

Application of a Neural Network Classifier to Radiofrequency-Based Osteopenia/Osteoporosis Screening

Johnathan Adams^{1*}, *Student Member, IEEE*, Ziming Zhang¹, *Member, IEEE*, Gregory M. Noetscher^{1,2},
Member, IEEE, Ara Nazarian^{3,5†}, and Sergey N. Makarov^{1,2,4†}, *Senior Member, IEEE*

Abstract—This preliminary study reports application of a neural network classifier to the processing of previously collected data on low power radiofrequency propagation through the wrist with the goal to detect osteoporotic/osteopenic conditions. The data set used includes 67 subjects (23-94 years old, 50 females, 17 males, 27 osteoporotic/osteopenic, 40 healthy). We process the entire spectrum of the propagation coefficient through the wrist from 30 kHz to 2 GHz, with 201 sampling points in total. We found that the dichotomic diagnostic test of raw non-normalized radiofrequency data performed with the trained neural network approaches 90% specificity and ~70% sensitivity. These results are obtained without inclusion of any additional clinical risk factors. They justify that the radio transmission data are usable on their own as a predictor of bone density. With the inclusion of additional clinical risk factors, both specificity and sensitivity improve to 95% and 76% respectively. Our approach correlates well with the available DXA measurements and has the potential for screening patients at risk for fragility fractures, given the ease of implementation and low costs associated with both the technique and the equipment.

Clinical Relevance— Dichotomic diagnostic test of raw non-normalized radiofrequency data performed with the trained neural network approaches 90% specificity and ~70% sensitivity. With the inclusion of other clinical risk factors, specificity and sensitivity increase to 95% and 76% respectively.

I. INTRODUCTION

Approximately 50% of women and 20% of men over the age of 50 will suffer from a fragility fracture in their remaining lifetime [1]. Hip fracture is one of the most serious and debilitating outcomes of osteoporosis [2], with a 14–36% mortality rate during the first-year post fracture [3]. Hip fracture incidence rates are known to increase exponentially with age in both women and men [4]. The number of fractures is predicted to double or triple by 2040 [5].

The World Health Organization (WHO) has defined individuals at risk for these fractures based on their *areal* bone mineral density (aBMD, g/cm²) relative to that of a normal young adult, as measured by Dual-energy X-ray Absorptiometry (DXA). The disadvantages of DXA include: exposing patients to ionizing radiation doses of up to 0.86

mrem [6]; the surrounding soft tissues can introduce relevant measurement errors [7]; bone mineral density (BMD) measurements are affected by variations in bone size [8]; and cortical and trabecular bone cannot be separated [9]. Additionally, fracture predictions based on aBMD have been shown to be neither sensitive nor specific [10],[11].

Microwave or radiofrequency imaging of (heel) bone was first introduced by Dr. Keith Paulsen and his research group at Dartmouth College approximately ten years ago as an alternative non-ionizing diagnostic method to assess bone health [2]. Due to the well-known complexity and poor spatial resolution of the standard microwave imaging setup [12] used in these studies, no clinically applicable results have been generated to date. However, the underlying physical idea of this method is simple and powerful. In osteoporosis, bone mass decreases and pore size increases. The lost bone mass is replaced by a mixture of yellow bone marrow. Such substantial changes in physical properties alter electromagnetic tissue properties [13] and generate a significantly different radiofrequency (RF) channel through the bone.

To measure this, we have selected the wrist, a body compartment where bone constitutes a significant fraction of the total tissue volume and is easily accessible. We have measured radio wave propagation through this compartment and compared our results with osteoporotic and osteopenic (low bone density) conditions established via DXA and through a history of bone fracture [14].

In the present study, we have included additionally collected subject data (7 new subjects) and have employed a neural network approach to process the previously obtained and new data for improved results, especially those obtained without the inclusion of clinical risk factors. We hypothesize that the incorporation of a neural network analysis will significantly improve the predictive power of the presented system compared to the current method based on a simple threshold binary classifier approach.

¹Department of Electrical and Computer Engineering, Worcester Polytechnic Institute, Worcester, MA 01609, USA

²Neva Electromagnetics, LLC., Yarmouth Port, MA 02675, USA

³Musculoskeletal Translational Innovation Initiative, Carl J. Shapiro Department of Orthopaedic Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA

⁴Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA

⁵Department of Orthopedic Surgery, Yerevan State Medical University, Yerevan, Armenia

[†]These authors have contributed equally as senior authors.

*Correspondence to jwadams2@wpi.edu

This work was supported by the NIH SBIR Phase I Grant R43AR071220 NIAMS and NSF CCF-2006738.

