

Deep Learning Prediction of Osteoporosis using Clinical Patient Data

Bedis Elkamel
School of Medicine
Wake Forest University, USA
bedis.elkamel@wfusm.edu

Ali Zahedi
School of Biological Science
University of California Irvine, USA
azahedi1@uci.edu

Hedia Fgaier
Department of Math and Science
Full Sail University, USA
hfgaier@alumni.uoguelph.ca

Marwen Elkamel
Department of Industrial Engineering and Management Systems
University of Central Florida, USA
marwen.elkamel@ucf.edu

Abstract

Osteoporosis is a global public health concern which is characterized by reduced bone mineral density and increased fracture risk, particularly among the elderly population. Early and accurate identification of at-risk individuals is critical for timely clinical intervention. This study investigates the application of deep learning for osteoporosis binary classification using a comprehensive clinical dataset of 1,503 patients which contains demographic information, bone mineral density measurements, serum biochemical indices, therapy records, and comorbidity history. Two deep learning architectures are employed for the classification task: a Multilayer Perceptron (MLP) network and an Autoencoder + Classifier. Model performance was assessed using 5-fold cross validation, and the MLP was superior in the classification task, achieving a test AUC of 0.848. The model found that the most influential variables for Osteoporosis classification were total proximal femur T-score and femoral neck T-score. These results are consistent with clinical diagnostic criteria for osteoporosis, and these findings demonstrate the potential of deep learning as a clinical decision support tool for osteoporosis identification.

Keywords
Deep Learning, Osteoporosis Prediction, Bioinformatics, Bone Mineral Density, Disease Prediction

© IEOM Society International

1. Introduction

Osteoporosis is a chronic skeletal disorder that leads to reduced bone mineral density and bone tissue deterioration, which leads to an increased risk of bone fractures and reduced quality of life (Sözen et al. 2019). Due to its asymptomatic nature during the early stages, osteoporosis is often not diagnosed until significant bone loss or fractures occur. Early detection is essential to reduce long-term complications, healthcare costs, and patient morbidity (Kling et al. 2014). Recent technological advancements have led to increased data collection in healthcare, leading to the development of Electric Medical Records (EMR) systems, which enable the collection and storage of large volumes of patient health data (Wang et al. 2003). These developments have progressed bioinformatics, creating increased opportunities to leverage artificial intelligence (AI) and machine learning (ML) techniques for disease prediction and clinical decision support. In recent years, the application of machine learning algorithms across various domains has increased, including energy forecasting (Elkamel et al. 2022), traffic prediction (Rabelo et al. 2025), materials science (Ulus et al. 2026), financial modeling (Kumar et al., 2025), supply chain management (Balan et al. 2025), and agriculture (Botero-Valencia et al. 2025). This has also led to the ability to process large and complex healthcare datasets, allowing for the identification of relationships among clinical variables, biomarkers, and patient characteristics. Deep learning models have demonstrated strong performance in medical prediction tasks due to their ability to model nonlinear relationships that exist within biomedical data.

In recent years, ML and AI techniques have been increasingly applied in the healthcare sector to support disease diagnosis, risk assessment, patient monitoring, and clinical decision-making. Predictive models have been successfully used in applications such as cancer detection (Meeradevi et al. 2025), cardiovascular disease prediction (Olawade et al., 2025), diabetes diagnosis (Mittal et al. 2025), and medical image analysis (Chen et al. 2025). These methods assist healthcare professionals in analyzing large-scale patient data more efficiently and identify patterns that exist in clinical data, leading to improved patient outcomes, through pattern recognition and early risk identification.

ML and AI techniques have been applied in recent years for osteoporosis prediction and classification. Traditional approaches are very efficient and use dual-energy X-ray absorptiometry (DXA) scanning to measure bone mineral density (BMD), diagnosing osteoporosis and determining an individual's risk of bone fracture. However, DXA is unable to fully capture the complexity of individual fracture risk, as it measures bone density and not bone quality, and can have limited diagnostic capability in certain patient populations. A wide variety of techniques (statistical, machine learning, deep learning) have been utilized to classify osteoporosis diagnosis for patient populations using a wide variety of data. Input data included: DXA derived BMD measurements and T-Scores, demographic variables, biochemical and hematological markers, comorbidity indicators, lifestyle factors, and medical imaging. A Gaussian Process model used to classify osteoporosis diagnosis had an AUC of 0.866 (Amin et al. 2023). A stacked ensemble model that combines logistic regression, adaboost, and gradient boosting was used for osteoporosis diagnosis, and included chronic disease data from 10,000 patients achieved an AUC of 0.76 and identified age, gender, lipid metabolism discords, cancer, and COPD as the top predictive features (Tu et al. 2024). Another study for osteoporosis risk prediction used demographic and clinical data from 8,134 patients to compare several machine learning algorithms (random forest, artificial neural network, k nearest neighbors, support vector machine) with a deep neural network (DNN) and found that the DNN achieved the highest accuracy with an AUC of 0.848, and found that weight, age, gender, and grip strength were the most important predictors (Qiu et al. 2024). A machine learning framework utilized a dataset that included bone density, blood and physical tests of 1493 patients to predict osteoporosis risk, and found that the most accurate model (89% accuracy) used a Forward Feature Selection algorithm (Khanna et al. 2023). Another study focused on identifying osteoporosis risk in postmenopausal woman, and used a dataset of 142 patients, and found that a support vector machine model that uses ANOVA-based feature selection was 94.3% accurate (Bilecik et al. 2026). Other works focused on incorporating image data to classify osteoporosis from knee X-ray images and proposed a ResNet-50 transfer learning framework for a classification task, and achieved an accuracy of 90% (Qureshi et al. 2025). Despite these advances, limited attention has been given to advanced tabular deep learning or to generative approaches like autoencoders for feature learning, particularly on heterogeneous clinical datasets.

