3-Satisfiability Reverse Analysis Method for Breast Cancer Detection

Samaila Abdullahi[1], Mohd Asyraf Mansor[2]

[1]Department of Mathematics, Sokoto State University Sokoto, PMB 2134 Sokoto State, Nigeria

[2]School of Distance Education, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia

Abstract: **Accurate breast cancer screening is essential to ensure patient with such symptom can be treated accordingly. Medical screening is quite complicated since every patient sign and symptoms will be screened and when the number of features increases the medical practitioner will not able to be screened appropriately. 3-Satisfiability Reverse Analysis Method (3-SATRA) incorporated with Hopfield neural network is a new approach for the early detection in breast cancer medical dataset. 3-SATRA has proposed to extract the best logic rule that will representing the attribute of breast cancer dataset since the conventional data extraction techniques focus only on standalone neural network. The proposed method is applied to Breast Cancer dataset obtained from UCI machine learning repository. To pursue that, the results of the analysis will promote the early detection stage used for medical practitioners. The simulation will be executed using Dev C++ 5.11 as a tool for training, testing and validating the performances of the proposed method. The performance of the method was measured based on Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Sum of Squared Error (SSE), and Computational Time. The performance and accuracy of the results obtained have shown the effectiveness of 3SATRA in medical data mining.**

Keyword: **3-Satisfiability problem; Hopfield neural network; Breast cancer; logic programing; Data mining.**

I INTRODUCTION

Breast cancer is the most common invading cancer in women and the second leading cause of death among them after lung cancer by Salama *et al*. (2012). The medical specialist is unable to screen the disease properly when the number of instances that attributed to the disease increased. The best and efficient way to deals with uncertainty is to apply data mining techniques incorporated with Hopfield neural network and logic mining which are complemented to each other and has the capacity to identify the relationship and compete with the system. To achieved that, data mining model will be formulated to serve as a logic mining tool. The 3-statisfiability problem will be considered as a mapping problem which assign a logic programming in conjunctive normal form (CNF) to truth values. 3-SAT can be defined as a formula in conjunctive normal form with a collection of clauses where each comprises or strictly 3-literals per clause by Kutzkov (2007). Logic mining was justifying consistently in mapping as well as representing the attribute of data set into the symbolic logic form by Mansor *et al.* (2018). Thus, the major issue with logic mining is the generating logical rule which will be embedded in the Hopfield network or any machined learning models. Therefore, the 3-satisfiability reversed analysis as a logical rule has been used to generate the best logic rule and pattern that represent the attribute of the real data set according to Mansor and Sathasivam (2016). Furthermore, the robust nature of data mining technique can be hybridized with 3-SAT logic mining. However, the new reform reverses analysis method for extracting valuable information among the attribute of a real data set by using the conjunctive normal form (CNF) logical rule by Sathasivam and Wan Abdullah (2011).The conceptual modern day Hopfield neural network inspired by the human biological nervous system to mock the computations employed by the human brain by Rojas (2013). The Hopfield neural network (HNN) consists of bipolar threshold unit which is full superimposed in dual direction and the connections are symmetric which cause the network to settle to minimum energy. furthermore, The Hopfield neural network composed of exceptional features such as better stability, recurrent for faster execution and efficient content addressable memory. The discrete Hopfield neural network computation is performed by collection of activated neurons by Mansor and Sathasivam (2016). The data mining is used to extract valuable information of the real dataset. Based on this paper, the data mining techniques is designed by integrating 3-SAT reverse analysis method (3-SATRA) with Hopfield network to execute the breast cancer detection and screening. The method will be simulated using Dev C++ software 5.11. the main idea is to check the performance and efficiency of the proposed method in training, testing and adopting the breast cancer dataset as the level of complexities increased.

II. 3-SATISFIABILITY PROBLEM

Theoretically, 3-satisfaibility can be classify as NP- hard problem by Johnson (1989). Generally SAT is a Boolean logic that composed of three literals which allow choices of values for each literals by Mansor *et al.* (2018). Hence, three satisfiability (3-SAT) will define as a mapping problem to assign a Boolean logic in conjunctive normal form (CNF) to truth values. The 3-SAT can be summarized as follows:

- The three satisfiability in Boolean form consist a set of n variable $p_1, p_2, \dots, p_n$ for each clause, $n = 3$.
- The set of literals can be either positive or negative and the set of d clause in Boolean logic formula H. $\exists d : H = c_1 \wedge c_2 \wedge, \dots, c_d$.

- The clause is combined with $OR(\vee)$ operator and connected with

$AND(\wedge)$ operator.

