Breast Cancer Classification: Using Hybrid Technique (Association Rules and Deep Neural Network)

Hajer Kamel Ahmed¹, Dhahir Abdulhade Abdullah¹ and Jamal Mustafa Al-Tuwaijari¹

¹ Department of computer science, College of science, University of Diyala, Baqubah, Iraq
E-mail: hajerkamel33@gmail.com

Abstract: Early detection of any disease is very important since it aids curable with a few of effort. A lot of people fail to detection their disease before it be chronic. This causes an increase in mortality about the world. One of these diseases is breast cancer that can be cured when identified in the early stages before it spreads throughout the body. Develop techniques that can aide physicians to get accurate diagnosis is significantly important in early detection for this disease. The goal is design a hybrid approach (class association rules and deep neural network). In this paper, we design efficient methodology for classifying breast cancer using hybrid approach techniques. Where used a CARs to discover all the interesting relationship in a large database, while the DNN is used for classification purpose. In this study, use Wisconsin Breast cancer dataset from UCI machine learning repository to evaluate the performance of the proposed system. The experiment show that proposed system achieves good results, with high accuracy of 100% and less mean square error rate 0.0002.

Keywords: Classification, Data Mining, Class Association Rules, Deep Neural Network and Breast Cancer Diagnosis.

1. Introduction

Data mining is a necessary step in the process of knowledge discovery in databases (KDD) in which intelligent ways are applied for extract patterns [1]. Data mining is Collection of tools and techniques applied on the nontrivial process of displaying and extracting previously unknown, implicit knowledge, humanly comprehensible, potentially beneficial, from large data sets [2, 3]. The concept intelligent data mining is the application of automatic learning ways to enumerate and discover present patterns in the data [4]. Data Mining is classified into various models based on what the data mining algorithm under into consideration. A rough characterization of the various data mining models can be accomplished by splitting them to predictive and descriptive models. A predictive method tries to set a value to a database fields or undefined value of other variables or future, while description tries to epitomize the information in the database (DB) and extract pattern. One of the descriptive mining models is the mining rule association. A rule, out of the collection of association rules, is one descriptive pattern a compact characterization for a very small subset of the all data. In other words, association rules are used to analyse relationships among data in large databases. Typical predictive models are regression and classification [5]. Classification is a prediction the target class of every sample in data where the classes are which previously known so it is a supervised learning technique.
It is a type of supervised learning algorithm by means of inputting some data to training and another one to test the data that is various from training data. The output from training data is a classifier. Regression maps are examples to a real-valued variable. Now, can be integrated both classification rule mining and association rule mining to form a framework called “Associative Classification” and these rules indicated as Class Association Rules [5]. Classification applications comprises the detecting faults in industry, image and pattern recognition, medical diagnosis and many more. Classification applications comprise the detecting faults in industry, image and pattern recognition, medical diagnosis and many more. To find efficiency of classification model using two parameters are the interpretability and accuracy [4]. Breast cancer is the most widespread cancer in women worldwide. It is also the main reason of death from cancer among women globally. The most efficient way to reduce breast cancer mortality is early detection. Early diagnosis requires a reliable and careful diagnosis procedure that allows doctors to distinguish malignant breast tumours from benign without going for surgical biopsy [7]. The paper is organized as follows: section 2 consist the studies related; Class association rule and deep neural network algorithm is clarified in section 3 and section 4. Followed the proposed system in section 5. Section 6 illustrates the implementation of the proposed system. Conclusion and future work given in section 7.

2. Related Work

- Murat Karabatak and M. Cevdet Ince [8] this research design an expert system for breast cancer detection. Used the Association rules (AR) to decrease the dimensions of the dataset. The researchers develop AR1 and AR2 to decrease the features. Minimize one feature from 9 by AR1 and minimize 5 features out of 9 by AR2. They use the conventional neural network for classification in both AR1 and AR2. This framework achieved 95.2% accuracy on all nine features with threefold cross-validation schema, 95.6% accuracy on AR1 and 97.4% on AR2.

- Jürgen Schmidhuber [9] the researcher provides an overview of deep learning in neural networks. Prove the experimental results that the deep learning algorithms that increase the accuracy and decrease the error rate with respect to training of algorithm.

