

3-Satisfiability Data Mining Technique with Hybrid Clonal Selection Algorithm for Autistic Spectrum Disorder Screening

Mohd Asyraf Mansor^{1*}, Mohd Shareduwan Mohd Kasihmuddin², Aslina Baharum³, Saratha Sathasivam⁴

¹*School of Distance Education, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia*

^{2,4}*School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Pulau Pinang, Malaysia*

³*Faculty of Computing and Informatics, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia*

Autistic Spectrum Disorder (ASD) is a variant of neurodevelopment and behavioural impairments that mostly affected the children and adolescence. The screening of ASD in adulthood is regularly more challenging as it may overlap with the mental-health disorder symptoms. In this paper, 3-Satisfiability data mining approach with hybrid clonal selection algorithm in Hopfield neural network is proposed. The core impetus of this research is to investigate the feasibility and effectiveness of the proposed model in the task of Autistic Spectrum Disorder detection. The performance of the proposed approach will be compared with the state-of-the-art 3-Satisfiability data mining with exhaustive search algorithm in Hopfield neural network. The simulations were carried out on the Autistic Spectrum Disorder (ASD) among adults' data sets provided by UCI machine learning repository. The comparison results depicted overall improvements of accuracy and robustness by the proposed model in Autistic Spectrum Disorder (ASD) diagnosis. This research will provide the useful insight as it can be further applied in the diagnosis of more vigilant diseases.

Keywords: 3-Satisfiability, Autistic Spectrum Disorder, Data Mining, Clonal Selection Algorithm

I. INTRODUCTION

Disease detection analytics have attracted prolific works since the era of industrial revolution 4.0 in recent years. The machine learning and neural network approach became a staple in assisting the health professional to do earlier screening before pursuing with the conventional clinical diagnosis. Thus, the screening of Autistic Spectrum Disorder (ASD) among adult via 3-Satisfiability data mining and metaheuristic algorithm with Hopfield network will set the tone of this research.

The research on ASD screening has been flourishing

starting from Ahmadrlo (2010), who proposed a wavelet-chaotic neural network as a paradigm to process the EEG signal of ASD. Recently, Thabtah et al. (2018) inaugurated the variable analysis paradigm in early ASD screening process via "feature to class" correlation approach with acceptable accuracy. Thus, we require a logic extraction method to understand the behaviour of the data to facilitate the ASD screening process.

Precisely, 3-Satisfiability is a variant of Boolean logic that can utilized to translate a specific value into the binary decision system (Mansor et al., 2017). Hence, the flexibility of 3-Satisfiability to map the real data into

*Corresponding author's e-mail: asyrafman@usm.my

logical form is our core impetus of using 3-Satisfiability reverse analysis paradigm in ASD data mining.

Hopfield neural network (HNN) is well known as a standard computing model that impersonates the way of human brain processing the information (Hopfield, 1982). Theoretically, the Hopfield neural network is a powerful network due to the flexibility of the network to works with metaheuristics and data mining approach (Kasihmuddin et al., 2018). The capabilities of the Hopfield neural network to truncate the outliers and store the data effectively via Content Addressable Memory are essential in knowledge discovery or logic mining (Sathasivam, 2010). However, on the theoretical side, Cabrera and Sossa (2018) have enhanced the stable states of HNN to be a robust network.

The research on the clonal selection algorithm (CSA) is motivated by the capability of the optimization operator in performing faster computational and searching task (Layeb, 2012). CSA is a class of evolutionary algorithm and metaheuristic method, inspired by the mechanism in the immune system fighting the pathogen (Dasgupta et al., 2011).

In this paper, we will propose a 3-Satisfiability data mining technique namely 3-Satisfiability reverse analysis paradigm (3-SATERAP) incorporated with hybrid HNN models in training and testing ASD screening data. The hybrid RDASHNN model with clonal selection algorithm is abbreviated as RDASHNN-3SATCSA and being compared with hybrid HNN with exhaustive search, RDASHNN-3SATES. The RDASHNN-3SATCSA is developed due to the capability of clonal selection algorithm in improving the convergence of the existing reverse analysis data mining technique in Hopfield neural network as proposed by Sathasivam (2010).

II. 3-SATISFIABILITY PROBLEM

3-Satisfiability (3-SAT) can be defined as a Boolean algebra based logical rule, each clause containing strictly 3 literals or variables (Brueggemann and Kern 2004). For example:

$$P = (\bar{R} \vee S \vee T) \wedge (U \vee \bar{V} \vee W) \wedge (X \vee Y \vee \bar{Z}) \quad (1)$$

Equation (1) shows a 3-SAT formula, P with 3 clauses

associated by *AND* operators. Therefore, each clause containing 3 different literals connected by OR operators. Henceforth, the number of neurons corresponds to the total number of literals involved in a 3-SAT logic programming (Fu et al., 2018).

