An Improved and Optimal Prediction of Bone Disease Based On Risk Factors

M.Saranya¹ M.Sc., M.Phil., Dr. K.Sarojini² M.C.A., M.Phil., Ph.D.

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: Bone disease prediction is the most required task in the real world environment which needs to process well to attain the improved prediction of bone diseases. For analysing the bone disease related data sets, various data mining approaches need to process to obtain the improved prediction of the disease. One of the main data mining approaches which is used for predicting the risk factor of bone diseases are machine learning algorithm. Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. The main disadvantage of existing method is that, it will not produce accurate result of bone loss rate due to limited number of features. This drawback is solved in our proposed method by using the relevance based feature selection which tries to predict the result with less availability of feature. Risk factors are fine-tuned using Deep belief network. In this algorithm 2 phases of process are passed out. It includes 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, were 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 to calculate Accuracy. The experimental result of this work proves that the proposed methodology of this work provides performance improvement than the existing methodology in terms of more accuracy.

Keywords: Big data, Data mining, Healthcare.

I INTRODUCTION

In Bone disease any of the diseases or injuries that affect human bones. Diseases and injuries of bones are major causes of abnormalities of the human skeletal system. Although physical injury, causing fracture, dominates over disease, fracture is but one of several common causes of bone disease, and disease is in fact a common cause of fracture. Bone diseases and injuries were formerly regarded as conditions that were more mechanical than metabolic. An improved understanding of the dual mechanical and chemical function of bone, however, has permitted a more integrated biological view. Osteoporosis is a condition that weakens bones, making them fragile and more likely to break. It's a fairly common condition that affects around three million people in the UK. More than 300,000 people receive hospital treatment for fragility fractures (fractures that occur from standing height or less) every year as a result of osteoporosis. Wrist fractures, hip fractures and fractures of the vertebrae (bones in the spine) are the most common type of breaks that affect people with osteoporosis. However, they can also occur in other bones, such as in the arm, ribs or pelvis. There are usually no warnings you've developed osteoporosis and it's often only diagnosed when a bone is fractured after even minor falls.

A. Motivation of the research

Health care department's plays most important role in the real world environment which needs to be concerned more for treating the patients in time. One of the main diseases which play a greatest role in health care department is the bone disease prediction. Bone diseases cannot be identified in the easiest manner because of numerous number of risk factors. There are various researches has been conducted in the real.

B. Research Objective

The main objective of this thesis is to predict the risk natures present in the human body by analysing the electronic health records based on the different nature of features available online. This is done based on the relevancies present between the different features that are available online. This work attempts to predict the bone related diseases that are present in the system in terms of various risk factors such bone loss rate, osteoporosis risk feature sets and so on. Improved detection of the bone related disease prediction by adapting the learning mechanism in terms of relevancy factor using Deep Belief Network (DBN) to predict bone loss rate.

C. Problem definition

The main problem that resides in the real world environment is the bone disease prediction which may cause the serious issues in the human's body. Bone disease prediction would be more difficult process in the real world environment where there are various parameters which cause the diseases. This should be analysed and predicted in the clear way for the clear identification of the bone diseases. The prediction of bone diseases by using the electronic health records available online becomes most complex process which needs to be established well to prevent the serious risk natures. The numerous number of risk factors might poses the complexity of detection of risk factors which is focused in this work.
D. Problem Solution

The issues mentioned above are resolved in this work by introducing the relevance based feature selection approach and the Deep Belief Network based classification approach which is used to predict the bone diseases in terms of various risk factors. The relevance based feature selection is used to select the more relevance features that are present in the bone disease risk factors. The DBN approach is used to accurately classify the risk level of the bone disease risk factors so that prediction accuracy can be improved considerably.

E. Framework of Proposed System

II METHODOLOGY

A. Existing System

Traditionally, Factors selection research has focused on searching for relevant features. Strong relevance of a feature indicates that the feature is always necessary for an optimal subset; it cannot be removed without affecting the original conditional class distribution. Weak relevance suggests that the feature is not always necessary but may become necessary for an optimal subset at certain conditions. Irrelevance indicates that the feature is not necessary at all. An optimal subset should include all strongly relevant features, none of irrelevant features, and a subset of weakly relevant features. However, it is not given in the definitions which of weakly relevant features should be selected and which of them removed.

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.

B. Proposed System

The issues mentioned above are resolved in this work by introducing the relevance based feature selection approach and the Deep Belief Network based classification approach which is used to predict the bone diseases in terms of various risk factors. The relevance based feature selection is used to select the more relevance features that are present in the bone disease risk factors. The DBN approach is used to accurately classify the risk level of the bone disease risk factors so that prediction accuracy can be improved considerably.

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.

C. Architecture of proposed work
D. Relevancy based feature selection algorithm

Input: S(F1,F2, ...,FN,C) // a training data set
δ // a predefined threshold

Output: Sbest // a selected subset

1 begin
2 for i = 1 to N do begin
3 calculate SUi,c for Fi;
4 if (SUi,c > δ)
5 append Fi to S'list ;
6 end;
7 order S'list in descending SUi,c value;
8 Fj = getFirstElement(S'list);
9 do begin
10 Fi = getNextElement(S'list ,Fj);
11 if (Fi <> NULL)
12 do begin
13 if (SUj,c >= SUi,c)
14 remove Fi from S'list ;
15 Fi = getNextElement(S'list ,Fi);
16 end until (Fi == NULL);
17 Fj = getNextElement(S'list ,Fj);
18 end until (Fj == NULL);
19 Sbest = S'list ;
20 end;

E. 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 hl-1 do
3: clamp t and run CDz to update Mt 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 g1
9: calculate partial gradient of c with respect to M
10: updateM
11: calculate cost (c') on holdout set
12: if c' is larger than c' - 3 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 (g1) 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 (g2) 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.

III 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.

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 2

Figure 2 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.

IV CONCLUSION

Healthcare industry is facing challenges now, and recent development in advanced technologies has broad opportunities for confronting such challenges. In this paper risk factor analysis is the process of finding bone diseases in various stages. In the proposed methodology risk factors are analysed in 2 levels. In first level disease prediction is done with relevancies present in different risk factors and the next level is DBN Algorithm is applied on 2 specific prediction tasks. They are osteoporosis and bone loss rate. All risk factors as the input and model parameters as the output the experimental result of this work proves that proposed methodology of this work provides performance improvement then the existing methodology in terms of more accuracy.

V FUTURE WORK

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

REFERENCE

[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.