

Evaluating Machine Learning techniques used to predict Bone Mineral Density T-scores from patient-reported Clinical Risk Factors

Ed Sykes¹, Ravi Jain², Neil Sano¹, Ha Neul Moon¹, Jennifer Weldon², Rohit Shanker³, Volodymyr Voytenko¹, Jansen Sullivan³, Doug Sauer³

¹Centre for Mobile Innovation, Sheridan College, Oakville, ²Ontario Osteoporosis Strategy, Osteoporosis Canada, Toronto, ³Inovex, 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. However, some patients do not have ready access to diagnostic imaging facilities or might experience extended wait times. Predicting T-scores could work as an initial screening tool to save time and resources for identifying patients who are likely to be at high risk.

Purpose of the study:

- Create a Machine Learning (ML) model that predicts the BMD T-score based on patients' self-reported features and identifies feature importance in the tested models.
- Feature permutation importance identifies which features impacted the predictions the most. Feature permutation importance shows how much performance metric such as Root Mean Squared Error (RMSE) would change when that certain feature was shuffled in the training set.

METHODS

Sample:

- Feature analysis was performed on a remote dataset of 7801 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.
- Polynomial features were created and experimented with for features like the patient's age, sex, and BMI.
- Numerical features have been scaled with Scikit-Learn's Standard Scaler.
- Permutation importance was calculated in these models to distinguish which features would have significance in predictions.

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.
- The following ML models were explored to compare predicted T-score values to actual values: Linear Regression, Ridge Regression, Bayesian Ridge, CatBoost Regression, Random Forest Regressor, and Deep Neural Networks (DNN).

DATA PROCESSING

Data Quality:

- The cleaning process consisted of removing duplicates and 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 7801 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 CAROC and FRAX calculation. CAROC uses 5 features while FRAX uses 12 features. Although both risk assessment tools incorporate the previous fracture as one of the risk factors, they can be used for patients without fracture history.
- All models included the patient's age, weight, height, sex, alcohol intake, tobacco use, and oral steroids use as initial features. In some models, features were split between the most common fracture sites (i.e., wrist and shoulder).

PREDICTIONS

- DNN with the wrist fracture site with a dropout rate of 20% performed the best with an RMSE of 0.85, the lowest of all models.
- Linear regression models had low R-Squared scores between 0.13-0.16, which showed that the features that were being provided to the model have only a minor correlation to a patient's t-score.
- Permutation importance revealed that in average sex (0.54), age (0.48), BMI (0.46), and weight (0.45) affected models' performance the most.

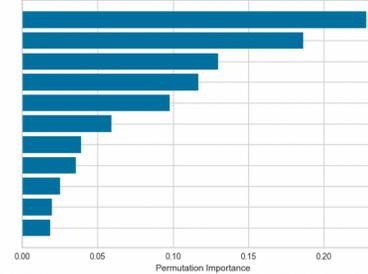
CONCLUSIONS

- The best performing model was DNN with the wrist fracture site with a dropout rate of 20%.
- The models rely heavily on the patient's age, height, weight, and sex in making predictions.
- Three of the above 4 mentioned main features are numerical values and had the greatest impact on a patient's predicted t-score.
- Considering the RMSEs, R-Squared scores, and the permutation feature importance, the models require more numerical data other than the patient's age, height, and weight to improve predictions.

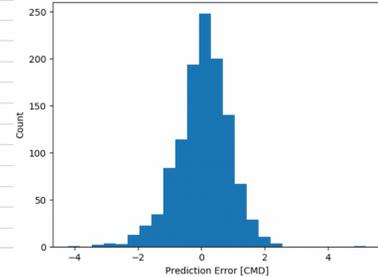
Acknowledgement: This study was supported through funding from NSERC (grant #06227). The Ontario Osteoporosis Strategy is funded by the Ontario Ministry of Health. The views expressed are those of the stakeholders and do not necessarily reflect those of the Ministry. Fracture Screening and Prevention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

RESULTS

Permutation Importance in RMSE (wrist fracture site with a dropout rate of 20%)



Prediction Error Count (wrist fracture site with a dropout rate of 20%)



Models and Metrics (RMSE and R-Squared)

Model	RMSE	R-Squared
Linear Regression	0.940	0.130
Ridge Regression	0.940	0.131
Bayesian Ridge	0.940	0.131
CatBoost Regression	0.922	0.163
Random Forest Regressor	0.926	0.156
DNN (wrist)	0.848	
DNN (shoulder)	0.944	
DNN (female)	0.896	
DNN (male)	1.000	

Sheridan | Centre for Mobile Innovation

ONTARIO Osteoporosis Strategy

inovex