This study investigates the use of deep learning techniques to predict osteoporosis using patient clinical information, bone mineral density measurements, and biochemical indicators. The objective is to develop an accurate predictive framework that is capable of supporting early diagnosis and assisting healthcare professionals in identifying individuals at elevated risk of osteoporosis. Early prediction can enable timely intervention, treating acute conditions before they become chronic, reducing patient risk, ensuring improved patient outcomes, and reducing burden on healthcare systems. The objectives of this study are to identify which patient variables and clinical biomarkers have statistically significant associations with osteoporosis diagnosis. In addition, two deep learning architectures for

© IEOM Society International

osteoporosis diagnosis classification will be developed and compared, and these include a MLP network and an Autoencoder + Classifier. Lastly, the most influential predictors of osteoporosis diagnosis will be determined, providing clinically meaningful insights to support provider decision-making.

2. Methods

2.1 Data Collection

A dataset was obtained from the Harvard Database repository containing clinical and biochemical information associated with osteoporosis diagnosis and bone mineral density (BMD) assessment (Linfeng 2022). The dataset contains 1,537 patient records and 40 variables categorized into patient information (Table 1), bone mineral density measurements (Table 2), serum biochemical indices (Table 3), osteoporosis therapy information (Table 4), clinical diagnoses (Table 5), and personal history data (Table 6). Each row in the dataset represents an individual patient, and columns correspond to recorded demographic, physiological, and clinical variables. The dataset contains continuous and binary variables, where continuous variables include patient demographic data, BMD measurements, and biochemical indices; binary variables include therapy-related variables, clinical diagnosis categories, and personal history information. Osteoporosis is included among the clinical diagnosis variables, and is denoted using a binary indicator, representing whether a patient was diagnosed with osteoporosis, and the same binary indicator denotes whether a patient has a clinical diagnosis or uses osteoporosis therapy medications. This variable is later utilized as the target variable for developing classification models aimed at predicting whether a patient has osteoporosis or not. In the clinical diagnosis portion of the dataset, there is a variable for whether a patient has Osteoporosis or not, and this can later become the target variable when building a classification model.

Table 1. Patient Information

Variable	Description
Gender	Biological sex of the patient encoded numerically (Female = 1, Male = 2)
Age	Patient age in years
Height	Height of patient (cm)
Weight	Weight of patient (kg)
BMI	Body Mass Index calculated from height and weight and used as an indicator of body composition

Table 2. Bone Mineral Density Measurements

Variable	Description
L1-L4 (g/cm^2)	Bone mineral density of lumbar vertebrae L1-L4
L1-L4 (T-score)	Standardized T-score comparison of lumbar bone density with healthy reference values
Femoral Neck (g/cm^2)	Bone mineral density of the femoral neck region
Femoral Neck (T-score)	Standardized T-score comparison of femoral neck with healthy reference values
Total Proximal femur (g/cm^2)	Bone mineral density of the total proximal femur
Total Proximal Femur (T-score)	Standardized T-score comparison of the total proximal femur region with healthy reference values

Table 3. Serum Biochemical Indices

Variable	Description
ALT (U/L)	Alanine Transaminase - liver enzyme associated with liver function
AST (U/L)	Aspartate Transferase - liver enzyme used to assess liver health
BUN (mg/dL)	Blood Urea Nitrogen, indicator of kidney function
CREA (*μ*mol/L)	Creatinine concentration associated with kidney performance and muscle metabolism
URIC (μmol/L)	Uric acid concentration in blood serum
FBG	Fasting blood glucose level

© IEOM Society International

HDL-C	High-density lipoprotein cholesterol
LDL-C	Low-density lipoprotein cholesterol
Ca	Serum calcium concentration
P	Serum phosphorus concentration
Mg	Serum magnesium concentration

Table 4. Osteoporosis Therapy Variables

Variable	Description
Calcium	Calcium supplementation status
Calcitriol	Calcitriol treatment information
Bisphosphonate	Use of bisphosphonate medication
Calcitonin	Calcitonin treatment information

Table 5. Clinical Diagnosis Variables

Variable	Description
Hypertension	Presence of hypertension diagnosis
COPD	Presence of chronic obstructive pulmonary disease/chronic bronchitis
DM	Diabetes mellitus diagnosis
Hyperlipidemia	Elevated blood lipid levels
Hyperuricemia	Elevated uric acid levels
Peripheral Atherosclerosis	Presence of arterial narrowing or blockage
Venous Thromboembolism	Presence of blood clot formation in veins
Valve Degeneration	Presence of valve degeneration diagnosis
Osteoporosis	Osteoporosis diagnosis status
Conakry Artery Disease	Presence of coronary artery disease
Chronic Kidney Disease	Presence of kidney disease diagnosis
Fracture	History of fractures

Table 6. Personal History Variables

Variable	Description
Smoking	Smoking history/status
Drinking	Alcohol consumption history/status

2.2. Data Exploration

The dataset was explored to assess variable distributions, identify missing values, detect potential outliers, and identify relationships that may exist between variables, and their relationship with Osteoporosis diagnosis. Table 7 presents summary statistics of patient characteristics, grouped into three categories: overall, no osteoporosis, and osteoporosis. There is a total of 1,537 patient records, where 37% are diagnosed with osteoporosis, and 63% are not. There is a class imbalance ratio of 1.7:1 of osteoporosis patients compared to those without it, which needs to be addressed when building a classification model. Most of the patients in the dataset were females (61.5%), and the average patient age was 59.85. Table 7 lists the summary statistics for all the continuous variables in the collected dataset, and Table 8 lists the percentage of patients with a certain diagnosis, and it is broken down by those who also have osteoporosis, and those who do not.

Table 7. Summary statistics of continuous variables.

