

# Using Machine Learning Models Trained on Lab Tests to Predict Osteopenia, Osteoporosis, and Spine Fractures

George Asrian<sup>1</sup>, Fikry Nuhuman<sup>1</sup>, Kathy Jia<sup>1</sup>, Zhihan Xiab, Rashad Madi, Bella Xia<sup>1</sup>, Chamith S. Rajapakse, PhD<sup>1</sup>

<sup>1</sup>University of Pennsylvania, Philadelphia, PA, 19104  
[gmasrian@sas.upenn.edu](mailto:gmasrian@sas.upenn.edu)

## Disclosures: N/A

**INTRODUCTION:** Osteopenia, which is a weakening of the bones due to decreased bone mineral density, and osteoporosis, which is a more severe stage of bone mineral density loss causing more brittle bones, are conditions that become much more prevalent with age. As spine fractures represent a significant burden on health systems and often lead to poor outcomes. Vertebral compression fractures significantly increase the risk of death, and as the US population ages, the prevalence of spine fractures will continue to increase. It is the goal of this research to investigate whether lab test values can be used to train machine learning classification algorithms that predict osteopenia, osteoporosis, and spine fractures 5 and 10 years into the future. The coefficients assigned to each lab test were assessed to evaluate which tests are most closely associated with each outcome.

**METHODS:** The machine learning multivariate classification models trained for this study were the random forest classifier, decision tree classifier, gradient boosting classifier, K nearest neighbors classifier, adaptive boosting classifier, light gradient boosting machine classifier, extra trees classifier, logistic regression, and linear discriminant analysis classifier. The input data set included over 100,000 patients admitted to the Beth Israel Deaconess Medical Center hospital system between 2008 and 2019. However, because less than 2% of these patients experienced either osteopenia, osteoporosis, or spine fractures within 10 years of the lab test date, oversampling was employed to correct for the imbalanced data. 117 unique lab tests were evaluated for osteoporosis and osteopenia, and 89 unique lab tests were evaluated for spine fractures. Age was also used as an input variable. Data was sourced from MIT's MIMIC IV (Medical Information Mart for Intensive Care) database. All patient data used was previously de-identified.

**RESULTS:** For predicting osteopenia and osteoporosis, the gradient boosting classifier (GBC) performed best. For the 5-year prediction, the training set accuracy of 92.4% and AUC of 0.722. This model achieved similar performance on the test set, with accuracy of 92.7% and AUC of 0.720. When training GBC to predict osteopenia and osteoporosis cases 10 years into the future, performance decreased only modestly, with training set accuracy of 91.7% and AUC of 0.734. The blood tests with the greatest predictive power for osteoporosis and osteopenia were iodine levels, magnesium levels, bicarbonate, anion gap, red blood cell distribution width (measuring variability in red blood cell volume), creatinine, sodium, basophils, and mean corpuscular volume. The linear discriminant analysis algorithm performed best for predicting spine fractures, though results were more modest. For the 5-year prediction, the training set accuracy of 67.7%, AUC of 0.725, sensitivity of 64.9%, and specificity of 67.2%. This model achieved similar performance on the test set, with accuracy of 67.2% and AUC of 0.724. When training LDA to predict spine fractures 10 years into the future, performance decreased only modestly, with training set accuracy of 66.4% and AUC of 0.713. This 10-year prediction model achieved test set accuracy of 66.1% and AUC of 0.702. The blood tests with the greatest predictive power for spine fractures were specific gravity urine test, reticulocyte count, serum protein, absolute monocyte count, magnesium, globulin, transferrin, absolute eosinophil count, basophil count, and bilirubin.

**DISCUSSION:** The results show that relatively strong accuracy and predictive power for forecasting osteoporosis and osteopenia, with the gradient boosting classifier performing the best, whereas significantly weaker performance in predicting spine fractures. This is likely due to external causes of spine fractures, such as accidents and sports injuries, which are not related to physiological factors described by lab tests. The coefficients used in the prediction models that performed the best can be used to identify which lab tests are most predictive of these outcomes 5 years and 10 years into the future. The results suggest lab test data and machine learning may be able to aid in predicting and staging diseases of bone mineral density loss with more specificity. The limitations of this study include missing lab test values (which were handled by imputing means), using significantly imbalanced data, and not stratifying the results based on more fundamental severity criteria such as age or comorbidities.

**SIGNIFICANCE:** Knowledge of the lab tests that are predictive of osteopenia, osteoporosis, and spine fractures can help clinicians better understand which medical conditions and states can lead to weakening bones and a significant decrease in bone mineral density. By better understanding the patients who are at a greater risk of brittle bones and fractures, more precautions can be taken for at-risk patients, and more fractures can be prevented. This would reduce the burden on the health system and reduce mortality.