The prescribed formula for 3-satisfiability is given by:

$$P_{3-SAT} = \overset{n}{\underset{i=1}{\wedge}} c_i \quad (1)$$

where c_i is given by:

$$c_i = \overset{n}{\underset{i=1}{\vee}} c_i(x_i, y_i, z_i) \quad (2)$$

The 3-satisfability formula can be denoted as:

$$H = (K \vee L \vee M) \wedge (N \vee O \vee P) \wedge (Q \vee R \vee S) \quad (3)$$

The equation (3) described the 3-Satisfiability problem in conjunction normal forma (CNF) and it also be formulated in various combination as the number of literals assigned at random. The main objective of the 3- SAT formula is to simplify the output of the given problem.

III. DISCRETE HOPFIELD NEURAL NETWORK

The discrete Hopfield neural network is one of the famous neural networks with a wide range of applications by Hopfield (1982). According to Little (1974) the dynamics of this model is asynchronous that is each neuron updating their state at once. Basically, the Hopfield neural network composed of the highly superimposed element (neurons) that form a network. The HNN consist of good remarkable properties which include parallel execution, faster computation, better stability, and higher memory capacity by Sathasivam (2011).The neuron state in HNN is crafted by $S_i(t)$ where $i = 1, 2, ..., N$ and it is termed bipolar in nature, $S_i \in (-1, 1)$. The bipolar representation of neuron during firing in Hopfield neural network is denoted by:

$$S_i = \begin{cases} 1 \ , \ if = \sum_j W_{ij} S_j > \theta \\ -1 \ , \ otherwise \end{cases} \quad (4)$$

where W_{ij} is the synaptic weight from unit *i* to *j*. S_j is the state of neuron *j* and θ is predefined threshold value. Both S_i and S_j have bipolar representation {1, −1}. The network is symmetrical since the synaptic weight $W_{ij} = W_{ji} = 0$ and the synaptic weight will represent the connection between the variable and the clauses in $3SAT$ formula. The connection model can be generalized to comply with higher order connection and that modifies the field to:

$$h_i(t) = \sum_j \sum_k W_{ijk} S_j S_k + \sum_j W_{ij} S_j + W_i \quad (5)$$

Therefore, since the synaptic weight in HNN is maintained symmetrically the updating rule maintains as:

$$S_i(t+1) = \mathrm{sgn}\left[h_i(t)\right] \quad (6)$$

Whereby sgn is refers to sign function of state of neuron. As indicated in equation (6) the updating rule decreases monotonically with the dynamics. The state of the system evolves from any initial state to a final state where it refers to local minimum of the Lyapunov function by Sathasivam (2006).The final state of the neurons will be examined by using Lyapunov energy function for the discrete HNN in 3-satisfiability clauses is donated by:

$$E_{Lyapunov} = -\frac{1}{3}\sum_i \sum_j \sum_k W_{ijk} S_i S_j S_k - \frac{1}{2}\sum_i \sum_j W_{ij} S_i S_j - \sum_i W_i S_i \quad (7)$$

Where W_{ij} is the synaptic weight, S_i and S_j is refers to state of neurons. The energy value obtained from the equation (7) will be verified as global or local minimum energy and is designed only for 3-SAT in HNN.

IV. IMPLIMENTATION

The 3-Satisfiability reversed analysis (3-SATRA) is incorporated with HNN for doing data mining. Thus, the implementation of the proposed method will be validated during the training and testing of medical data set. The simulations of the designed method will be test on Breast Cancer Coimbra dataset obtained from UCI machine learning repository. The procedure of are summaries as follows:

Step 1: Select the data set to be trained and tested by the proposed method. In this paper, we consider 60% for the training and 40% portion for testing.

*Step 2:*Convert all attribute of the date set into bipolar and values to the neuron state that will generate the best logic, P_{best}.

*Step 3:*Derive a cost function of 3-SAT logic, $E_{3-SAT.}$ by considering $P = \frac{1}{2}(1 + S_P)$ and $\neg P = \frac{1}{2}(1 - S_P)$. Hence the neuron state can be verified as $S_P = 1$ or $S_P = -1$.

Step 4:

Check the clauses satisfaction of $E_{3-SAT.}$ the model will be trained by exhaustive search (ES) method and best pattern that represent the data set will be store in CAM of HNN.

Step 5: Compute the synaptic weight of P_{best} using (wan Abdullah 1993) standard weight management techniques in HNN by comparing the energy function and cost function.

Step 6: Compute the local field, h_i and apply hyperbolic tangent activation function to determine the final state neurons.

Step 7: Generate the induced logic, $P_{induced}$ to compare the output of the network and the actual target output from the data set which will verified as Success or Failure.

Step 8: Compute the corresponding performance evaluation metrics namely, RMSE, MAE, SSE, Accuracy and Computational time of model.