Variable	Overall	No Osteoporosis	Osteoporosis
Total	1537	969	568
Age (years)	59.85 ± 12.94	59.39 ± 12.97	60.63 ± 12.87

© IEOM Society International

Height (cm)	165.82 ± 8.09	165.78 ± 8.01	165.90 ± 8.22
Weight (kg)	67.15 ± 11.99	67.10 ± 12.06	67.23 ± 11.89
BMI (kg/m²)	24.31 ± 3.32	24.31 ± 3.33	24.32 ± 3.29
L1-4 BMD (g/cm²)	1.14 ± 0.19	1.14 ± 0.18	1.14 ± 0.19
L1-4 T-score	-0.55 ± 1.53	-0.56 ± 1.50	-0.54 ± 1.58
FN BMD (g/cm²)	0.87 ± 0.15	0.93 ± 0.13	0.76 ± 0.10
FN T-score	-1.30 ± 1.11	-0.82 ± 0.98	-2.12 ± 0.80
TPF BMD (g/cm²)	0.93 ± 0.15	1.00 ± 0.13	0.82 ± 0.12
TPF T-score	-0.93 ± 1.15	-0.44 ± 0.99	-1.78 ± 0.88
ALT (U/L)	23.55 ± 16.62	25.31 ± 18.44	20.56 ± 12.40
AST (U/L)	22.66 ± 9.46	22.91 ± 9.56	22.24 ± 9.29
BUN (mg/dL)	5.62 ± 3.32	5.58 ± 3.87	5.68 ± 2.05
Creatinine (µmol/L)	73.97 ± 25.70	73.80 ± 24.59	74.27 ± 27.52
Uric Acid (µmol/L)	348.11 ± 96.83	353.34 ± 100.16	339.18 ± 90.26
FBG	5.33 ± 1.54	5.43 ± 1.72	5.15 ± 1.15
HDL-C	1.25 ± 0.38	1.23 ± 0.38	1.28 ± 0.38
LDL-C	2.60 ± 0.90	2.65 ± 0.91	2.52 ± 0.88
Calcium (Ca)	2.24 ± 0.16	2.24 ± 0.12	2.24 ± 0.21
Phosphorus (P)	1.04 ± 0.21	1.04 ± 0.23	1.03 ± 0.18
Magnesium (Mg)	0.87 ± 0.10	0.87 ± 0.09	0.87 ± 0.10

Table 8. Summary statistics of binary variables.

Variable	Overall	No Osteoporosis	Osteoporosis
Gender (Female)	945 (61.5%)	597 (61.6%)	348 (61.3%)
Calcium supplement	226 (14.7%)	78 (8.0%)	148 (26.1%)
Calcitriol therapy	263 (17.1%)	71 (7.3%)	192 (33.8%)
Bisphosphonate therapy	94 (6.1%)	29 (3.0%)	65 (11.4%)
Calcitonin therapy	88 (5.7%)	11 (1.1%)	77 (13.6%)
Hypertension	842 (54.8%)	515 (53.1%)	327 (57.6%)
COPD	379 (24.7%)	195 (20.1%)	184 (32.4%)
Diabetes Mellitus (DM)	506 (32.9%)	312 (32.2%)	194 (34.2%)
Hyperlipidaemia	601 (39.1%)	387 (39.9%)	214 (37.7%)
Hyperuricemia	268 (17.4%)	178 (18.4%)	90 (15.8%)
Atherosclerosis (AS)	1152 (75.0%)	678 (70.0%)	474 (83.5%)
Venous Thromboembolism	29 (1.9%)	11 (1.1%)	18 (3.2%)
Valve Degeneration	112 (7.3%)	54 (5.6%)	58 (10.2%)
Coronary Artery Disease	312 (20.3%)	159 (16.4%)	153 (26.9%)
Chronic Kidney Disease	59 (3.8%)	32 (3.3%)	27 (4.8%)
Fracture	31 (2.0%)	11 (1.1%)	20 (3.5%)
Smoking	396 (25.8%)	253 (26.1%)	143 (25.2%)

© IEOM Society International

Drinking	350 (22.8%)	253 (26.1%)	97 (17.1%)

Statistical tests were conducted to determine if there was a statistical difference between Osteoporosis patients and non-Osteoporosis patients for each of the input variables. A summary of the most relevant clinical variables is shown in Figure 1, where the difference between the two groups is shown, and whether there is statistical significance between the two groups, denoted by *, **, or ***, or ns, for no significance. Among the bone density measurements, the difference between groups was statistically significant at $p < 0.001$ for femoral neck T-score, and total proximal femur T-score. However, for lumbar spine T-score, there was no significant difference between the two groups, suggesting that femoral measurement sites are more discriminative for osteoporosis diagnosis for this group of patients. For biochemical markers, ALT was significantly ($p < 0.001$) lower in osteoporosis patients, and uric acid and fasting blood glucose differed significantly between both patient groups. Comorbidities, including atherosclerosis, COPD, and coronary artery disease were significantly ($p < 0.001$) more prevalent among osteoporosis patients, which is consistent with established links between systemic vascular and inflammatory conditions and reduced bone mineral density. In terms of the four osteoporosis therapy variables in the dataset, calcium supplementation, calcitriol, bisphosphonate, and calcitonin were all significantly ($p < 0.001$) more prevalent in the osteoporosis group, which is consistent with clinical practice, where these medications are prescribed to patients following confirmed osteoporosis diagnosis.

2.3 Missing Data

Once the dataset was investigated, we found that 14 variables contained missing variables, ranging from 0.1% to 2.3% of total observations (Figure 2). We noticed there were two types of missing data within the dataset. The first group of missing data was related to 34 patients, where there were no recorded values for any of the four demographic variables (age, height, weight, BMI). Using data imputation techniques was considered not appropriate from a clinician point of view, and these data points were omitted. This left us with 1,503 datapoints. The second group of missing data were related to biochemical lab values, where variables such as HDL, FBG, and LDL had missing values. Multiple Imputation Chained Equations (MICE) (Van Buuren and Groothuis-Oudshoorn 2011) were utilized as the data imputation technique to address these missing data values, and to prepare the dataset for a classification task.

