

Prediction of Bone Loss Rate Based on Osteoporosis Risk Features

M. Saranya M.Sc., M.Phil¹, Dr. K.Sarojini M.C.A., M.Phil., Ph.D.²

¹Research Scholar, Computer Science, S.N.R. Sons College, Coimbatore, India

²Head of the Department, Department of Computer Science, L.R.G Govt. Arts College for Women, Tirupur, India

Abstract: Osteoporosis is found from the set of electronic health care data and the bone loss rate is calculated from the available osteoporosis risk feature sets. The main disadvantage arise in existing work is that, bone loss rate cannot be predicted accurately due to limited number of features. This is resolved in our proposed work by introducing the relevance based feature selection which helps to predict the related features with less availability of feature. Osteoporosis measured by bone mineral density, bone fracture risk is determined by the bone loss rate and various factors such as family history and life style. . The Deep belief network is used for fine tuning of risk factors. In this learning process, two stages of process are carried out. They include pre-training and fine tuning. In the pre-training phase, most important risk factors with model parameters are used to calculate contrastive divergence and it minimizes the record size. In the fine tuning phase comparison is made with the results achieved in the previous phase with the ground truth value g1 and again the same comparison done with ground truth value g2, where g1 is refer to as osteoporosis and g2 is refer to as a bone loss rate. The final results are applied to confusion matrix to describe the performance of classification model based on the comparison results, the following are calculated: Accuracy, Precisions, Recall, and F-Measure.

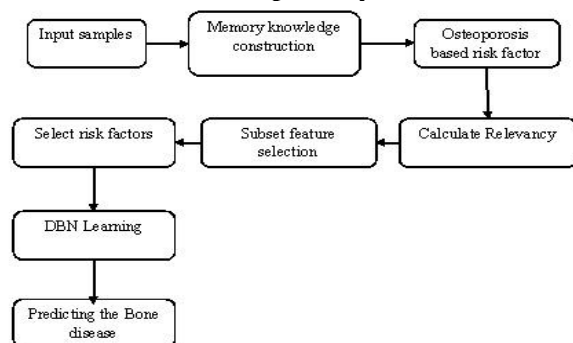
Keywords: Data mining, Healthcare

1. Introduction

Risk Factor (RF) analysis based on patients' Electronic Health Records (EHRs) is a crucial task of epidemiology and public health. Bone disease is one of the major challenging disease or injury that affects human bones. Diseases and injuries of bones are the major causes of abnormalities of the human skeletal system. Out of the major bone diseases over the past few decades, osteoporosis has been recognized as an established and well-defined disease that affects more than 75 million people all over the world [7]. Different from osteoporosis measured by Bone Mineral Density (BMD), bone fracture risk is determined by the bone loss rate and various factors such as demographic attributes, family history and life style. With a faster rate of bone loss, people have a higher risk of fracture [8]. In data mining feature selections and classifications algorithm are used for this analysis.

disease RF selection according to the significance. For clear understanding, we define disease memory (DM) as a model trained by a specific group of samples aiming to memorize the underlying characteristics for this group. Our model is separately trained using diseased samples and non-diseased samples to distinguish their different properties. Bone disease memory (BDM) is a type of DM model which is trained by diseased samples and so it only memorizes the characteristics of those patients who suffer from bone diseases. Similarly, the non-disease memory (NDM) is a model which is trained by the non-diseased samples and memorizes their attributes. The main disadvantages of existing work are The Less availability of features present in the system might leads to the failure of the system in which prediction of bone loss rate would be more difficult. More time complexity in case of presence of missing values. Decrease accuracy of better prediction of risk. The major objective of a new framework is to efficiently find the optimal subset. This can be achieved through a new framework, shown in Figure 1 is composed of two steps: First, relevance analysis that determines the subset of relevant features by removing irrelevant ones, and second, Deep Belief Network (DBN) learning is used. In this learning process, 2 stages of process are carried out. They include pre-training and fine tuning. In the pre-training phase, most important risk factors with model parameters are used to calculate contrastive divergence and it minimizes the record size. In the fine tuning phase, comparison is made with the results achieved in the previous phase with the ground truth value g1 and again we compare that result with ground truth value g2, where g1 is referred to as osteoporosis and g2 is referred to as the bone loss rate. The Final results are applied to the confusion matrix to describe the performance of classification model. Based on the comparison results, the following are calculated; Accuracy, Precision, Recall and F-Measure.