II. MATERIALS AND METHODS

A. Device Concept

The device concept is illustrated in Fig. 1a. Two antennas (Fig. 1b) [14] are placed on the superior and inferior flat sides of the wrist adjacent to the position of the ulnar head, under an applied controlled pressure of 1 kg force. The radiofrequency signal in the 0 – 2GHz band travels from the transmit antenna through bone, cartilage, and soft tissue to the receiving antenna while being attenuated and scattered. The total amount of attenuation and scatter is measured via the microwave transmission coefficient $S_{21}(f)$ and is correlated to osteopenic and osteoporotic conditions. The antenna width across the wrist is 2 cm, and the antenna length along the wrist is 5 cm; facilitating good contact between the two surfaces. Details of the design of the device including numerical simulations have been previously published in [14].

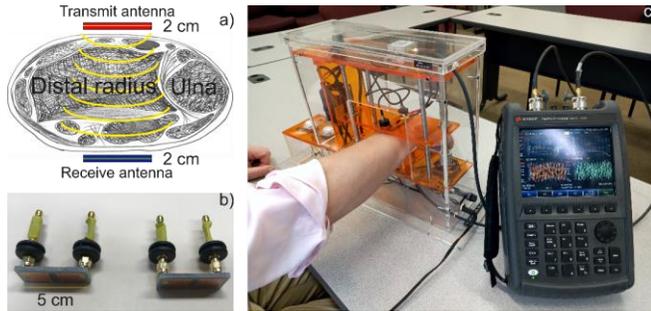


Fig. 1. a) – Idealized diagram illustrating antenna placement on both sides of a human wrist. b) – Transmit and receive dual antiphase patch antennas with individual lumped-component matching networks. c) – Wrist tester device demonstration.

B. Data Set

After receiving Institutional Review Board (IRB) approval through Worcester Polytechnic Institute, written informed consent was obtained from 80 subjects to participate in this study (age range 23-94 years old, 60 female, 20 male, 3 African American, 4 Hispanic, 64 Caucasian, 9 declined to state). All measurements were further performed in accordance with the relevant IRB guidelines and regulations. 72 subjects were measured in a previous study [14], and 8 new subjects were added for this study. From 80 subjects, we selected 67 subjects suitable for a dichotomous diagnostic set:

1. **Group 1 osteopenic/osteoporotic** (DXA T-score between -1.0 and -2.4 for osteopenic subjects and below -2.4 for osteoporotic subjects (within the last year) and prescribed medications, 55-90 years old, mean $77.5/STD$ 10.1). **27 subjects** in total (24 female, 3 male).
2. **Group 2 healthy** (low risk category, 23-94 years old, mean $60.2/STD$ 16.6). Unknown bone density (no DXA data) but either young adults or having no history of bone fractures, no medication, and no family history of osteoporosis. **40 subjects** in total (26 female, 14 male). We are comfortable considering these subjects at low risk without explicit BMD information because the clinical risk factors above can have a larger impact on fracture risk than one standard deviation decline in bone density [15].

C. Raw Radiofrequency Data

Fig. 2A shows the magnitude of the transmission coefficient, $|S_{21}(f)|$ for 201 frequency sampling points between 300 kHz and 2.0 GHz. Group 1 is plotted in red and Group 2 is plotted in blue. There is significant overlap between the two groups data, especially between 1.8 and 2.0 GHz.

