Prediction of Hip Fracture in Post-menopausal Women using Artificial Neural Network Approach

Thao P. Ho-Le[a], *Member, IEEE*, Jacqueline R. Center[b,c], John A. Eisman[b,c,e], Tuan V. Nguyen[a,b,c,d,e,*], Hung T. Nguyen[a,*], *Senior Member, IEEE*

***Abstract*— Hip fracture is one of the most serious health problems among post-menopausal women with osteoporosis. It is very difficult to predict hip fracture, because it is affected by multiple risk factors. Existing statistical models for predicting hip fracture risk yield area under the receiver operating characteristic curve (AUC) ~0.7-0.85. In this study, we trained an artificial neural network (ANN) to predict hip fracture in one cohort, and validated its predictive performance in another cohort. The data for training and validation included age, bone mineral density (BMD), clinical factors, and lifestyle factors which had been obtained from a longitudinal study that involved 1167 women aged 60 years and above. The women had been followed up for up to 10 years, and during the period, the incidence of new hip fractures was ascertained. We applied feed-forward neural networks to learn from the data, and then used the learning for predicting hip fracture. Results of prediction showed that the accuracy of model I (which included only lumbar spine and femoral neck BMD) and model II (which included non-BMD factors) was 82% and 84%, respectively. When both BMD and non-BMD factors were combined (Model III), the accuracy increased to 87%. The AUC for model III was 0.94. These findings indicate that ANNs are able to predict hip fracture more accurately than any existing statistical models, and that ANNs can help stratify individuals for clinical management.**

I. Introduction

In older women, osteoporotic hip fracture is a serious public health problem concern. This is true because hip fracture is associated with an increased risk of mortality. For women aged 50 years, the risk of hip fracture is ~15% during their remaining lifetime [1]. More importantly, hip fracture causes considerable morbidity, leads to excess mortality, and incurs significant financial burden on societies [2]. The identification of high-risk individuals for fracture is critically important, because it could facilitate early intervention to reduce the burden of hip fracture in the general population.

[a] Faculty of Engineering and Information Technology, University of Technology, Sydney, Broadway, NSW 2007, Australia. (Le.P.Ho@student.uts.edu.au; Hung.Nguyen@uts.edu.au; TuanVan.Nguyen@uts.edu.au)

[b] Bone Biology Division, Garvan Institute of Medical Research, NSW, Australia (J.Center@garvan.org.au; J.Eisman@garvan.org.au; T.Nguyen@garvan.org.au)

[c] St Vincent Clinical School, UNSW, Australia

[d] School of Public Health and Community Medicine, UNSW, Australia

[e] School of Medicine, Sydney University of Notre Dame, Australia

* Dual senior authorship

Prediction of hip fracture is very difficult, because it is affected by multiple risk factors. Risk factors for hip fracture include low bone mineral density, history of hip fracture and fall, being female, advancing age, lower body weight, lack of physical activity, low muscle strength, high alcohol consumption, and cigarette smoking [3]. However, the most important and clinically relevant risk factor is low BMD. Based on the risk factors, a number of statistical models including the World Health Organization fracture risk assessment tool (FRAX®) [4] and the Garvan Fracture Risk Calculator [5] have been developed to assess the risk of hip fracture. The fitness of these models is assessed by the area under the receiver operating characteristic curve (AUC) which reflects the concordance between the model-predicted probability of fracture and actual fracture status. Previous comparison [6, 7] showed that the AUC of these models ranged from moderate (0.7) to good (0.85). Current models for predicting hip fracture do not consider potential interactions between risk factors, and this is a significant weakness of these models. Thus, there is room for improving the accuracy of hip fracture prediction.