- S. Karthik, R. Srinivasa Perumal and P. V. S. S. R. Chandra Mouli [10] they proposed a technique used DNN to learn deep features of data. They classify the breast cancer data using DNN with multiple layers of processing. The performance of this system is based on sensitivity, accuracy, precision, specificity and recall. The experimental results show that the accuracy obtained from this system was found 97.66%, which is better than other existing methods.

- Taysir Hassan A. Soliman, Randa Mohamed and Adel A. Sewissy [11] they propose a hybrid technique (DNNs and AHP) multi criteria decision making to deal with large datasets and improve classification accuracy. Measure the performance of the hybrid techniques by use three different breast cancer datasets, involving Wisconsin breast cancer datasets. In a lot of cases, the hybrid techniques of applying DNN Back-propagation with three hidden layers and AHP shown best results are recall rate 86.6%, accuracy 84.33%, F-measure of 90.2% and precision 95.9%.

3. Class Association Rules Mining

This stage progress a method for finding the Class Association Rules from the data. CARs contains the following related theory.

- Association Rules
ARs is one of the important unsupervised learning techniques in data mining. The goal of this technique is finding and describing relationships between various items in a large data set. This relationship is in the form "IF condition Then consequence" [12].

Determine:

\[ D = \{d_1, d_2, \ldots, d_m \} \] is a database that have set of m data and \( d \subseteq J \)
\[ J = \{j_1, j_2, \ldots, j_n \} \] is a set of all items that show in D

The association rule has a form, which is \( X \rightarrow Z \) by support value = S\% and Confidence value = C\% by \( X, Z \subseteq J \) and \( X \cap Z = \emptyset \) following condition:
- Support value is frequency of number of data that consist of \( X \) and consist of \( Z \) or \( p(X \cup Z) \)
- Confidence value is frequency of number of data that consist of \( X \) and consist of \( B \) or \( p(Z | X) \)
- Support value and confidence value must show in percentage.

- **Class Association Rules (CARs)**

CARs are the particular subset of the association rules whose right-hand-side are restricted to the classification class attribute [5].

According to this, a CARs is of the form

\[ X \rightarrow a_j \]

Where \( a_j \) is class attribute and \( X \) is \{ \( a_1, \ldots, a_{j-1}, a_{j+1}, \ldots, a_m \) \}.

Consider:

\[ D = \{d_1, d_2, \ldots, d_j \} \] is a database that consist collection of data that have m attribute and class by \( d = \{a_1, a_2, \ldots, a_m, z_n \} \)

where \( h=1,2,\ldots n \).

\[ I = \{X_1, X_2, \ldots, X_j \} \] is collection of all items that show in \( D \)

\[ Z = \{z_1, z_2, \ldots, z_n \} \] is collection of class label n class

A class association rule is a imply from the form \( X \rightarrow z \), where \( X \subseteq I \) and \( z \subseteq Z \). A rule \( X \rightarrow z \) holds in \( D \) with Support=S\% and Confidence=C\% following these conditions:
- Support value is the frequency of number of all data that consist of item set \( X \) and class label \( z \)
- Confidence value is the frequency of number of data that consist of class label \( z \) when data have item set \( X \).

Some formats should be considered before formulating model for support and confidence, which is as following:

For CAR,\( \text{conset} \rightarrow z \), where conset is an itemset and \( z \subseteq Z \).
- **ruleitem has the format**
  - \( \langle(\text{conset}, \text{condsupcount}), (z, \text{rulesupcount})\rangle \).  
  - \( h\text{-ruleitem is ruleitem that conset consist of conset} \)
  - \( \text{condsupcount is the number of data that consist of conset} \)
  - \( \text{rulesupcount is the number of data that consist of conset and has the label z} \).

**Then**

\[
\text{Support} = \frac{r}{|D|} \times 100\% \quad (1)
\]
\[
\text{Confidence} = \frac{r}{c} \times 100\% \quad (2)
\]

The proposed algorithm, illustrated below will use this notations and formulae to find class association rules.