2.4 Data Preprocessing

Following data imputation, a correlation analysis was conducted to determine if there were any correlations present between variables, and to address this and to avoid having multicollinearity. We found correlation values above 0.9 between raw bone mineral density measurements and their corresponding T-scores (Figure 3). T-scores represent the clinically standardized diagnostic measure, thus these variables were retained, while the raw scores were removed. Following the correlation analysis, the dataset was split using an 80/20 train-test split.

© IEOM Society International

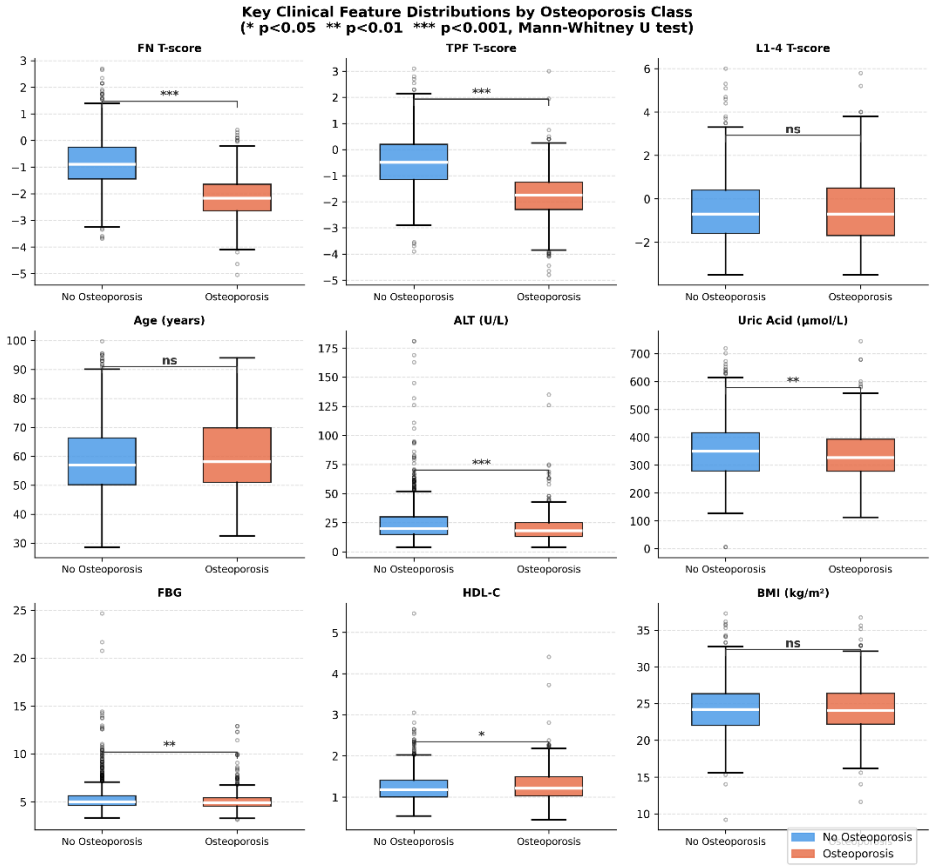

Figure 1. Box plots of relevant clinical variables, comparing patients with and without Osteoporosis.

3. Deep Learning Models

3.1 Multilayer Perceptron (MLP)

The first deep learning model that was developed to classify whether patients have osteoporosis or not, is a feedforward MLP model that consists of four hidden layers. Each hidden layer has a fully connected linear layer, batch normalization, ReLU activation function, and dropout regularization. The developed deep neural network compresses features through hidden dimensions of 256, 128, 64, and 32 neurons respectively, before mapping to a single output classification layer. Dropout rates decrease from 0.4 in the first hidden layer to 0.1 in the final hidden layer, applying stronger regularization in the earlier layers, where the risk of overfitting is greater. The model is trained using Adam optimization with a learning rate of $1x10^{-3}$ and L2 weight decay of $1x10^{-4}$. Training was conducted using a batch size of 32 and 200 epochs, and early stopping was implemented.

© IEOM Society International

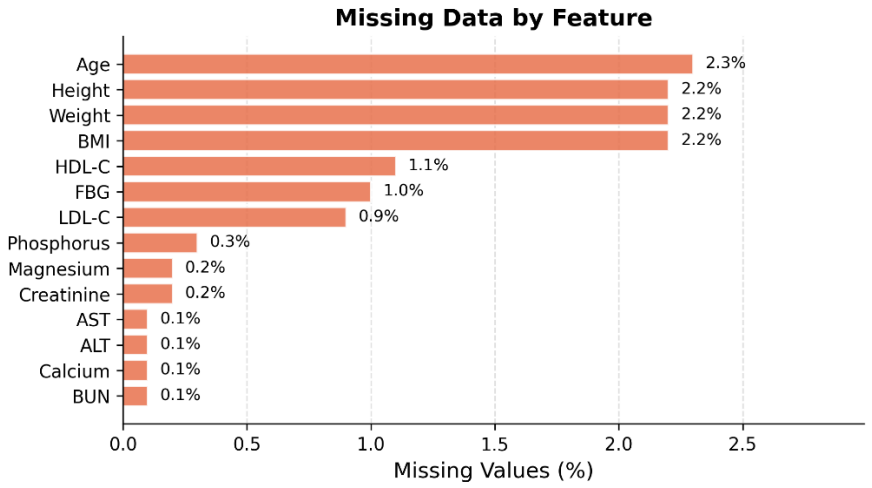

Figure 2. Missing Data present in the dataset sorted by Variable.