V. RESULT AND DISCUSSION

The platform used in simulating the medical data set is Dev C++ 5.11 for training and testing the proposed method in order to validate the model. In this research, 60% of the data is allocated for the training data set and the remaining 40% as the testing data. The main task is to extract logical rule for Breast Cancer Coimbra data set. Basically, the proposed network will assist the medical practitioners to identify the presence or absence of the disease during screening. The simulations were carried out until $NC = 8$ for simplicity to check the complexity. the HNN-3SAT model with 3-SATRA has successfully extracted the best logical rule for data set.

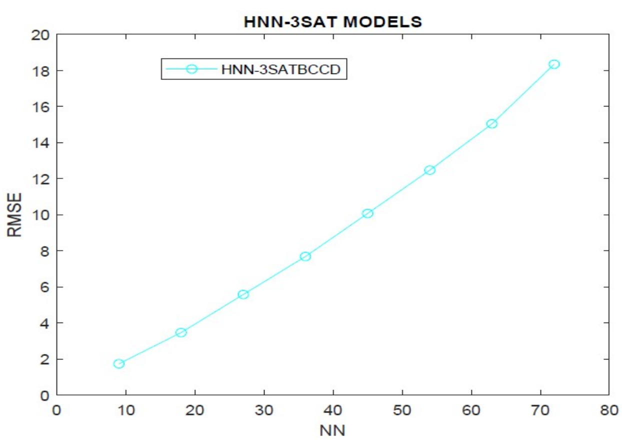

Figure 1. RMSE For HNN-3SATBCD

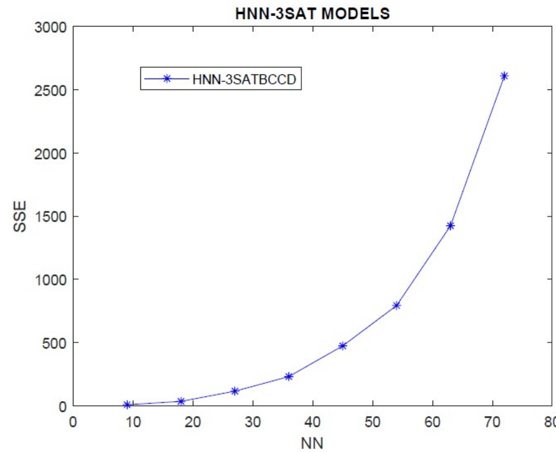

Figure 3: SSE For HNN-3SATNBCD

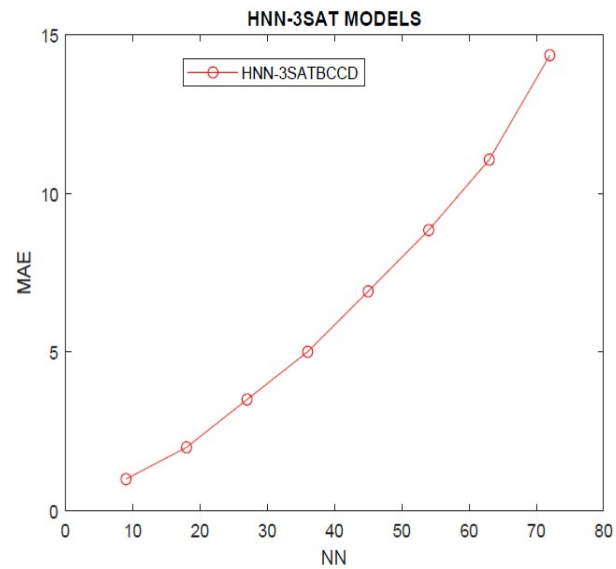

Figure 2: MAE for HNN-3-SATBCD

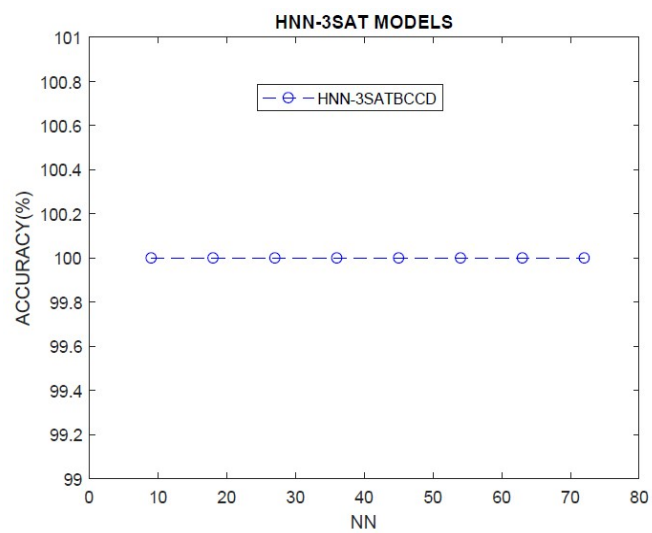

Figure 4: Accuracy for HNN-3-SATBCD

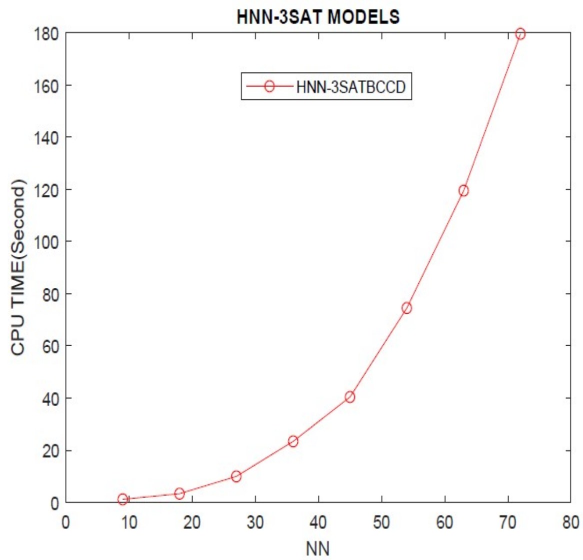

Figure 5: CPU Time for HNN-3-SATBCD

The above graphs present the result of 3 performance evaluation metric with computational time that will measure the efficiency, accuracy and stability of HNN-3SAT in doing data mining. Fig 1: The performance Root Mean Squared Error (RMSE) values obtained for different complexities which shows the proposed model performed consistently better as the neuros increased. The Mean Absolute Error (MAE) values shown in Fig. 2 is proven the positive correlation since the MAE values are truly positive and indicate the result is quite near to the optimal solution. In Fig 3, The Sum of Squared Error (SSE) values obtained by HNN-3SATBCD with different array of number of clause parameters. Hence, the proposed method was performed consistently better in training the breast cancer data set. The robustness of our 3-SATRA algorithms can be approximately demonstrated base on the effectiveness of the entire computational process. As the number of neurons increases, the CPU Time also increases for all the hybrid networks. The ACCURACY recorded for HNN-3SATBCD were 100%. The HNN-3SAT with 3-SATRA approach to the Breast cancer Coimbra is proven the performance of the proposed model and it can be implemented by medical practitioners as an early screening tool.

VI. CONCLUSIONS