Artificial neural network (ANN) could be useful in the prediction of fracture due to its effective performance and high computation rates [8]. By imitating human brain functions, ANN can model complex real-world relationships, including interacting variables. ANN consists of three interconnected layers: input layer for input risk factors, hidden layer, and output layer. The layers are tied together by weighted connections, and are trained on the observed data to derive classification rules. However, in the field of osteoporosis research, ANN has not been widely used. Recent studies have applied ANN to detect vertebral fracture among postmenopausal osteoporosis women [9], and predict mortality following a hip fracture [10]. The two studies have shown that ANN prediction is more accurate than traditional statistical methods such as the logistic regression model, because ANN can model the complex interaction effects between risk factors.

Our underlying hypothesis is that by modeling the interactions between risk factors, ANN algorithm can improve the accuracy of the prediction compared with traditional statistical models. In this study, we sought to develop ANN models in a cohort of hip fracture patients, and then validated the models in a separate cohort.

II. Methodology

A. Study design and settings

The present study is a subset of the Dubbo Osteoporosis Epidemiology Study (DOES), which was designed as a

978-1-5090-2809-2/17/$31.00 ©2017 IEEE

population-based longitudinal investigation. Details of protocol and study design have been described elsewhere [11]. In 1989, all men and women aged 60 years and older in Dubbo city (98.6% Caucasian background) were invited to participate. Dubbo was selected because its population closely resembles the Australian population in terms of age and sex distribution. The study's procedure and protocol were conducted under the approval of the St Vincent's Campus Research Ethics Committee. Written informed consent was obtained from all participants.

B. Measurements

This study was limited to 1167 post-menopausal women, who have been followed for up to 10 years. Extensive data were collected at baseline and subsequent visits. The primary outcome of this study is non-trauma hip fracture, which had been continuously ascertained from 1989, via X-ray reports from the two local radiology Centres. Only fractures occurring following low trauma (e.g., fall from standing heights or less) were included in the analysis. Fractures clearly caused by major trauma (e.g., motor vehicle accident), underlying diseases (e.g., cancer or bone-related diseases) were excluded.

BMD measurements. Bone mineral density (BMD; g/cm^2) at the lumbar spine and femoral neck was measured at study entry. The measurement was done with the dual-energy X-ray absorptiometry (DXA) using a DPX densitometer (GE LUNAR., Madison, WI, USA). Based on the femoral neck BMD, the femoral neck BMD T-score was calculated for each individual as number of standard deviation (SD) different from the young normal level (ideal or peak bone mineral density). We used the "young normal" BMD and SD from reference ranges report of a sample of 100 Australian women aged between 20 to 29 years [12].

Clinical risk factors. Data concerning fracture history (after the age of 50), frequency of falls during the previous 12 months, calcium intakes, alcohol consumption, and cigarette smoking were collected by structured questionnaire. Physical activity was assessed by the metabolic equivalent index, which was the average number of hours per day spent in each of five levels of activity and the weighting factor based on associated oxygen consumption for each of levels. Height and weight without shoes of participants were also measured and recorded.

C. Building of ANN

We used a two-layer feed forward neural network to model the observed clinical data for predicting 10-year risk of hip fracture. In order to assess the contribution of different combination of risk factors, we considered three ANN models: Model I included femoral neck BMD and lumbar spine BMD only; Model II included non-BMD clinical factors (e.g., age, weight, height, history of fall, and previous hip fracture); and Model III included both of BMD and non-BMD. Each model has one output node, set as 1 for *Event Reached* (at least one hip fracture occurred within 10 years) and set as 0 for *Event Not Reached* (no incident hip fracture within 10 years). The number of hidden nodes was chosen from 2 to 10 based on trial and error.

In order to assess the generalization, the overall dataset was randomly divided into two separate cohorts for training (60%) and test (40%). The predictive models were built in the training dataset using Levenberg-Marquardt algorithm under 5-fold cross-validation, repeated 5 times to maximize performance and avoid over-fitting. Due to the low incidence of hip fracture, we used Cohen's Kappa coefficient [13] as the metric for optimization. Cut-off values were selected based on Youden J-index. The performance ANN models were assessed in terms of sensitivity, specificity, and AUC [14]. These predictive performance metrics were obtained by averaging 50 running times.