**TABLE I:** Algorithm1 used to find the class association
rules
Algorithm1: CARs Generate
$T_1 = \{\text{large 1-ruleitems}\}$;
CAR$_1$.genRules($T_1$);
For (h = 2; $T_{h-1} = \emptyset$; h++)
C$_h$ = candidateGen($T_{h-1}$)
For each data case d $\in$ D
C$_d$ = ruleSubset(C$_h$, d);
For each candidate c $\in$ C$_d$
c.condsupcount++;
if d.class = c.class then
c.rulesupCount++;
End
End
End
d
Th = \{ c $\in$ Ch $|$ c.rulesupCount $\geq$ minsup\};
CAR$_h$.genRules (Th);
end
CARs = $\bigcup_h$CAR$_h$;

4. Deep Neural Networks
Deep Neural Networks are an offshoot of artificial intelligence that been lately applied in differen
domains, such as speech recognize, object detection, image recognition and natural language
processing [13]. DNNs represent a Conventional Multilayer Perceptron (MLP) with increase (often
more than two) hidden layers. With six layers that include of input layer, four hidden layers and an
output layer.

![DNN Diagram](image)

Figure1. DNN with input layer, four hidden layer and output layer

DNNs can be trained based on standard back propagation algorithm [14], in the application of the
back-propagation algorithm; two different passes of calculation are distinct. The first pass is referred
to as the forward pass, and the second is referred to as the backward pass [15].
In the forward pass, the synaptic weights rest unaltered throughout the network, and calculate the function signals of the network on a neuron-by-neuron basis. Appearing the function signal at the output of neuron $k$ is calculate as:

$$Y_k(m) = f_k(v_k(m))$$  \hspace{1cm} (3)$$

Where $v_k(n)$ is the induced local field of neuron $k$, defined by

$$v_k(m) = w_{ok} + \sum_{i=1}^{n} w_{ki}(m) y_i(m)$$  \hspace{1cm} (4)$$

where $n$ is the total number of inputs (excluding the bias) applied to neuron $k$; $w_{ok}$ is the synaptic weight connecting neuron $i$ to neuron $k$; $w_{oi}$ is bias value; and $y_i(m)$ is an input signal of neuron $k$, or equivalently, the function signal appearing at the output of neuron $i$, if neuron $k$ is in the first hidden layer of the network, then $n=n0$ and the index $i$ refers to the $i$th input terminal of the network, for which we write

$$y_i(m) = x_i(m)$$  \hspace{1cm} (5)$$

Where $x_i(m)$ is the $i$th element of the input vector (pattern). If, on the other hand, neuron $k$ is in the output layer of the network, then $n=nL$ and the index $k$ refers to the $k$th output terminal of the network, for which write

$$y_k(m) = o_k(m)$$  \hspace{1cm} (6)$$

Where $o_k(n)$ is the $k$th element of the output vector of the multilayer perceptron. Compares this output with the desired response $d_k(n)$, the error signal $e_k(n)$ obtain for the $k$th output neuron as show in equation (7). Thus, the forward phase of calculation begins at the first hidden layer by submitting it with the input vector and finishes at the output layer by calculating the error signal for each neuron of this layer.

$$\varepsilon_k(n) = d_k(n) - y_k(n)$$  \hspace{1cm} (7)$$

Backward pass, begins at the output layer by passage the error signals leftward cross the network, layer by layer, and recursively calculating the (i.e., the local gradient) for every neuron. This recursive process allow the synaptic weights and (biases) of the network to undergo changes in accordance with the delta rule, defined by

$$\Delta w_{ik} = \alpha \delta_i(n) y_k(n)$$  \hspace{1cm} (8)$$

$$\Delta b_{ok} = \alpha \delta_k(n)$$  \hspace{1cm} (9)$$

The $\delta$ is equal to the error signal of that neuron multiplied by the first derivative of its nonlinearity as illustrate in equation (10). Hence, use equation (8) to calculation the changes to the weights of whole the connections feeding into the output layer. Given the $\delta$s for the neurons of the output layer, calculate the $\delta$s for all the neurons in the penultimate layer and thus the changes to the weights of all connections feeding into it by use equation (11). The recursive calculation is continued, layer by layer, by propagating the changes to all synaptic weights in the network [15].