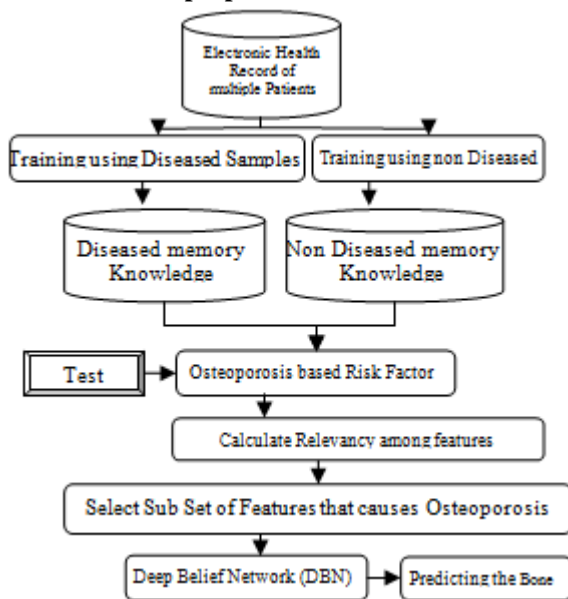
A. Framework of Proposed System



2. Methodology

In existing work, it proposes a novel approach for the study of bone diseases in two aspects: bone disease prediction and

A. Architecture of proposed work



B. Relevancy based feature selection algorithm

Input: $S(F_1, F_2, \dots, F_N, C)$ // a training data set

δ // a predefined threshold

Output: S_{best} // a selected subset

```

1 begin
2 for i = 1 to N do begin
3 calculate  $SU_{i,c}$  for  $F_i$ ;
4 if  $(SU_{i,c} > \delta)$ 
5 append  $F_i$  to  $S'_{list}$ ;
6 end;
7 order  $S'_{list}$  in descending  $SU_{i,c}$  value;
8  $F_j = \text{getFirstElement}(S'_{list})$ ;
9 do begin
10  $F_i = \text{getNextElement}(S'_{list}, F_j)$ ;
11 if  $(F_i \neq \text{NULL})$ 
12 do begin
13 if  $(SU_{i,j} \geq SU_{i,c})$ 
14 remove  $F_i$  from  $S'_{list}$ ;
15  $F_i = \text{getNextElement}(S'_{list}, F_i)$ ;
16 end until  $(F_i = \text{NULL})$ ;
17  $F_j = \text{getNextElement}(S'_{list}, F_j)$ ;
18 end until  $(F_j = \text{NULL})$ ;
19  $S_{best} = S'_{list}$ ;
20 end;
    
```

For a data set S with N features and class C , the algorithm finds a set of predominant features S_{best} . In the first step (lines 2-7), it calculates the SU value for each feature, selects relevant features into S'_{list} based on a predefined threshold δ , and orders them in a descending order according to their SU values. In the second step (lines 8-18), it further processes the ordered list S'_{list} to select predominant features. A feature F_j that has already been determined to be a predominant feature can always be used to filter out other features. The iteration starts from the first element in S'_{list} (line 8) and continues as follows. For all the remaining features (from the one right next to F_j to the last one in S'_{list}), if F_j happens to form an approximate value for F_i (line 13), F_i will be removed from S'_{list} . After one round of filtering features based on F_j , the algorithm will take the remaining feature right next to F_j as the new reference (line 17) to

repeat the filtering process. The algorithm stops until no more predominant features can be selected.

C. DBN Based Classification Algorithm

Input: All risk factors, learning rate ϵ , Gibbs round z ;

Output: Model parameters $M(W; a, b)$;

Pre-training Stage:

- 1: Randomly initialize all $W; a; b$;
- 2: for t from layer V to h_{l-1} do
- 3: clamp t and run CDz to update M_t and $t+1$
- 4: end for

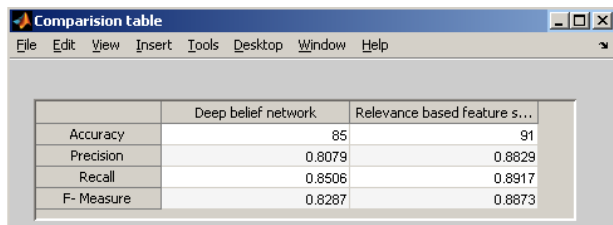
Fine-tuning Stage:

- 5: randomly dropout 30% hidden units for each layer
- 6: loop
- 7: for each predicted result (r) do
- 8: calculate cost (c) between r and ground truth g_1
- 9: calculate partial gradient of c with respect to M
- 10: update M
- 11: calculate cost (c') on holdout set
- 12: if c' is larger than c'_{-1} for 5 rounds then

As shown in DBN Algorithm, the training procedure for DM concentrates on two specific prediction tasks (osteoporosis and bone loss rate) with all Risk Factors (RFs) as the input and model parameters as the output. It includes a pre-training stage and a fine-tuning stage. The first stage is the unsupervised pre-training stage. Here applying the layer-wise Contrastive Divergence (CD) learning procedure for putting the parameter values in the appropriate range for further supervised training. So the result of the pre-training procedure establishes an initialization point of the fine tuning procedure inside a region of parameter space in which the parameters are henceforth restricted. In the second stage, the fine-tuning stage, it takes the advantage of information to train our model in a supervised fashion. In this way, the prediction errors for both prediction tasks will be minimized. Specifically, the parameters from the pre-training stage to calculate the prediction results for each sample and then back propagate the errors between the predicted result and the ground truth (g_1) about osteoporosis from top to bottom to update model parameters to a better state. The another type of information, then repeat the fine-tuning stage by calculating errors between the predicted result and another ground truth (g_2) about bone loss rate. After the two-stage training procedure, our Diseased Memory DM is well trained and can be used to predict osteoporosis and bone loss rate.

3. Implementation

The experimental tests were conducted in the MATLAB simulation environment between the existing and the proposed methodology in terms of performance measures called the accuracy, precision, recall and the F-measure values. It is done to prove the effectiveness of the proposed approach than the existing approach.



	Deep belief network	Relevance based feature s...
Accuracy	85	91
Precision	0.8079	0.8829
Recall	0.8506	0.8917
F- Measure	0.8287	0.8873

Figure 1 Comparison of Existing and proposed Method for Accuracy, Precision, Recall and F-Measure

A. Accuracy Comparison

The accuracy is defined as the correction prediction of the bone diseases with reduced misclassification rate. The accuracy of the proposed method called relevance based feature selection should be more than the existing method called the deep belief network. Accuracy is evaluated as,

$$\text{Accuracy} = \frac{(\text{True positive} + \text{True negative})}{(\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative})}$$

The Accuracy comparison is depicted as graphical notation in the following figure 3

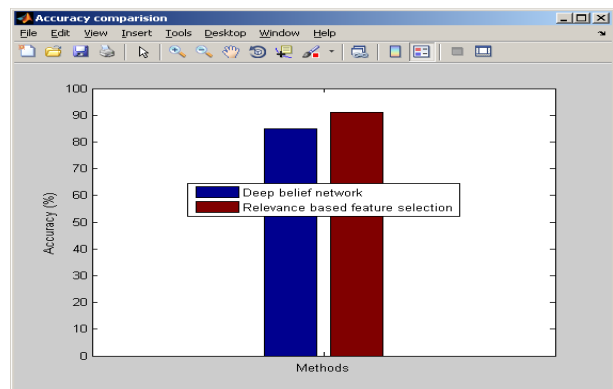


Figure 3 Accuracy comparison of Existing method and proposed Relevance based feature Selection with DBN

In the above graph, proposed and existing research methodologies are compared and evaluated for the performance evaluation. In the x axis methodologies are depicted and in the y axis accuracy in % is depicted. This comparison graph proves that the proposed approach leads to the high accuracy rate of correct prediction of bone disease than the existing approach.