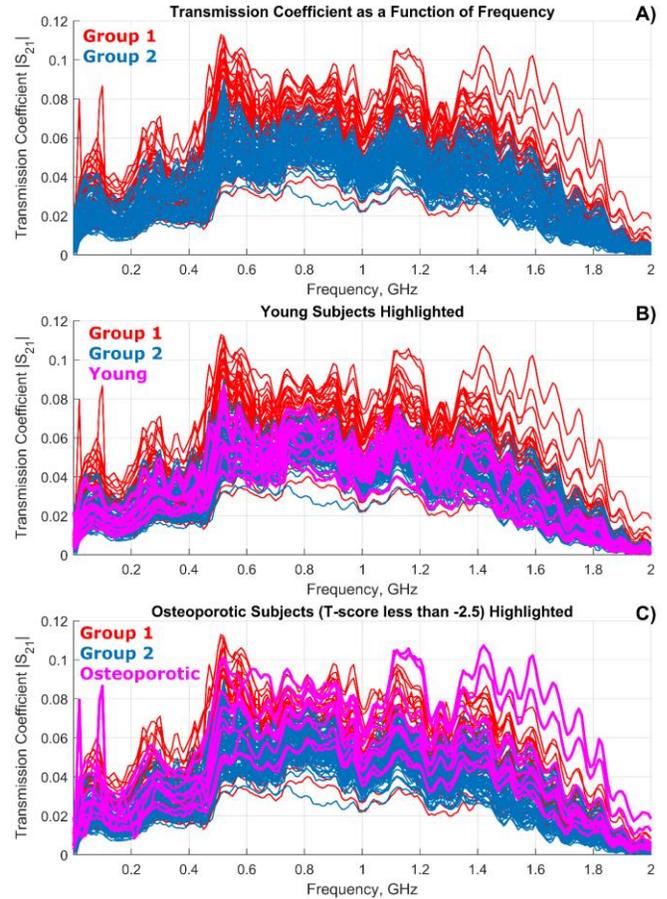


Fig. 2. a) – Transmission coefficient $|S_{21}(f)|$ between the two antennas and through left and right wrists the frequency range 0-2 GHz for all subjects from Group 1 (osteopenic/osteoporotic) and Group 2 (healthy). Red color corresponds to Group 1 while blue color corresponds to Group 2. 160 frequency curves (both arms for all 80 subjects) in total are shown in the figure. b) – The same as in a) but with the data for seven young adults highlighted in magenta. c) – The same as in a) but with the data for five osteoporotic subjects (T score below -2.5) highlighted in magenta.

Additionally, this study also used the normalized transmission coefficient data shown in Fig. 7 of [14]. The normalized data are similar in shape but have less overlap between Groups 1 and 2 due to multiplication by the predictors (clinical risk factors) of osteoporosis as shown in Eq. (1). The normalized data also provide great differentiation between the younger and older healthy subjects [14], where no such differentiation is present in the raw data in Fig. 2B. Use of the raw data, therefore, is a more challenging problem due to the overlap between extremes of Groups 1 and 2 (Fig. 2B, Fig. 2C).

$$N_{21} = \frac{Age}{BMI} |S_{21}| \quad (1)$$

D. Topology of Neural Network

The neural network used to generate the binary classifier was a multilayered network based on MLP (Multi-Layer Perceptron) classifier implemented using the MATLAB Deep Learning Toolbox™ (MathWorks, Inc, Natick, MA, USA). A flow diagram for the entire neural network is shown in Fig. 3 and is explained below.

The input in Fig. 3 (featureinput) was provided as a matrix with each row consisting of the features for a given subject. The input data was then sent to the first dropout layer (dropout_2), which set its features to 0 randomly with a probability of 50%. The dropout layers prevent all weights of the following layer from being updated simultaneously, which helps prevent overfitting. The first fully connected layer (fc_2) reduced the number of features by an order of magnitude. For input data with length 201, it would output 20 features and for input data with length 402 it would output 40 features, as specified by dividing the number of input features by 10 and rounding the result to the nearest integer. Next, a relu activation function was applied to these reduced features. The following layer (dropout_1) had a 50% probability, identical to the first dropout layer. The result of the dropout was fed to the second fully connected layer (fc_1), which combined the result of the dropout into a binary classifier. The softmax function was applied to this result, and then it was classified using cross-entropy loss.

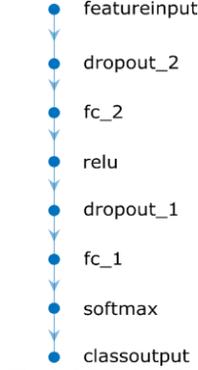


Fig. 3. MLP classification neural network flow diagram featuring two fully connected layers.

E. Input Data

The data used to train the network was derived from the transmission coefficient measured through the subject’s wrist. It came in four configurations per frequency point:

1. The complex number representation of the transmission coefficient with the real part and imaginary part of each frequency point as separate features.
2. The magnitude of the transmission coefficient
3. The phase of the transmission coefficient
4. The magnitude and phase of the transmission coefficient.

All input data series were generated by averaging together the left and right arm spectra for each subject, to match [14].

First, we used a conventional training scheme which split the data between a single training and single validation set with training sets chosen between 20 and 50 subjects (30% to 75% of data) in size. Due to the small size of our dataset, we did not use a test set. This scheme was tested with all four input data configurations. Training was stopped manually when the best observed accuracy was achieved.

Because the first scheme needed more data to produce a consistent result, we additionally attempted a leave-one-out cross-validation scheme [16]. Under a leave-one-out scheme, the data are randomly distributed into subsets with approximately proportional numbers of Group 1 and Group 2

subjects in each set. Each data subset takes a turn being the validation data, while the remaining sets are the training data. This provides a number of individual results equal to the number of subsets. Statistics then indicate the overall performance of the classifier.

