

# Assessment of Machine Learning Classifiers for Fracture Risk Prediction compared to FRAX without T-score

Ed Sykes<sup>1</sup>, Ravi Jain<sup>2</sup>, Neil Sano<sup>1</sup>, Ha Neul Moon<sup>1</sup>, Jennifer Weldon<sup>2</sup>, Rohit Shanker<sup>3</sup>, Volodymyr Voytenko<sup>1</sup>, Jansen Sullivan<sup>3</sup>, Doug Sauer<sup>3</sup>  
<sup>1</sup>Centre for Mobile Innovation, Sheridan College, Oakville, <sup>2</sup>Ontario Osteoporosis Strategy, Osteoporosis Canada, Toronto, <sup>3</sup>Inovex, Oakville, Canada

## INTRODUCTION

### Background:

- Fracture risk assessments are essential to evaluate and prevent osteoporotic fractures. Bone Mineral Density (BMD) T-score is one of the inputs for fracture risk calculators like FRAX or CAROC used in Canada. In the absence of T-scores, FRAX calculation can still be used. However, this requires up to 11 variables to be entered into the calculator. To overcome this, Machine Learning (ML) models using patients' self-reported clinical risk factors excluding T-scores were used to predict fracture risk levels. The predictions from the above models and FRAX without T-score were compared to the actual fracture risk level from FRAX with T-score.

### Purpose of the study:

- Create an ML model that rivals FRAX without T-score in high fracture risk level prediction among fracture patients.

## METHODS

### Sample:

- Feature analysis was performed on a preliminary sample dataset of 818 patients from the Ontario Fracture Screening and Prevention Program (FSPP) database provided by the Ontario Osteoporosis Strategy to determine the linear relationship of the features and the T-score.
- Permutation importance was calculated in these models.
- Selenium was used to calculate patients' FRAX risk level if missing.
- Synthetic Minority Oversampling Technique (SMOTE) was used to balance training data.

### Analysis:

- The sample data augmented by other Ontario FSPP data reports, osteoporosis research articles and advice from osteoporosis experts were used to create a Cleaning Script that was used to process the 29 features found in the dataset.
- All models included the patient's age, weight, height, sex, alcohol intake, tobacco use, and oral steroids use as initial features.
- The following ML algorithms were explored to be compared to FRAX without T-score: Random Forest Classifier (RFC), Support Vector Classifier (SVC), and Stacking Classifier consisting of RFC, SVC, Multi Layer Perceptron (MLP), and Logistic Regressor (LR).

## DATA PROCESSING

### Data Quality:

- The cleaning process consisted of removing duplicates, eliminating irrelevant data, performing type conversion, fixing syntax errors, filling in missing values with mean, mode, and 0, and dealing with outliers. After the cleaning, there were 818 patient records available for ML training and testing.

### Selecting Features:

- Although 29 features were selected for development, we primarily focused on data features used for FRAX calculation. FRAX uses 12 features to determine risk level which are age, sex, weight, height, previous fracture, parent fracture, smoking, glucocorticoids, arthritis, secondary osteoporosis, alcohol intake, and T-score.
- To compare the performance of FRAX without T-score, we used the same 11 features in our models.

## PREDICTIONS

- The RFC model had a sensitivity of 93.8% for high risk level, which is significantly higher than 68.7% of FRAX without T-score.
- The RFC model had a precision of 55.6% for high risk level due to moderate risk levels getting overpredicted to high risk level.
- The RFC model had the precisions of 92.1% for low and 79.2% for moderate risk level.
- While SVC and Stacking models performed better in predicting moderate patients correctly, they tended to lose sensitivity in low and high risk level in comparison to RFC and FRAX without T-score.
- FRAX without T-score's performance weakened as the risk level went higher, where precision and sensitivity for moderate and high risk level dropped 20-30% compared to low-risk level.

## CONCLUSIONS

- The RFC model offered the most balanced performance and was best at classifying high-risk patients correctly.
- The RFC model performed 25.1% higher than FRAX without T-score in classifying high risk patients correctly.
- Although RFC's precision in high risk level is 55.6%, this is due to over diagnosing moderate risk level patients to high risk level, which is preferable to underdiagnosing high risk level patients to moderate risk level.
- Our next step is to test this against FSPP data extract (~ 7800 patients) to see if the model retains its edge against FRAX without T-score.

## RESULTS



Model	Precision			Sensitivity		
	Low	Moderate	High	Low	Moderate	High
RFC	82.1	79.2	55.6	91.4	35.2	93.8
SVC	56.6	83.5	53.1	65.2	79.5	50.7
Stacking	60.5	79.0	51.9	50.0	77.1	59.7
FRAX <u>without T-score</u>	82.4	63.0	61.0	90.0	51.8	68.7

\* Precision and Sensitivity for risk levels in %

Acknowledgement: This study was supported through funding from NSERC (grant #06227). The Ontario Osteoporosis Strategy is funded by the Ontario Ministry of Health. The views expressed are those of the stakeholders and do not necessarily reflect those of the Ministry. Fracture Screening and Prevention Program (Ontario FLS) is implemented by Osteoporosis Canada.

Sheridan Centre for Mobile Innovation

ONTARIO Osteoporosis Strategy

inovex