CNN-Based Evaluation of Wrist Bone Health and Fracture Risk Using Raman Spectroscopy of the Proximal Phalanx

Anthony Yosick, MS[1,2], Mohammad Hosseini[3], Sadia Afrin[1], Sophia Turbide, MD[4], Andrew Berger, PhD[1,3], Hani Awad, PhD[1,2]

[1]Department of Biomedical Engineering, [2]Center for Musculoskeletal Research, [3]The Institute of Optics, University of Rochester, Rochester, NY, [4]Department of Internal Medicine and Pediatrics, University at Buffalo, Buffalo, NY

ayosick@ur.rochester.edu

Disclosures: Anthony Yosick (N), Mohammad Hosseini (N), Sadia Afrin (N), Sophia Turbide (N), Andrew Berger (N), Hani Awad (N)

INTRODUCTION: Dual-energy X-ray absorptiometry (DXA) is clinically used for quantifying bone mineral density (BMD) loss. The T-score, which measures the number of standard deviations from normal healthy BMD, is the World Health Organization's (WHO) gold standard for diagnosing osteoporosis (OP). Alternatively, the Fracture Risk Assessment Tool (FRAX) considers additional patient factors to estimate a 10-year probability of a major osteoporotic fracture (wrist, hip, humerus, and spine). Although established as clinical standards, these tools overlook the compositional qualities that contribute to fracture toughness. To address these limitations, Raman spectroscopy (RS) has long been investigated as a method for non-ionizing, non-destructive molecular fingerprint of bone. Traditional RS analysis often focuses on parametrizing select spectral components, such as mineral to matrix ratio and carbonate to phosphate ratio. While these parameters capture key features of bone quality, much of the spectral information remains unused, limiting a comprehensive assessment of bone quality. With the advent of artificial intelligence (AI) approaches, such as convolutional neural networks (CNN), we hypothesized that the full-spectrum RS can be leveraged to improve OP diagnosis and fracture risk assessment. Specifically, this study aims to evaluate the sensitivity and specificity of CNN-based analysis of the full-spectrum RS of the proximal phalanx bones to predict the wrist (1/3 radius) T-scores and wrist fracture toughness, compared to RS parameter-based regression approaches. The proximal phalanx bones were chosen for their accessibility, minimal soft tissue coverage, and suitability for rapid, non-invasive Raman measurements, making them highly practical for clinical translation if their composition reflects systemic skeletal properties at major sites of fragility fractures.

METHODS: Human cadaveric specimens were obtained through Anatomy Gifts Registry (AGR). Given the higher osteoporosis risk in biological females, this cohort comprised of 25 female (mean age of 70.2±14.8 years and a mean BMI of 26.5±7.7 kg/m^2) non-paired cadaver forearms disarticulated at the elbow. BMD classification was determined from distal radius DXA scans (1/3 radius) performed on a Horizon Ci DXA System (Hologic) by Advanced Radiology, which determined that n=8 were normal (N), n=6 were osteopenic (OPE), and n=11 were OP. First, biomechanical testing protocols simulating a fall on an outstretched hand (FOOSH) were adapted from previous studies [1,2]. Forearms were mounted in pronation with 15° radial abduction and loaded at a rate of 3.3 mm/s using an Instron ElectroPuls E10000. These consistently resulted in clinically relevant radial fractures, from which the force-displacement data were used to determine the work to fracture as a measure of fracture risk. Following the biomechanical testing, the second proximal phalanx specimens were dissected and rehydrated in PBS for 2-hours prior to collecting spatially offset Raman spectroscopy (SORS) using an 830 nm laser delivering 100 mW with scattered light acquisition at 0- and 3-mm offsets. Raw spectra were preprocessed and fluorescence removed using an iterative algorithm and normalized to the phosphate (ν_1PO_4) peak [3]. Peak area ratios were calculated for phosphate mineral to matrix (PTMR; 924-986 cm^{-1}), carbonate to matrix (CTMR; 1054-1098 cm^{-1}), and carbonate to phosphate (CTPR) with Amide I (1596-1720 cm^{-1}) as the matrix component [2]. Predictive modeling included a partial least square regression (PLSR) leave-one-out cross-validation (LOOCV) model using PTMR, CTMR, CTPR, age, and BMI. Alternatively, a CNN adapted from a previously optimized RS model [4] was trained using a LOOCV approach on the full Raman spectrum (744-1740 cm^{-1}) with age and BMI also included as inputs, replacing the classification layer with regression for continuous output (Figure 1A). Both models directly predicted T-score and work to fracture per donor. Donor-level values were obtained by averaging all spectral results and classifying into N (T-score > -1), OPE (-2.5 < T-score ≤ -1) and OP (T-score ≤ -2.5), as well as low (≥19 J) versus high (<19 J) wrist fracture risk, with fracture thresholds measured from the FOOSH simulation [5]. Averaged 0- and 3-mm spectra for wrist T-score classification and fracture risk are illustrated in Figure 1B–C. <u>Statistics:</u> T-score and work to fracture were predicted through PLSR-LOOCV and CNN-LOOCV models and evaluated using root-mean-squared error of cross-validation (RMSECV) and coefficient of determination (R^2). Classification performance for WHO categories and low versus high fracture risk was assessed using receiver operating characteristic (ROC) curves with area under the curve (AUC) as the evaluation metric.