Because subset size could influence the results, cross validation was performed both with 7 subsets of nominally 10 subjects and 10 subsets of nominally seven subjects. In both cases, 3 subsets were undersized by one subject. Leave-one-out cross-validation was performed on data configurations 1 and 2 only.

In all cases, two sets of networks were trained: one using the raw transmission coefficient data and the other using the normalized transmission coefficient data featured in [14]. Networks trained using “non-normalized” data used the raw magnitude and/or phase information directly, whereas networks trained using “Normalized” data first applied Eq. (1) to the magnitude and/or phase spectra before putting the data through the neural network.

III. RESULTS

A. Leave-One-Out Cross-Validation

The neural network setup described in the previous section was applied to the transmission coefficient dataset for 67 subjects suitable for the dichotomous diagnostic set as described in Section II.B above.

Table 1 shows the aggregate of the results from the first leave-one-out testing using all 7 subsets for different input data. The same network was also trained with phase data and with magnitude and phase data simultaneously, but the resulting network was severely over-fitted after training for 1,000 epochs. Similar results have been obtained for the second leave-one-out testing, using each of the 10 subsets interchangeably as the validation data.

Table 1. Training results for MLP neural network using leave-one-out with 7 subsets of 10 subjects each. The numbers are the mean across 7 trials, where each trial used a different single subset as the validation data.

Input Data		Epochs	Sensitivity	Specificity	Accuracy
Real/ Imag.	Raw	1,000	0.690	0.905	0.824
	Norm.	1,000	0.761	0.952	0.882
Mag.	Raw	1,000	0.630	0.905	0.803
	Norm.	1,000	0.797	0.905	0.867

B. Fixed Training Sets

The network trained from normalized magnitude data from the 30-subject training set produced a mean sensitivity of 0.85 and mean specificity of 0.91 as seen in Table 2. Normalized data performed better in accuracy than non-normalized data by about 0.05. Table 3 shows increasing sensitivity correlates with training set size.

Table 2. Training results using a fixed 30-subject training set and 37-subject validation set with early-stop training. Mean across 5 trials with the same training and validation data.

Input Data		Epochs	Sensitivity	Specificity	Accuracy
Real/ Imag.	Raw	175	0.571	0.852	0.746
	Norm.	800	0.686	0.948	0.849
Mag.	Raw	175	0.571	1.000	0.838
	Norm.	175	0.857	0.913	0.892

Standard deviations for all entries in Table 2 were 0.0 except for the complex normalized data specificity (0.0238) and accuracy (0.0148).

Table 3. Training results for raw data using early-stop training on magnitude data. Mean across 5 trials with the same training and validation data.

Training	Epochs	Sensitivity	Specificity	Accuracy
20	350	0.533	0.986	0.813
30	175	0.571	1.000	0.838
40	225	0.600	1.000	0.852
50	125	0.667	1.000	0.882

IV. DISCUSSION

In this study, we have found that the neural network trained with the entire frequency spectrum of radio wave propagation through the wrist may serve as a promising predictor tool for detecting osteopenic/osteoporotic conditions. Raw non-normalized data for the transmission coefficient through both wrists have been used as an input, in contrast to our previous study [14], where the processed data included clinical risk factors as well. In [14], a simple threshold binary classifier was used, which is functionally equivalent to checking the area under the entire frequency curve for every subject.

A. Results for non-normalized (raw) data

In the leave-one-out testing, the trained neural network provides sensitivity and specificity values of ~70% and 90%, respectively. The specificity compares favorably to the sensitivity and specificity provided by data that included clinical risk factors (both 87%), presented in a prior study [14]. The increase in the specificity obtained in the present study is a significant advantage due to the increased correctness when predicting the healthy condition, thereby improving utility for prescreening. The inclusion of the phase data by the neural network serves to increase its sensitivity compared to a network trained using only magnitude data (Table 1).

B. Results for normalized data

Note that, when instead of the raw dataset, the neural network is applied to the normalized dataset including other clinical risk factors as in [14], a meaningful improvement is obtained. Normalizing the data provides a 5.5% increase in overall accuracy. This boost appears to be derived mostly from increased sensitivity. It appears, therefore, that inclusion of additional clinical risk factors will be complementary to the ability of the transmission data to reliably differentiate between healthy and diseases patients.

C. Results for smaller training sets

Due to the trend of increasing accuracy with increased training set size (Table 3), it can be assumed that additional data will lead to better classification results.