3.2 Autoencoder

The second deep learning classification model that was built used a standard autoencoder that was followed by classifier layer for fine-tuning. The two-stage approach begins by learning a compressed representation of the input variables in an unsupervised manner, and then the encoder extracts meaningful information from the data without relying on class labels, potentially providing a more informative starting point for the classifier. The full autoencoder model architecture consists of an encoder and a decoder, where the encoder progressively compresses the input variables through hidden layers (32 and 16), producing an 8-dimensional latent representation of the input data. Each encoder layer applies a linear transformation, batch normalization, ReLU activation function, and dropout. A final ReLU activation function is applied on the latent layer, which constrains the representation to non-negative values. Next, the decoder mirrors the structure, expanding from 8 dimensions back through to 16 and 32 dimensions respectively, and finally to the original dimensions of the input data. The decoder output doesn't have an activation function, and this allows for unconstrained reconstruction of the standardized continuous values. During this stage, the full autoencoder is trained using a mean squared error (MSE) reconstruction loss, and the Adam optimizer is applied with a learning rate of $1x10^{-3}$ and a weight decay of $1x10^{-4}$, and early stopping is implemented. This stage of the autoencoder was trained for a maximum of 150 epochs, and the objective is for the encoder to learn a compact latent representation of the input data that captures the most informative structure.

The second stage of the autoencoder was the classifier fine-tuning. This stage occurs following pre-training, where the decoder is is discarded and the encoder weights are frozen, to prepare for the classification stage. The classification head is attached to the 8-dimensional encoder output, which consists of two hidden layers (16 and 8 neurons) with ReLU activation functions and dropout, and this is followed by a single output. During the second stage, only the classification head parameters are updated, and it is trained using binary cross-entropy, incorporating the same class weights as the MLP model. In this stage, the Adam optimizer was utilized, however a reduced learning rate of $5x10^{-4}$ was applied. This stage was also trained for a maximum of 150 epochs, and early stopping was implemented to prevent overfitting. The objective of freezing the encoder at this stage was to preserve the unsupervised representation of the input data that was learned by the encoder in first stage and to prevent the initialized classification head from corrupting the pretrained encoder weights in early training. The two-stage autoencoder approach forces the encoder to learn which variables are the most informative, representing the overall patient profiles without the risk of overfitting due to the smaller dataset. The compressed 8-dimensional latent representation of the input data allows us to retain the most relevant clinical features while discarding noise. The frozen encoder then provides a stable and information-rich input feature for the second stage, which contains the binary classifier, classifying patients as those with Osteoporosis or not.

© IEOM Society International

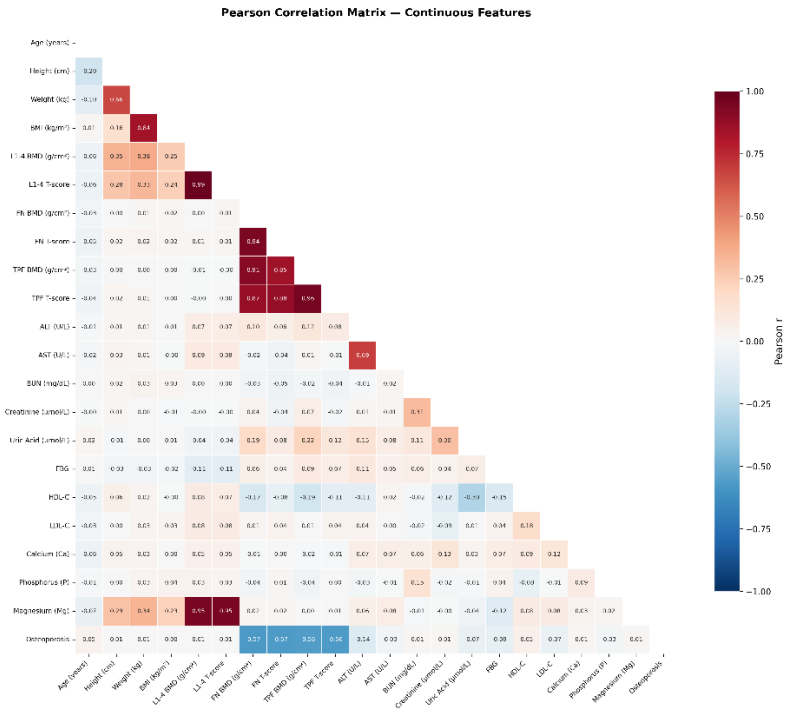

Figure 3. Correlation Matrix of Continuous Variables.

3.3 Model Evaluation

Model performance was evaluated using several classification metrics, including AUC-ROC, F1-Score, accuracy, precision, sensitivity (recall), and specificity. Accuracy measures the overall proportion of correctly classified instances, however, in the presence of class imbalance, it can provide misleading results because a model may achieve a high accuracy while failing to correctly identify minority-class cases. Therefore, additional metrics were incorporated to provide a more comprehensive evaluation of model performance. Precision measures the proportion of predicted positive cases that were correctly identified, while sensitivity (recall) quantifies the model's ability to correctly detect patients with osteoporosis. Sensitivity is particularly important in clinical applications because false negative predictions may lead to missed diagnoses and delayed treatment. Specificity evaluates the ability of the model to correctly identify patients without osteoporosis, reducing unnecessary interventions and false alarms. The F1-score was included because it balances precision and recall and provides a useful assessment when class distributions are unequal. Additionally, AUC-ROC was used to evaluate the model's discriminative capability across different classification thresholds by measuring the tradeoff between true positive and false positive rates. AUC-ROC is used to evaluate the ability of the model to distinguish between classes, where higher AUC values indicate better discrimination capability between osteoporosis and non-osteoporosis cases. Equations (1)-(5) describe the different model evaluation metrics that are used to assess the performance of the model, where True Positives (TP) represent patients that are correctly classified as having osteoporosis, while True Negatives (TN) correspond to patients that are correctly identified as not having osteoporosis. False Positives (FP) represent patients that were incorrectly predicted as having osteoporosis when they did not have it, and False Negatives (FN) correspond to patients with osteoporosis that were incorrectly classified as healthy. In clinical applications, it is important to minimize false negatives, as undetected diagnosis may lead to delayed treatments and worsening patient conditions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Sensitivity\ (Recall) = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