The proposed algorithm, HNN-3SATES has demonstrated a good performance based on RMSE, MAE, SSE and CPU time obtained during training and testing process. as such, the proposed hybrid HNN method can be utilized as classification tool for medical data set. Collectively, the hybrid HNN models with 3-SATRA is proven as a robust paradigm according to the experimental results. Overall, this paperwork has authenticated the capability of our proposed hybrid HNN as a tool for data mining. Thus, in the future research, the 3-SATRA data mining can be carried out in other variants of neural network such as convolution neural network (CNN), Wavelet neural network, Deep neural network and a probabilistic neural network. The model can be extended via metaheuristic training method to accelerate classification of complex data set such as streaming and time series data set.

REFERENCE

[1] G. I. Salama, M. B. Abdelhalim, and M. A. Zeid, "Using Multi-Classifiers," 2012.
[2] K. Kutzkov, "New upper bound for the #3-SAT problem," *Inf. Process. Lett.*, vol. 105, no. 1, pp. 1–5, 2007.
[3] M. A. Mansor, S. Sathasivam, and M. S. M. Kasihmuddin, "3-Satisfiability Logic Programming Approach for Cardiovascular Diseases Diagnosis," *AIP Conf. Proc.*, vol. 1974, 2018.
[4] M. A. Mansor and S. Sathasivam, "Accelerating Activation Function for 3- Satisfiability Logic Programming," *Int. J. Intell. Syst. Appl.*, 2016.
[5] S. Sathasivam and W. A. T. Wan Abdullah, "Logic mining in neural network: Reverse analysis method," *Comput. (Vienna/New York)*, 2011.
[6] R. ul Rojas, *Neural Networks: A Systematic Introduction Springer*. 2013.
[7] J. L. Johnson, "A neural network approach to the 3-satisfiability problem," *J. Parallel Distrib. Comput.*, 1989.
[8] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities Biophysics," *Proc. Natl. Acad. Sci.*, 1982.
[9] W. A. Little, "The Existence of Persistent States in the Brain I . INTRODUCTION In this paper we examine the long term behavior of a neuronal network such as the human brain . We will start from the assumption that the state of the brain at any time may be described by," vol. 120, no. 1974, pp. 145–146, 1974.
[10] S. Sathasivam, "Application of Neural Networks in Predictive Data Mining," *2011 2nd Int. Conf. Bus. Econ. Res.*, pp. 371–376, 2011.
[11] S. Sathasivam, "Logic Mining in Neural Networks," 2006.