

# Predicting Hip Fractures Within 5 and 10 Years Using Machine Learning Models Trained on Lab Test Data

George Asrian<sup>1</sup>, Kathy Jia<sup>1</sup>, Fikry Nuhuman<sup>1</sup>, Zhihan Xaib, Rashad Madi, 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:** Hip fractures represent one of the most serious fall injuries, presenting significant challenges for aging populations. Each year over 300,000 older people (aged 65 and up) are hospitalized from hip fractures. As the US population ages, the prevalence of hip fractures will continue to increase. It is the goal of this research to investigate whether lab test data can be used to predict future hip fractures using machine learning. The incidence of a hip fracture occurring 5 years and 10 years after the time of each lab test was predicted using multiple machine learning classification algorithms, and the coefficients assigned to each lab test were assessed to evaluate which tests are most closely associated with future hip fractures.

**METHODS:** 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 were used to model prediction algorithms. The input data set included 109,228 unique patients admitted to the Beth Israel Deaconess Medical Center hospital system between 2008 and 2019. However, because less than 1% of these patients experienced a hip fracture 5 years or 10 years from the lab test date, oversampling was employed to correct for the imbalanced data. 89 unique lab tests were evaluated, and these tests plus age were used as features in the modeling. Data was sourced from MIT's MIMIC IV (Medical Information Mart for Intensive Care) database. All patient data used was previously de-identified.

**RESULTS:** Among the machine learning models tested, the logistic regression classifier performed best. For the 5-year prediction, the training set accuracy of 72.4%, AUC of 0.763, sensitivity of 69.6%, and specificity of 70.7%. This model achieved even greater performance on the test set, with accuracy of 72.4% and AUC of 0.806. When training logistic regression to predict hip fractures 10 years into the future, performance remained strong, with training set accuracy of 71.4% and AUC of 0.763, sensitivity of 68.6%, and specificity of 70.7%. This 10-year prediction model achieved test set accuracy of 71.1% and AUC of 0.750.

**DISCUSSION:** The results show that relatively strong accuracy and predictive power can be achieved by using machine learning models trained on lab tests and patient age, with logistic regression performing the best. The coefficients used in the prediction models that performed the best can be used to identify which lab tests are most predictive of hip fractures 5 years and 10 years later. The limitations of this study include missing values in the data set (which were handled by imputing means), using significantly imbalanced data, and not stratifying the results based on a more fundamental severity criteria such as age or comorbidities.

**SIGNIFICANCE:** Knowledge of the lab tests that are predictive of hip fractures can help clinicians better understand which medical conditions may increase the risk of hip fractures and which elevated biomarkers clinicians should look out for. By better understanding the patients that are at a greater risk of hip 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 following a fall.