© IEOM Society International

$$F1 = 2\ x\ \frac{Precision\ x\ Recall}{Precision +\ Recall} \tag{5}$$

4. Results and Discussion

4.1 MLP Model

A stratified 5-fold cross-validation framework was used to evaluate the MLP model. The training curve shows consistent decrease in training loss over 31 epochs before early stopping is triggered, and validation AUC was between 0.885 and 0.920. Once training was complete, cross-validation yielded a mean AUC of 0.891, a mean F1-score of 0.770, and low standard deviation across folds, indicating that the model had a stable and consistent performance. The MLP model was then evaluated using the test set (Figure 6), and it achieved an AUC of 0.849, sensitivity of 0.777, specificity of 0.762, F1-score of 0.713, and an accuracy of 0.767. Investigating the confusion matrix, we find that there are 87 TP, 25 FN, 144 TN, and 45 FP. 92 of 112 osteoporosis patients were correctly diagnosed, while there were 46 false positives (Figure 4).

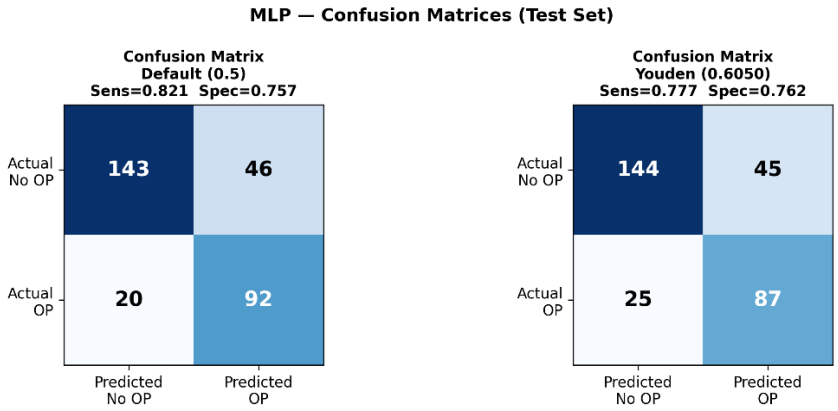

Figure 4. Confusion Matrix for MLP Model.

4.2 Autoencoder and Classifier

The Autoencoder and Classifier model was evaluated using a stratified 5-fold cross-validation followed by a final evaluation on the held-out test set. During the first stage, the autoencoder reconstruction loss converged smoothly across all five folds, and the average validation MSE was consistently below the training MSE, indicated that the encoder successfully learned a compressed representation of the clinical feature space without overfitting. In the 2nd stage, featuring the classifier head, there was rapid convergence, with validation AUC stabilizing within the first 5-10 epochs and training loss continued to decrease over 35 epochs before early stopping was triggered. Cross-validation results yielded a mean AUC of 0.869 and a mean F1-score of 0.729, indicating reasonable and consistent classification performance across folds. When assessing the model on the test set, the model achieved an AUC of 0.814 (Figure 6), sensitivity of 0.670, specificity of 0.778, F1-score of 0.655, and an accuracy of 0.738. The confusion matrix shows 75 TP, 37 FN, 147 TN, and 42 FP, out of 301 test patients. This model was able to correctly identify 89 out of 112 Osteoporsis patients, however, there were 56 false positive predictions among 189 non-osteoporosis patients (Figure 5).

4.3 Comparative Analysis

Table 9 summarizes the testing performance of both models presented in this paper, the MLP and the Autoencoder with a classifier. The MLP model achieved a higher AUC, higher sensitivity, and higher F1-score. The Autoencoder was slightly better in terms of specificity. Both models perform well in classifying osteoporosis diagnosis, especially when we consider the limited training set size of 1,202 samples. When comparing the cross-validation results with the test set, we find that the MLP has better generalization, compared to the autoencoder, and this is likely due to the benefit of using a supervised training approach, compared to the two-stage autoencoder, where the model has to reconstruct the compressed data before classification. The ROC of both models is shown in Figure 6.

© IEOM Society International

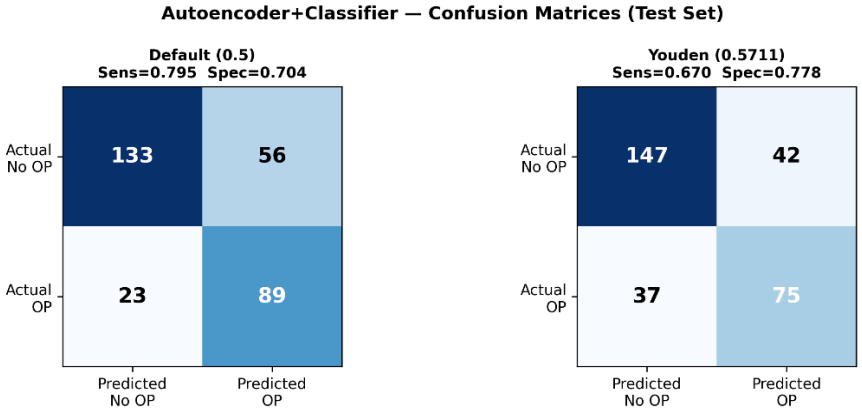

Figure 5. Confusion Matrix for Autoencoder + Classifier Model.