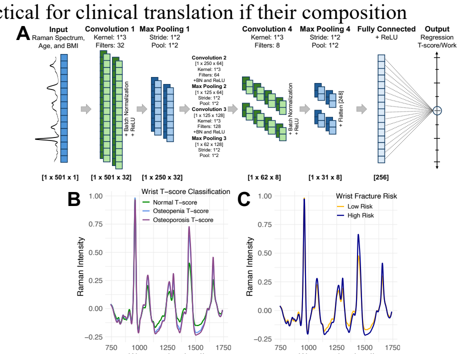

Figure 1. (A) Convolutional neural network (CNN) model architecture for Raman spectral analysis. Average proximal phalanx Raman spectra grouped by (B) wrist T-score classification and (C) wrist fracture risk.

RESULTS: PLSR model (PTMR, CTMR, CTPR, age, BMI) and CNN model (full Raman spectra 744-1740 cm^{-1}, age, BMI) predictions of T-score and work to fracture are presented in Figures 2A–B and 3A–B. PLSR achieved the strongest prediction of T-score (R^2 = 0.779; RMSECV = 0.923), while CNN-LOOCV showed lower accuracy for T-score prediction (R^2 = 0.664; RMSECV = 1.137) but outperformed PLSR in predicting work to fracture (R^2 = 0.151; RMSECV = 7.926 J). Both models differentiated WHO categories with high accuracy, as illustrated in Figure 2C, with PLSR achieving higher overall accuracy. Importantly, classification of N versus OP donors yielded nearly identical performance for the two approaches, with AUCs of 1.000 for PLSR and 0.989 for CNN. For fracture risk classification illustrated in Figure 3C, CNN achieved better results, with an AUC of 0.873 compared to 0.833 for PLSR.

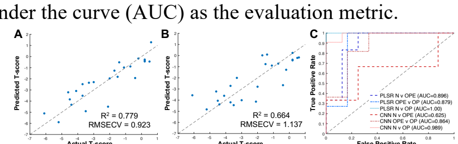

Figure 2. Predicted vs. actual wrist T-scores using (A) PLSR and (B) CNN models of proximal phalanx SORS. (C) Receiver operating characteristic (ROC) curves for WHO classifications, showing high discrimination for N vs. OP specimens (AUC = 1.000 for PLSR, 0.989 for CNN).

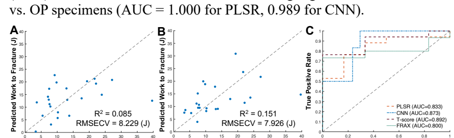

Figure 3. Predicted vs. actual wrist fracture risk using (A) PLSR and (B) CNN models of proximal phalanx SORS. (C) Receiver operating characteristic (ROC) curves for low vs. high fracture risk classification, showing improved discrimination with CNN (AUC = 0.873) compared with PLSR (AUC = 0.833).

DISCUSSION: Our findings demonstrate that both parameter-based and full-spectrum SORS of the proximal phalanx are effective for predicting wrist T-score and fracture risk. Both models achieved a near-perfect AUC when comparing normal versus osteoporotic specimens. While prediction of work-to-fracture values was less accurate, the full-spectrum CNN model outperformed FRAX and was comparable to T-score for fracture risk classification (Figure 3C). Notably, RS measurements obtained remotely from the phalanx accurately predicted wrist T-score and fracture risk. The accuracy observed is particularly promising for extending this approach to other major sites of fragility fractures including the femoral neck and the spine. While limitations include the use of exposed phalanx samples and a single FOOSH loading orientation, ongoing work in our lab is focused on enabling accurate transcutaneous SORS acquisition of bone spectra. Altogether, these results provide compelling evidence for SORS within clinical workflows for osteoporosis screening, longitudinal bone health monitoring, and early fracture risk assessment.

SIGNIFICANCE/CLINICAL RELEVANCE: It is estimated that 1 out of 3 women over the age of 50 are expected to experience a fragility fracture, however, only approximately 20% of women over the age of 50 receive a DXA screening [6,7].

ACKNOWLEDGEMENTS: We would like to acknowledge Natnael Ayalew for assistance and funding from NIH/NIAMS (R01AR07061).

REFERENCES: [1] Lochmüller, E. M., et al., J Bone and Miner Res, (2002) 17(9), 1629–1638. [2] Massie, C., et al., J Biomech, (2023) 161, 111852. [3] Lieber, C. A., et al., Appl Spectrosc (2003), 57(11), 1363–1367. [4] Wan, Y., et al., Spectrochim Acta A Mol Biomol Spectrosc, (2025), 329, 125608. [5] Abdolshah, S., et al., J Musculoskelet Neuronal Interact, (2020) 20(2), 176–184. [6] Singer, A., et al., Arch Osteoporos, (2023) 18(1), 42. [7] Gillespie, C. W., et al., Am J Med (2017) 130(3), 306–316.