Furthermore, for the purpose of comparison, we also considered the logistic regression and other classification methods, K-Nearest Neighbors (KNN) and Support Vector Machine (SVM), for predicting hip fracture. In order to evaluate the relative importance of each risk factor to fracture prediction, we conducted the "*Weights*" method as described by Gevrey [15]. The method involves computing the product of the weights in input-hidden layer and hidden-output layer. The consultant measure was then standardized to have value ranging between 0 and 1. All analyses were conducted with R version 3.3.2 on the Window platform, mainly with the caret [16] and nnet [17] packages.

III. Results

A. Baseline characteristics

During the follow-up period, 90 women sustained a hip fracture. At baseline (Table I), women who subsequently sustained a hip fracture were, on average, older than those who did not (76.8 vs. 69.1 years). Women subsequently sustaining a hip fracture had lower body mass index (BMI, 23.26 kg/m^2) and physical activity index (PAI, 28.9), compared with those who did not, in which BMI and PAI were 26.08 kg/m^2 and 30.7, respectively. In addition, women having subsequent hip fracture were more likely to have had a prior fracture (15.5%) and experienced a fall during the previous 12 months (57.8%).

TABLE I. Key Baseline clinical characteristics stratified by 10-year hip fracture status

Parameters	No fracture (n=1077)	Hip fracture (n=90)	*p*
Age (yrs)	69.1 (6.4)	76.8 (7.5)	<0.001
Weight (kg)	66.8 (12.6)	56.7 (11.1)	<0.001
Height (cm)	160.1 (6.3)	155.9 (6.6)	<0.001
Body mass index (kg/m^2)	26.08 (4.8)	23.26 (4.1)	<0.001
Femoral neck BMD (g/cm^2)	0.80 (0.12)	0.64 (0.11)	<0.001
FN BMD Tscore	-1.7 (1.02)	-3.02 (0.9)	<0.001
Lumbar spine BMD (g/cm^2)	1.03 (0.19)	0.93 (0.20)	<0.001
Physical activity index	30.7 (2.9)	28.9 (2.7)	<0.001
Previous fracture	71 (6.5%)	14 (15.5%)	0.003
Fall in 12 months	406 (37.3%)	52 (57.8%)	<0.001

Notes: Differences between hip fracture and non-fracture group were tested by the unpaired Student's t-test for continuous variables and chi-square test for categorical variables; Values shown are mean and standard deviation (in brackets). BMD: Bone mineral density.

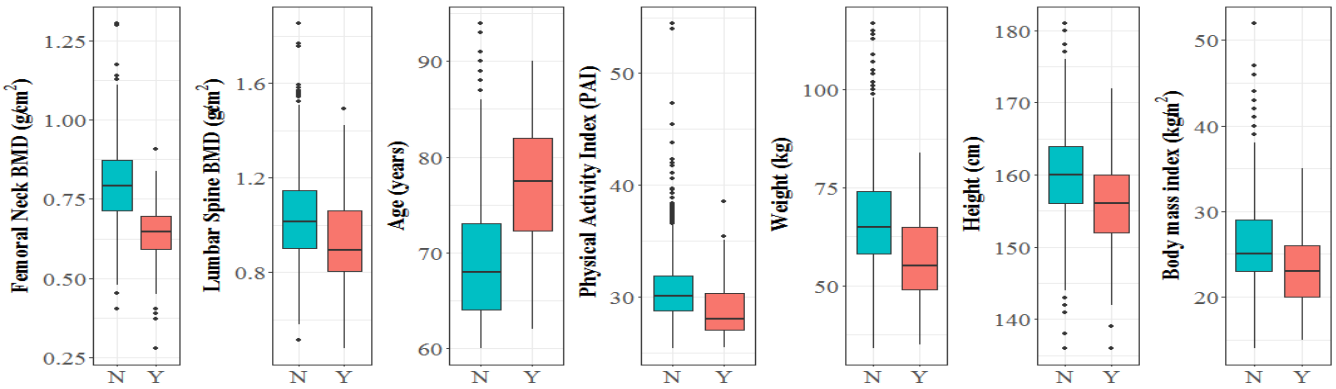

Figure 1: Boxplot of key risk factors stratified by hip fracture status; N: Non-fracture group; Y: Hip fracture group