$$\delta_k(n) = \varepsilon_k(n) f'(v_k(n))$$  \hspace{1cm} (10)$$

$$\delta_i(n) = f'(v_i(n)) \sum_j \delta_j(n) w_{ij}(n)$$  \hspace{1cm} (11)$$
Update the weights and bias of the network in layer $l$ according to the generalized delta rule

$$W_{ki}^{l}(n+1) = w_{ki}^{l}(n) + \rho[\Delta w_{ki}^{l}(n-1)] + \alpha_{k}^{l}(n)x_{i}^{l-1}(n)$$ (12)

and

$$b_{ik}^{l}(n+1)=b_{ik}^{l}(n)+ \rho[\Delta b_{ik}^{l}(n-1)] + \alpha_{i}^{l}(n)x_{i}^{l-1}(n)$$ (13)

5. Proposed System

This proposed work consists of two main parts: CARs and DNN-BP. Firstly, CARs has been used to create rules which have the class attribute as part of its outcome. The Candidate rules are the ones with high confidence values, because these rules are frequent patterns in data set and supposed to improve the accuracy and efficiency of DNNs. Secondly, DNN-BP is used to train the system. The main steps of the proposed system are illustrated below in figure2.

![Figure2. A flow chart of CARs-DNN System](image)
5.1. Pre-processing phase

Pre-processing techniques are applied for handle missing values. Breast cancer dataset contain of 9 attribute with 699 instances and every attribute have value rang 0-10. This dataset uses only two class labels named benign and malignant. Benign is to recognize patients without cancer, and to recognize patients with cancerous tumours is malignant. In dataset 16 instances have missing values, the system will replace all these 16 instances by value 0 To handle those values, before generate class association rules to improve stabilization the system.

5.2. Class Association Rules

At this stage, Association rule were applied for the purpose determine frequent patterns, association and correlations among sets of items in transactional databases. Finding the rules can be effective and interesting in the relational database and useful in many contexts. Generally, CARs leads to understand of the relationship of the class to predictive attributes. Firstly, Apriori algorithm has been applied to find the frequent patterns which can be used to generate the association rules. The relationship between the confidence value and the generated rules is illustrated in figure3. In this work, the value 2 is set to the minimum support parameter and the value 90 to minimum confidence. The output of this stage is 872 rules which represent the inputs to the DNN after converting them to binary values (0, 1) the value 1 refers to the attribute that available in the rule and the value 0 refers to unavailable attribute, Figure4 illustrating some these rules.

![Figure3](relation between confidence and rules)

![Figure4](Generated rules)
5.3. Classification Phase

This phase is the last phase of the proposed system (CARs-DNN) which can identify the patient’s state. The output of the CARs which described in section (3.2) will be the inputs to the DNN, DNN has been used for training system. The training process can be split into two stages: the first stage, represent forward propagation, the second stage, represent back propagation. In the first stage the neurons of input layer are take the input data and find the function signal for these data (by implementing equation (4) and tanh active function) which will be the input to the hidden layer. The output of the hidden layers will be the input in the output layer in which function signal can be calculated (by implementing equation (4) and softmax active function) and this forward propagation learning procedure has been completed. Function signal resulted by output layer would be the output of the forward propagation learning procedure. When the forward propagation output does not satisfy the target output and consist of: calculate the errors and passing these errors to the output layer to debug weights of each level according to error gradient descent. Finally, these errors propagation back to hidden layers and input layer. the forward and error back propagation consists of adjusting the weights for each level continuously and this the neural network acquires training and learning. this process will continuous until acceptable error has been obtain. This algorithm consists of the following steps:

1- Create neural network with an input layer has m input nodes.

2- Select the number of hidden layers required to train the data.

3- Select the bias and learning rate value for each node. Define the initialize weights (small random values not zero).

4- Select activation function

\[ \tanh (v) = \frac{e^v - e^{-v}}{e^v + e^{-v}} \]

5- Select the number of epochs to back propagation.

6- Training the network for specific set of data.

7- After training the network the data that specific for the testing will pass the training model to calculate the classification rate.

8- Train the network until achieved the expected output (or) complete the number of epochs.