B. Precision Comparison

Precision value is determined based on the retrieval of information at true positive prediction, false positive. In health care data precision is determined the percentage of positive outcome returned that are relevant.

$$\text{Precision} = \text{True Positive} / (\text{True Positive} + \text{False Positive})$$

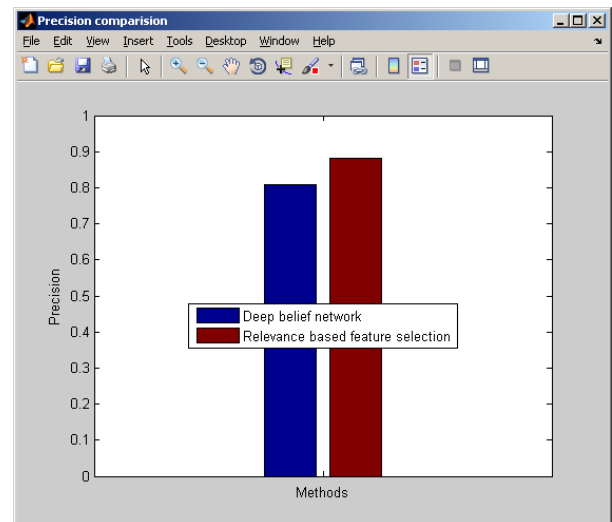


Figure 4: Precision comparison of True Positive Prediction for Proposed and Existing Method

In the above graph, proposed and existing research methodologies are compared and evaluated for the performance evaluation in terms of precision metric value. In the x axis methodologies are depicted and in the y axis precision value is depicted. This comparison graph proves that the proposed approach leads to the high precision rate of correct prediction of bone disease than the existing approach.

C. Recall Comparison

Recall value is determined based on the retrieval of information at true positive prediction, false negative. Recall in this context is also referred to as the True Positive Rate. In that process the fraction of relevant instances that are retrieved.

$$\text{Recall} = \text{True Positive} / (\text{True Positive} + \text{False Negative})$$

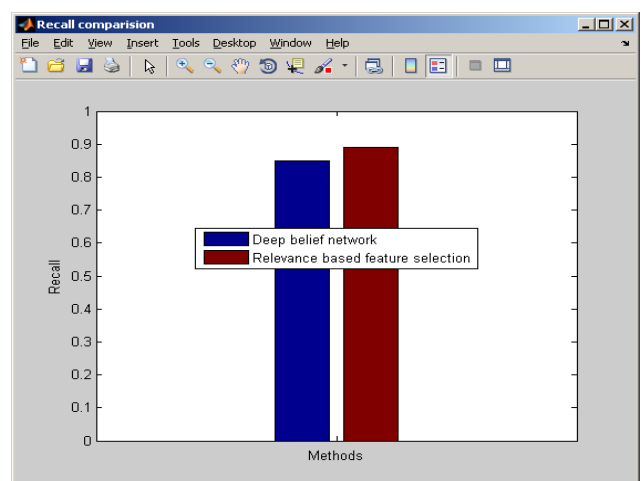


Figure 5 Recall comparison of True positive and False Negative of Proposed and existing methods

In the above graph, proposed and existing research methodologies are compared and evaluated for the performance evaluation in terms of recall metric value. In the x axis methodologies are depicted and in the y axis recall value is depicted. This comparison graph proves that the proposed approach leads to the high recall rate of correct prediction of bone disease than the existing approach.

D. F-Measure Comparison

The F-Measure computes some average of the information retrieval precision and recall metrics

$$\text{F-measure} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

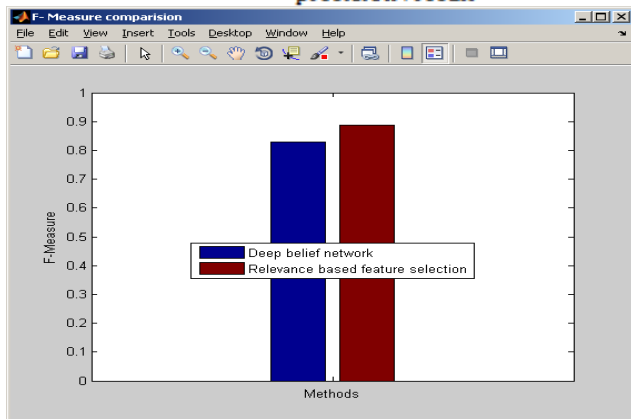


Figure 6 F-Measure Comparison of Precision and Recall of proposed and Existing Method

In the above graph, proposed and existing research methodologies are compared and evaluated for the performance evaluation in terms of F-measure metric value. In the x axis methodologies are depicted and in the y axis F-measure value is depicted. This comparison graph proves that the proposed approach leads to the high F-Measure rate of correct prediction of bone disease than the existing approach.