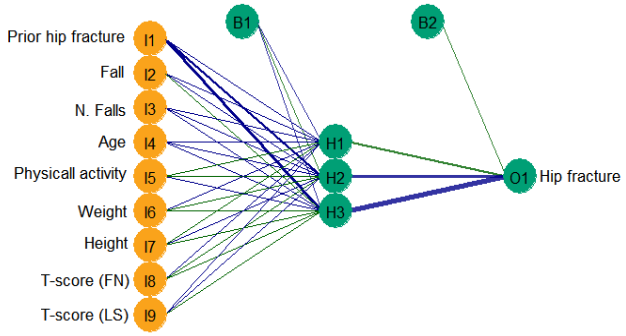

Figure 2: ANN model with BMD and non-BMD risk factors

More importantly, femoral neck BMD (0.64 g/cm^2) and lumbar spine BMD (0.93 g/cm^2) in hip-fracture group were significant lower than those without a subsequent fracture. Using the criteria of femoral neck T-scores ≤ -2.5, the prevalence of osteoporosis among hip fracture patients was 76%, significantly higher than the non-fracture group (22%). The difference in key risk factors between hip fracture and non-fracture groups are shown in Fig.1.

B. Prediction of hip fracture by ANN

Three models of ANNs performed well in the prediction of hip fracture in either training or test dataset (Table II). Model I (which included only two BMD measurements) with 7 hidden nodes yielded an accuracy rate of 79% in the training dataset, and 82% in the test dataset. The AUC for model I was 0.87. Model II (which included only non-BMD risk factors) yielded an accuracy of 86% in in the training dataset and 84% the test dataset. The AUC for model II was 0.92, significantly better than Model I. When BMD and non-BMD risk factors were combined in Model III (Figure 2), the accuracy was 86% in the training dataset and 87% in the test dataset. Compared with Model I and Model II, Model III had the highest AUC values (0.94).

The predictive performance for other models is shown in Table III. The sensitivity for ANN (83%) was greater than that for KNN (81%) and SVM (81%). The specificity and accuracy for KNN (same at 79%) and SVM (same at 82%) were also lower than ANN (88% and 87%, respectively). Analysis of relative importance revealed that the most consistent predictor of hip fracture were history of fracture is (relative importance 42 %), followed by BMD (23%), fall (17%), PAI (11%), age (4%), weight (2%), and height (1%).

IV. DISCUSSION

Predicting hip fracture is a challenging endeavor, because many factors interactively contribute to an individual's hip fracture susceptibility. A number of statistical models including FRAX® and Garvan Fracture Risk Calculator have been developed for assessing the risk of hip fracture, but these models have moderate to good discriminatory power. In this work, we considered hip fracture as a classification problem, and we have demonstrated that by modeling the multidimensional relationships ANNs yielded a better discrimination than previous statistical models. The present study suggests that ANN significantly outperformed other algorithms such as logistic regression, k-Nearest Neighbors, and Support Vector Machine. These findings deserve further elaboration.

In the presence of multiple risk factors, the number of possible interactions becomes very large, and traditional statistical models can not accommodate these interactions due to limited sample size. In ANN, multidimensional interactions are "learned" from the observed data, and this learning process can help better discriminate hip fracture from a non-hip fracture.