D. Future improvements

It is likely possible to further increase the overall accuracy by employing larger training and validation sets. Also note that subjects' data additionally includes the neglected circumference of their wrists, which could help further improve the sensitivity of the classification. Expected transmission varies with fat and muscle content in the wrist, and wrist circumference is an indicator of this. Considering the wrists separately could also lead to an improvement due to necessarily doubling the amount of data.

V. CONCLUSION

These results are obtained without inclusion of any additional clinical risk factors. We therefore conclude that the radio transmission data are usable on their own as a predictor of bone density, without the need to include any clinical risk factors into the calculations as done so previously [14].

Our approach correlates well with the available DXA measurements and has the potential for screening patients at risk for fragility fractures, given the ease of implementation and the low costs associated with both the technique and the equipment. Neural networks can identify and use characteristics of the data not readily apparent to the human eye to increase specificity of predictions. To get widespread acceptance and validation, this approach must be evaluated in a clinical trial, where patient data can be trained and tested against actual fracture cases.

REFERENCES

- [1]. Foundation IO. Facts and Statistics. <https://www.iofbonehealth.org/facts-statistics> (2019).
- [2]. Meaney, P. M. et al. Clinical Microwave Tomographic Imaging of the Calcaneus: A First-in-Human Case Study of Two Subjects. *IEEE Trans Biomed Eng.* **59**, 3304-13 (2012).
- [3]. Mundi, S., Pindiprolu, B., Simunovic, N. & Bhandari, M. Similar mortality rates in hip fracture patients over the past 31 years. *Acta Orthopaedica.* **85**, 54-9 (2014).
- [4]. Kannus, P., Natri, A., Paakkala, T. & Jarvinen M. An outcome study of chronic patellofemoral pain syndrome. Seven-year follow-up of patients in a randomized, controlled trial. *J Bone Joint Surg Am.* **81**, 355-63 (1999).
- [5]. Oden, A., McCloskey, E.V., Kanis, J.A., Harvey, N.C. & Johansson, H. Burden of high fracture probability worldwide: secular increases 2010-2040. *Osteoporos Int.* **26**, 2243-8 (2015).
- [6]. Lee, S. Y. & Gallagher, D. Assessment methods in human body composition. *Current Opinion in Clinical Nutrition and Metabolic Care.* **11**, 566-72 (2008).
- [7]. Lochmuller, E. M., Krefting, N., Burklein, D. & Eckstein F. Effect of fixation, soft-tissues, and scan projection on bone mineral measurements with dual energy X-ray absorptiometry (DXA). *Calcif Tissue Int.* **68**, 140-5 (2001).
- [8]. Lochmuller, E. M. et al. In situ femoral dual-energy X-ray absorptiometry related to ash weight, bone size and density, and its relationship with mechanical failure loads of the proximal femur. *Osteoporosis Int.* **11**, 361-7 (2000).
- [9]. Genant, H. K. et al. Noninvasive assessment of bone mineral and structure: state of the art. *J Bone Miner Res.* **11**, 707-30 (1996).
- [10]. Schuit, S. C. et al. Fracture incidence and association with bone mineral density in elderly men and women: the Rotterdam Study. *Bone.* **34**, 195-202 (2004).
- [11]. Cummings, S. R. et al. Improvement in spine bone density and reduction in risk of vertebral fractures during treatment with antiresorptive drugs. *Am J Med.* **112**, 281-9 (2002).
- [12]. Chandra, R., Zhou, H., Balasingham, I., Narayanan, R. M. On the opportunities and challenges in microwave medical sensing and imaging. *IEEE Trans Biomed Eng.* **62**, 1667-82 (2015).
- [13]. Gabriel, C. & Gabriel, S. Compilation of the dielectric properties of body tissues at RF and microwave frequencies. US Air Force <http://www.brooks.af.mil/AFRL/HED/hedr/reports/dielectric/home.html> (1997).
- [14]. Makarov, S.N., Noetscher, G.M., Arum, S., Rabiner, R., and Nazarin, A. Concept of a Radiofrequency Device for Osteopenia/Osteoporosis Screening. *Sci Rep* **10**, 3540 (2020).
- [15]. Kanis, J.A. et al. The use of clinical risk factors enhances the performance of BMD in the prediction of hip and osteoporotic fractures in men and women. *Osteoporos Int.* **18**, 1033-46 (2007).
- [16]. Molinaro, A. M.; Simon, R.; Pfeiffer, R. M. (2005-08-01). "Prediction error estimation: a comparison of resampling methods". *Bioinformatics.* **21** (15): 3301-3307. doi:10.1093/bioinformatics/bti499. ISSN 1367-4803. PMID 15905277