9- Compute the accuracy of the model using equation (14).

\[ \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \]  \hspace{1cm} (14)

6. Experiment Results

In this work the breast cancer dataset which used binary classification is used to evaluate the performance of the proposed system [16]. The parameter values used the momentum is 0.01, learning rate 0.001 for network training. The results show that the increasing in the epoch value will decrease the mean Square error value and tend to be stable, as illustrated in Figure 4.
The Testing Cases of Implemented DNN-BP algorithm is illustrated in table 1 and figure5.

Table 1. Performance evaluation using breast cancer dataset

| Number of hidden layer | Number of hidden node in each layer | Accuracy | MSE   |
|------------------------|-------------------------------------|----------|-------|
| 4                      | (4,2,2,3)                           | 91.63    | 0.0767|
| 5                      | (10,10,8,5,4)                       | 98.33    | 0.0164|
| 7                      | (10,10,10,7,5,4,3)                  | 100      | 0.0002|

Figure 6. relation between hidden layer and accuracy

7. Conclusions and future works

In this work, hybrid approach (CARs and DNNs) has been developed to build efficient and accurate classifier in order to deal with large datasets and get better accuracy. Wisconsin breast cancer dataset is used to evaluate the performance of the proposed work. The implementation results show that the system of applying CARs and DNN backpropagation with seven hidden layers gave best accuracy rate 100% and less MSE rate 0.0002. Future work includes hybridization of the DNN algorithm with one optimization techniques.
8. References

[1] Rani K.U 2010. Parallel Approach for Diagnosis of Breast Cancer using Neural Network Technique. International Journal of Computer Applications(0975 – 8887). Vol 10, p 3.

[2] Perichinsky G and Garc’ia-Mart´inez R 2000. Proc. Workshop Comput. Sc. Researchers (La Plata University Press, Buenos Aires, p 107.

[3] Perichinsky G, Servetto A, Garc´ia-Mart´inez R, Orellana R and Plastino A 2003, Taxonomic Evidence Applying Algorithms of Intelligent Data Mining Asteroid Families,comput. Sci., Software Eng., Information Technology, e-Bussines & Applications (Rio de Janeiro) p 308.

[4] Kharya S, Agrawal S and Soni S 2014. Naive Bayes Classifiers: A Probabilistic Detection Model for Breast Cancer, International Journal of Computer Applications (0975 – 8887) Vol 92 , p 10.

[5] Shekhawat P.B and Dhande S.S 2011. A Classification Technique using Associative Classification, International Journal of Computer Applications (0975 – 8887) Vol 20, p 5.

[6] Sharma R, Narayan S and Khatri S 2016. Medical Data Mining Using Different Classification and Clustering Techniques: A Critical Survey, Second International Conference on Computational Intelligence & Communication Technology.

[7] Kharya S 2012. Using data mining techniques for diagnosis and prognosis of cancer disease, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol 2, p 2.

[8] Karabatak M and Cevdet Ince M 2009. An expert system for detection of breast cancer based on association rules and neural network, Expert Systems with Applications, (36) 3465–3469.

[9] Schmidhuber J 2015. Deep learning in neural networks: An overview, Neural Networks, vol 61, pp 85–117.

[10] Karthik S, Perumal R. S and Mouli P. V. S. S. R. C 2018. Breast Cancer Classification Using Deep Neural Networks, Springer Nature Singapore Pte Ltd.

[11] Soliman T.H. A, Mohamed R and Sewissy A. A 2016, A Hybrid Analytical Hierarchical Process and Deep Neural Networks Approach for Classifying Breast Cancer, IEEE.

[12] Sermswatsri P and Srisa-an C 2006. A neural-networks associative classification method for association rule mining, Transactions on Information and Communication Technologies (WIT), vol 37.

[13] Hinton G, Deng L, Yu D, Dahl G, Mohamed A, Jaitly N, Senior A, Vanhoucke V, Nguyen P., Sainath T and Kingsbury B 2012. Deep Neural Networks for Acoustic Modeling in Speech Recognition, IEEE Signal Processing Magazine, vol 29, pp 82-97.

[14] Yu D and Deng L 2015. Automatic Speech Recognition, Chapter4. Signals and Communication Technology, Springer-Verlag, London.

[15] Haykin S 2009. Neural Networks and Learning Machines -3rd ed, McMaster University Hamilton, Ontario, Canada.

[16] https://archive.ics.uci.edu/ml/datasets.html