TABLE II. NEURAL NETWORK CLASSIFICATION RESULTS

Model	H	Training dataset			Testing dataset		
		Sens (%)	Spec (%)	Acc (%)	Sens (%)	Spec (%)	Acc (%)
FN+LS BMD	7	79.6	79.1	79.2	77.8	81.9	81.6
Non-BMD	7	88.9	85.8	86.0	80.6	84.4	84.1
FN+LS BMD + Non-BMD	3	88.9	86.1	86.3	83.3	87.7	87.3

Notes: BMD, Bone mineral density; FN, Femoral neck; LS, Lumbar spine; H, Hidden node; Sens,Sensitivity; Acc, Accuracy;

TABLE III. PREDICTIVE PERFORMANCE OF FOUR ALGORITHMS FOR PREDICTING HIP FRACTURE

Method	Training dataset			Test dataset		
	Sens (%)	Spec (%)	Acc (%)	Sens (%)	Spec (%)	Acc (%)
ANN	88.9	86.1	86.3	83.3	87.7	87.3
LR	90.7	86.4	86.7	77. 8	81.8	81.5
KNN	100.0	83.3	84.6	80.6	79.3	79.4
SVM	92.4	96.9	96.6	80.6	81.6	81.5

Notes: ANN, artificial neural network; LR, logistic regression; KNN; K-nearest neighbors; SVN, support vector machine. Sens, sensitivity; Spec, specificity; Acc, accuracy.

In this study, based on clinical reality, we considered three competing models: BMD only, clinical risk factors only, and a combination of both BMD and clinical risk factors. We found that the ANN model with clinical risk factors could perform as well as, or even better than, the ANN model with BMD alone. However, the "best" model was the one with both BMD and clinical risk factors. This finding is consistent with previous studies in which BMD alone accounted for less than 50% of all fracture cases. Thus, non-BMD factors are as important as, or even more important than, BMD in the prediction of fracture. In other words, prediction of hip fracture is better by considering BMD along with other clinical risk factors. However, the finding also suggests that in the absence of BMD, clinical risk factors are able to predict hip fracture as well as does BMD.

The risk factors considered in this study are all clinically relevant. Advancing age and low bone mineral density are well-known risk factors for hip fracture. In this study, we found that hip fracture patients had femoral neck BMD lower than non-fracture individuals by more than 1 standard deviation, and this difference was statistically significant. Thus, model with BMD alone can produce good discrimination. A prior fracture is also a well recognized risk factor for hip fracture. Moreover, fall, low body weight, and lack of physical activity are all known to be associated with hip fracture risk. The present study found that the use of non-invasive risk factors could predict hip fracture as well as the use of BMD alone, suggesting that in the absence of BMD, these clinical risk factors could be used to identify women at high risk of hip fracture.

Over-fitting is a concern for any model building. In this study, the number of events (i.e., hip fractures) per risk factors was >10, which minimizes the possibility of over-fitting. Moreover, we address the potential problem of over-fitting by applying applied 5-fold 5-times-repeated cross-validation on training dataset for reducing the instability problem caused by different local minima of training steps in each fold. The risk threshold that yielded maximum sensitivity and specificity was derived using the Youden J-index to best stratify individuals with or without hip fracture. The consistency between training and test results also suggest that there was no over-fitting in the models.

Previous predictive models (e.g. Garvan Fracture Risk Calculator and FRAX) were designed to predict total fractures, not necessarily focused on hip fracture. However, hip fracture is a very distinct disorder compared with non-hip fracture, because hip fracture patients tend to be older than other fracture patients. Moreover, fall is a very important risk factor for hip fracture, but not an important predictor of vertebral fracture. Therefore, one-size-fits-all models don't perform well for hip fracture. Our model was developed exclusively to predict hip fracture, and as a result, may not perform as well for non-hip fracture.

V. CONCLUSION

This paper has introduced a series of artificial neural network models for predicting hip fracture in post-menopausal women. The model with BMD measurements and non-invasive clinical risk factors yielded the highest discrimination and accuracy. Thus, neural networks can predict fracture for individual women more accurately than the existing predictive models such as FRAX® and the Garvan Fracture Risk Calculator. In future, we will validate our model in an external cohort. We are planning to incorporate genetic variants as input variables for predicting hip fracture, as approximately 50% of hip fracture susceptibility is attributable to genetic factors. We will also consider the application of deep learning to enhance the predictive performance of fracture prediction.