Table 9. Model Evaluation Criteria of Deep Learning Classification Models

Metric	MLP	Autoencoder + Classifier
CV AUC	0.8906	0.8692
Test AUC	0.8488	0.8149
Sensitivity	0.777	0.670
Specificity	0.762	0.778
F1-Score	0.713	0.655
Accuracy	0.767	0.738

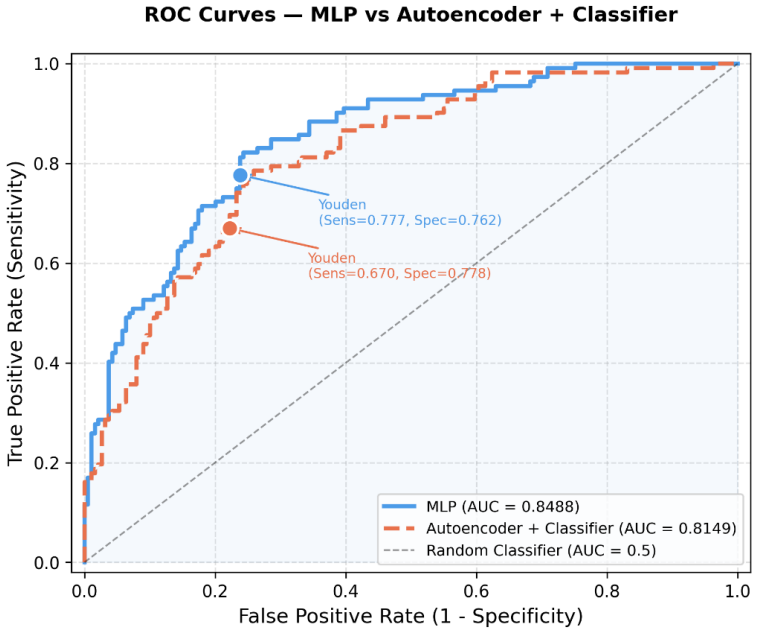

Figure 6. ROC Curve Comparison between MLP and Autoencoder + Classifier Model for Osteoporosis Classification.

4.4 Discussion

Following the development of the classification models for Osteoporosis diagnosis, the best performing model was taken, and variable importance was computed to determine the most influential variables. SHAP analysis was applied to the MLP model, and the most influential variables are shown in Figure 3. Total proximal femur T-score (0.129) and

© IEOM Society International

femoral neck T-score (0.104) were the most influential variables contributing a greater influence than the remaining input variables. Atherosclerosis and Calcitriol therapy were ranked third and fourth respectively. The results show the association between cardiovascular disease and reduced bone mineral density as well as the clinical practice of prescribing Calcitriol to patients with confirmed osteoporosis. ALT ranked fifth, and this supports the clinical hypothesis that liver function plays a role in bone metabolism through its involvement in vitamin D activation and bone health.

5. Conclusion

This study presented two deep learning models for osteoporosis classification using clinical patient data. The two deep learning architectures that were investigated included a MLP network and an Autoencoder + Classifier. Missing data was imputed using MICE imputation, and a stratified cross-validation approach was used during training. The MLP demonstrated superior performance with a test AUC of 0.8488, sensitivity of 0.777, and specificity of 0.762, while the Autoencoder + Classifier achieved an AUC of 0.8149 with higher specificity of 0.778. SHAP interpretability analysis confirmed that total proximal femur and femoral neck T-scores were the most influential predictors, validating the clinical relevance of the learned representations. Notably, lumbar spine measurements showed negligible predictive contribution despite being commonly used in clinical practice, suggesting that femoral site measurements may be more informative for automated risk stratification in this population.

Future work will explore additional deep learning architectures for osteoporosis diagnosis classification; including different configurations of MLP networks, autoencoder as, and transformer-based models. These models will be evaluated to provide a comprehensive comparison of deep learning approaches for clinical tabular data. Bayesian hyperparameter optimization will be applied across all models to determine the best network configuration for the osteoporosis classification task. SHAP analysis will be expanded to provide patient-level interpretability supporting clinical decision-making.

References

Amin, R., Reza, M. S., Maniruzzaman, M., Hasan, M. A. M., Lee, H. S., Jang, S. W., and Shin, J., Intensive statistical exploration to identify osteoporosis predisposing factors and optimizing recognition performance with integrated GP kernels, *IEEE Access*, vol. 11, pp. 131338-131350, 2023.

Balan, G. S., Kumar, V. S., and Raj, S. A., Machine learning and artificial intelligence methods and applications for post-crisis supply chain resiliency and recovery, *Supply Chain Analytics*, vol. 10, p. 100121, 2025. https://doi.org/10.1016/j.sca.2025.100121

Bilecik, N. A., Uğur, L., Öten, E., and Çapraz, M., Machine Learning-Based Automatic Diagnosis of Osteoporosis Using Bone Mineral Density Measurements, *Journal of Clinical Medicine*, vol. 15, no. 2, p. 549, 2026.

Botero-Valencia, J., García-Pineda, V., Valencia-Arias, A., Valencia, J., Reyes-Vera, E., Mejia-Herrera, M., and Hernández-García, R., Machine Learning in Sustainable Agriculture: Systematic Review and Research Perspectives, *Agriculture*, vol. 15, no. 4, p. 377, 2025.

Chen, C., Isa, N. A. M., and Liu, X., A review of convolutional neural network based methods for medical image classification, *Computers in Biology and Medicine*, vol. 185, p. 109507, 2025.

Elkamel, M., Ahmadian, A., and Zheng, Q. P., Impact of coronavirus disease 2019 on electricity demand and the unit commitment problem: a long–short-term memory-based machine learning approach, *Engineering Optimization*, vol. 54, no. 11, pp. 1835-1852, 2022.

He, L., Bone mineral density, *Literature Review/Technical Monograph*, 2022.

Khanna, V. V., Chadaga, K., Sampathila, N., Chadaga, R., Prabhu, S., K. S., S., ... and Bhat, D., A decision support system for osteoporosis risk prediction using machine learning and explainable artificial intelligence, *Heliyon*, vol. 9, no. 12, p. e22975, 2023.