4. Conclusion

Analysing risk factor is the process of finding the risk level of bone diseases in various stages. Risk factor analysis on bone diseases needs to be done with more concern which should be implemented with various analysis factors. In this work, osteoporosis based bone disease prediction is done by analysing the various risk factors that are present in the electronic health care data with the concern of different stages of bone diseases.

In future following research scenarios can be considered for the efficient prediction of bone related diseases. Different machine learning algorithms can be used to label the features of bone diseases accurately. The various prediction methodologies can be incorporated to predict the bone disease accurately. Additional risk feature sets can be incorporated with the available features to improve the detection accuracy. The relevancy based feature selection with DBN learning approach is used to improve the detection accuracy

References

- [1] Karamjit Kaur and Rinkle Rani, "Managing Data in Healthcare Information Systems: Many Models, One Solution", Published by IEEE computer society 2015, pp. 52-59.
- [2] Hui Li, Xiaoyi Li, Murali Ramanathan, and Aidong Zhang, "Prediction and Informative Risk Factor Selection of Bone Diseases", Published by IEEE/ACM

- Transactions On Computational Biology and Bioinformatics, Vol. 12, No. 1, January/February 2015, PP. 79-91.
- [3] Uma Srinivasan, *Capital Markets Cooperative Research Centre, Australia* "Anomalies Detection in Healthcare Services", Published by the IEEE Computer Society IT Pro November/December 2014, pp. 12-15.
- [4] Daniel R. Harris, Darren W. Henderson, Ramakanth Kavuluru, Arnold J. Stromberg, and Todd R. Johnson "Using Common Table Expressions to Build a Scalable Boolean Query Generator for Clinical Data Warehouses" Published by IEEE Journal of Biomedical and Health Informatics, Vol. 18, no. 5, September 2014. Pp.1607-1613.
- [5] Shancang Li, LiDaXu and XinhengWang "A continuous biomedical signal acquisition system based on compressed sensing in body sensor networks" Published by IEEE Transactions on Industrial Informatics, Vol. 9, No. 3, August 2013. PP 1764- 1771.
- [6] Carlos Ordonez, "Comparing Association Rules and Decision Trees for Disease Prediction", Proceedings of the international workshop on Healthcare information and knowledge management, Pages 17-24, 2006
- [7] World Health Organization, "WHO scientific group on the assessment of osteoporosis at primary health care level," Brussels, Belgium, Summary Meeting Rep., May 2004. Web link: <http://www.who.int/chp/topics/Osteoporosis.pdf>
- [8] B. J. Riis, "The role of bone loss," Amer. J. Med., vol. 98, no. 2, pp. 29S-32S, 1995. Web: <http://www.ncbi.nlm.nih.gov/pubmed/7709930>.
- [9] G.Lemineur, R.Harba, N.Kilic, O.Ucan, O.Osman, and L. Benhamou, "Efficient estimation of osteoporosis using artificial neural networks," in Proc. IEEE 33rd Annu. Conf. Ind. Electron. Soc., 2007, pp. 3039-3044.
- [10] Hui Li, Xiaoyi Li, Lawrence Bone, Cathy Buyea, Murali Ramanathan, Aidong Zhang, "3D Bone Microarchitecture Modeling and Fracture Risk Prediction", Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine, Pages 361-368, 2012
- [11] Dumitru Erhan, Aaron Courville, and Yoshua Bengio, "Understanding Representations Learned in Deep Architectures", <http://memkite.com/deep-learning-bibliography/>
- [12] Yichuan Tang and Chris Eliasmith, "Deep networks for robust visual recognition", Proceedings of the 27th International Conference on Machine Learning (ICML-10), Year: 2010, Pages: 1055-1062
- [13] Yichuan Tang, Ruslan Salakhutdinov, Geoffrey Hinton, "Robust Boltzmann Machines for Recognition and Denoising", Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on 16-21 June 2012
- [14] Nicolas Le Roux and Yoshua Bengio, "Representational Power of Restricted Boltzmann Machines and Deep Belief Networks", Neural Computation, Volume 20 Issue 6, June 2008, Pages 1631-1649
- [15] D. Black, M. Steinbuch, L. Palermo, P. Dargent-Molina, R. Lindsay, M. Hoseyni, and O. Johnell, "An assessment tool for predicting fracture risk in postmenopausal women," Osteoporosis. Int., vol. 12, no. 7, pp. 519-528, 2001.