VI. REFERENCES

[1] S. R. Cummings, D. M. Black, and S. M. Rubin, "Lifetime risks of hip, Colles', or vertebral fracture and coronary heart disease among white postmenopausal women," *Arch Intern Med,* vol. 149, pp. 2445-8, Nov 1989.

[2] B. C. Taylor, P. J. Schreiner, K. L. Stone, H. A. Fink, S. R. Cummings, M. C. Nevitt*, et al.*, "Long-term prediction of incident hip fracture risk in elderly white women: study of osteoporotic fractures," *J Am Geriatr Soc,* vol. 52, pp. 1479-86, Sep 2004.

[3] R. Marks, "Hip fracture epidemiological trends, outcomes, and risk factors, 1970-2009," *Int J Gen Med,* vol. 3, pp. 1-17, Apr 08 2010.

[4] J. A. Kanis, A. Oden, H. Johansson, F. Borgström, O. Ström, and E. McCloskey, "FRAX® and its applications to clinical practice," *Bone,* vol. 44, pp. 734-743, 5// 2009.

[5] N. D. Nguyen, S. A. Frost, J. R. Center, J. A. Eisman, and T. V. Nguyen, "Development of a nomogram for individualizing hip fracture risk in men and women," *Osteoporos Int,* vol. 18, pp. 1109-17, Aug 2007.

[6] P. N. Sambrook, J. Flahive, F. H. Hooven, S. Boonen, R. Chapurlat, R. Lindsay*, et al.*, "Predicting fractures in an international cohort using risk factor algorithms without BMD," *J Bone Miner Res,* vol. 26, pp. 2770-7, Nov 2011.

[7] S. K. Sandhu, N. D. Nguyen, J. R. Center, N. A. Pocock, J. A. Eisman, and T. V. Nguyen, "Prognosis of fracture: evaluation of predictive accuracy of the FRAX algorithm and Garvan nomogram," *Osteoporos Int,* vol. 21, pp. 863-71, May 2010.

[8] I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *J Microbiol Methods,* vol. 43, pp. 3-31, Dec 01 2000.

[9] C. Eller-Vainicher, I. Chiodini, I. Santi, M. Massarotti, L. Pietrogrande, E. Cairoli*, et al.*, "Recognition of morphometric vertebral fractures by artificial neural networks: analysis from GISMO Lombardia Database," *PLoS One,* vol. 6, p. e27277, 2011.

[10] C.-C. Lin, Y.-K. Ou, S.-H. Chen, Y.-C. Liu, and J. Lin, "Comparison of artificial neural network and logistic regression models for predicting mortality in elderly patients with hip fracture," *Injury,* vol. 41, pp. 869-873, 8// 2010.

[11] T. Nguyen, P. Sambrook, P. Kelly, G. Jones, S. Lord, J. Freund*, et al.*, "Prediction of osteoporotic fractures by postural instability and bone density," *Bmj,* vol. 307, pp. 1111-5, Oct 30 1993.

[12] M. J. Henry, J. A. Pasco, N. A. Pocock, G. C. Nicholson, and M. A. Kotowicz, "Reference ranges for bone densitometers adopted Australia-wide: Geelong osteoporosis study," *Australas Radiol,* vol. 48, pp. 473-5, Dec 2004.

[13] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement,* vol. 20, pp. 37-46, 1960.

[14] N. Nuryani, S. S. H. Ling, and H. T. Nguyen, "Electrocardiographic Signals and Swarm-Based Support Vector Machine for Hypoglycemia Detection," *Annals of Biomedical Engineering,* vol. 40, pp. 934-945, 2012.

[15] M. Gevrey, I. Dimopoulos, and S. Lek, "Review and comparison of methods to study the contribution of variables in artificial neural network models," *Ecological Modelling,* vol. 160, pp. 249-264, 2/15/ 2003.

[16] M. Kuhn, "Building Predictive Models in R Using the caret Package," *2008,* vol. 28, p. 26, 2008-11-10 2008.

[17] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4 ed.: Springer, 2002.