Kling, J. M., Clarke, B. L., and Sandhu, N. P., Osteoporosis Prevention, Screening, and Treatment: A Review, *Journal of Women's Health*, vol. 23, no. 7, pp. 563–572, 2014.

Kumar, S., Rao, A., and Dhochak, M., Hybrid ML Models for Volatility Prediction in Financial Risk Management, *International Review of Economics & Finance*, p. 103915, 2025.

Meeradevi, T., Sasikala, S., Murali, L., Manikandan, N., and Ramaswamy, K., Lung cancer detection with machine learning classifiers with multi-attribute decision-making system and deep learning model, *Scientific Reports*, vol. 15, no. 1, p. 8565, 2025.

© IEOM Society International

Mittal, R., Weiss, M. B., Rendon, A., Shafazand, S., Lemos, J. R., and Hirani, K., Harnessing machine learning, a subset of artificial intelligence, for early detection and diagnosis of type 1 diabetes: a systematic review, *International Journal of Molecular Sciences*, vol. 26, no. 9, pp. 3935, 2025.

Olawade, D. B., Soladoye, A. A., Omodunbi, B. A., Aderinto, N., and Adeyanju, I. A., Comparative analysis of machine learning models for coronary artery disease prediction with optimized feature selection, *International Journal of Cardiology*, vol. 436, p. 133443, 2025.

Qiu, C., Su, K., Luo, Z., Tian, Q., Zhao, L., Wu, L., ... and Shen, H., Developing and comparing deep learning and machine learning algorithms for osteoporosis risk prediction, *Frontiers in Artificial Intelligence*, vol. 7, p. 1355287, 2024.

Qureshi, M. B., Sani, M., Raza, A., Qureshi, M. S., Alghamdi, W., Akbar, S., ... and Sarker, M. R., Deep-learning based osteoporosis classification in knee X-ray images using transfer-learning approach, *Scientific Reports*, vol. 15, no. 1, p. 38448, 2025.

Rabelo, L., Rincón-Guio, C., Laynes, V., Gutierrez-Franco, E., Bhat, V., Zamora-Aguas, J., and Elkamel, M., Effective last-mile delivery using reinforcement learning and social media-based traffic prediction in underdeveloped megacities, *Discover Cities*, vol. 2, no. 1, p. 69, 2025.

Sözen, T., Özışık, L., and Çalık Başaran, N., An overview and management of osteoporosis, *European Journal of Rheumatology*, vol. 4, no. 1, pp. 46–56, 2019.

Tu, J. B., Liao, W. J., Liu, W. C., and Gao, X. H., Using machine learning techniques to predict the risk of osteoporosis based on nationwide chronic disease data, *Scientific Reports*, vol. 14, no. 1, p. 5245, 2024.

Ulus, N., Sahin, V., Elkamel, M., Orhan, O. Y., and Ersan, H. Y., Metal nanosites-confined hierarchical zeolite for enhanced formic acid dehydrogenation, *Molecular Catalysis*, vol. 591, p. 115714, 2026.

Van Buuren, S. and Groothuis-Oudshoorn, K., mice: Multivariate imputation by chained equations in R, *Journal of Statistical Software*, vol. 45, pp. 1-67, 2011.

Wang, S. J., Middleton, B., Prosser, L. A., Bardon, C. G., Spurr, C. D., Carchidi, P. J., ... and Bates, D. W., A cost-benefit analysis of electronic medical records in primary care, *The American Journal of Medicine*, vol. 114, no. 5, pp. 397-403, 2003.

Biographies

Bedis Elkamel is a medical student at Wake Forest University School of Medicine and a graduate of the University of Central Florida (UCF), where he earned a degree in Health Sciences with a 4.0 GPA. He was selected as a recipient of UCF's prestigious Order of Pegasus, the university's highest student honor, recognizing exceptional academic achievement, leadership, research, and community service. His interests include public health, preventive medicine, healthcare disparities, and scientific research. Bedis has also been actively involved in mentoring underserved youth, healthcare volunteering, and child advocacy initiatives.

Ali Zahedi is an undergraduate student studying Biological Sciences at the University of California, Irvine with interests in healthcare, public health, and computational biology. He has developed professional and research experience through work at UCI Health and as an NSF REU Intern at Iowa State University. In the future, Ali plans to pursue a career in healthcare focused on improving patient outcomes and advancing scientific innovation.

Hedia Fgaier is currently a Professor of Mathematics at Full Sail University. Prior to this she was a Lecturer at the University of Waterloo and an Assistant Professor of Applied Mathematics at Al-Ain University of Science & Technology. Dr. Fgaier holds a PhD and a Master degree in applied mathematics from the University of Guelph, ON, Canada. Her research interests lie in the areas of dynamical systems, computer simulation, parameter estimation, and optimal control with applications to biology and medicine. Dr. Fgaier envisions her research to be a blend of theoretical investigations, development of computational methods, and the building and analysis of mathematical models of nonlinear systems. She has published in peer review journals such as Journal of Theoretical Biology and Computers & Chemical Engineering. She has participated in national and international conferences and workshops.

Marwen Elkamel is a Research Associate in the Department of Industrial Engineering and Management Systems at the University of Central Florida (UCF), specializing in AI-driven infrastructure systems, smart supply chains, and sustainable energy systems. He earned his Ph.D. in Industrial Engineering from UCF and has conducted research in simulation, optimization, machine learning, and resilient urban systems. His work spans renewable energy, logistics, microgrids, and the food-energy-water nexus, with publications in leading journals and presentations at major

© IEOM Society International

international conferences. Marwen is also an active mentor and educator, having supervised numerous industry-sponsored engineering design projects in collaboration with aerospace, defense, and logistics partners.

© IEOM Society International