Therefore, Aiman and Asrar (2015) introduced the generalized form of 3-SAT formula is shown in Equation (2).

$$P = \bigwedge_{i=1}^n T_i \quad (2)$$

whereby T symbolizes the clause and I denote the number of clauses involved in 3-SAT formula P . In this research, the attributes of ASD data set will be represented by the literals of each 3-SAT clause.

III. HOPFIELD NEURAL NETWORK

Hopfield neural network (HNN) is classified as a recurrent neural network, extensively being implemented on logic programming and data mining. The fundamental neuron updating rule in HNN is given as follows:

$$S_i = \begin{cases} 1 & \text{if } \sum_j W_{ij} S_j > \psi_i \\ -1 & \text{Otherwise} \end{cases} \quad (3)$$

where W_{ij} is the synaptic weight from unit j to i . S_i is the state of unit I and ψ_i is the threshold of unit i . The important features are the connection in the HNN has no connection with itself, $W_{jj} = W_{ii} = 0$ and connections are symmetrical or bidirectional (Sathasivam & Abdullah, 2011). The neurons in HNN are bipolar and can be denoted as $S_i \in \{1, -1\}$ as the states. The spin points of the neuron flip until equilibrium is reached (Kasihmuddin et al., 2018). In terms of 3-SAT representation, each variable in 3-SAT formula will be represented in terms of N neurons. The synaptic weight resembles the connection between the variable and the clauses in 3-SAT formula (Pinkas, 1991). The connection model can be generalized to embrace higher order connection. Hence, the local field

function of HNN for 3-SAT logic is given:

$$h_i = \dots + \sum_j W_{ijk}^{(3)} S_j S_k + \sum_j W_{ij}^{(2)} S_j + W_i^{(1)}$$

(4) Following that, the updating rule maintains as follows:

$$S_i(t+1) = \text{sgn}[h_i(t)] \quad (5)$$

where h_i is the local field of the network. The final state of the neurons will be examined by using Lyapunov energy function $H_{P_{3-SAT}}$ which is given:

$$H_{P_{3-SAT}} = \dots - \frac{1}{3} \sum_i \sum_j \sum_k W_{ijk}^{(3)} S_i S_j S_k - \frac{1}{2} \sum_i \sum_j W_{ij}^{(2)} S_i S_j - \sum_i W_i^{(1)} S_i \quad (6)$$

The updating rule in Equation (5) guarantees the energy function in Equation (6) will decrease monotonically with the network (Mathias & Rech, 2012). The synaptic weight of the network is computed by using Wan Abdullah Method (Abdullah, 1993).

IV. CLONAL SELECTION ALGORITHM

The clonal selection algorithm (CSA) is a variant of artificial immune system algorithm (AIS), enthused by the process of clonal selection and hypermutation in B-cells in biological immune system. Hence, Layeb (2012) has inaugurates the systematic CSA as a searching mechanism regardless of the complexities. Therefore, the CSA algorithm has been proven to work well in 3-Satisfiability logic programming due to the robustness of somatic hypermutation operator (Mansor et al., 2017).

INPUT

100 strings of 3-SAT interpretation.

Step 1 (Initialization of B-Cells)

Initialize 100 B-cells (interpretations).

Step 2 (Affinity Calculation)

The affinity can be calculated by using:

$$f_{\text{affinity}} = c_1(x) + c_2(x) + c_3(x) \dots + c_{\text{total } NC}(x) \quad (7)$$

where c represents the number of clauses during learning phase and NC symbolizes the number of clauses embedded into the 3-SAT formula.

Step 3 (B-cells Selection)

Select the best 5 B-cells with the highest value of affinity.

Step 4 (Cloning)

The cloning begins by replicating the selected B-cells by using the conventional roulette wheel selection to the system (Layeb, 2012). Moreover, the number of cloned B-cells can be computed as follows:

$$\left(\begin{array}{l} \text{The number of} \\ \text{clone allowed} \end{array} \right) = \frac{\text{affinity}_i}{\sum \text{affinity}} \times \beta \quad (8)$$

where affinity_i denotes to the initial affinity observed by the system and β is the number of population clone that is introduced to the system. By taking the parameter from the work of Layeb (2012), $\beta = 200$ is selected.

Step 5 (Normalization)

The normalization of the B-cells is also termed maturation of the immune response. The ordinary formulation for B-cells normalization is shown in Equation (9).

$$\text{affinity } N_i = \frac{\text{affinity}_i - \min \text{affinity}}{\max \text{affinity} - \min \text{affinity}} \quad (9)$$

where $\text{affinity } N_i$ refersthe normalized affinity.

Step 6 (Somatic Hypermutation)

Somatic hypermutation operator will enhance B-cells to achieve the maximum affinity that will converge towards a feasible solution. Specifically, the B-cells (interpretation) flipping will enhance the solution to obtain the maximum affinity values (Ulutas & Kulturel-Konak, 2011). That mechanism works by randomly flipping the one or more string of the B-cells from -1 to 1 or vice versa. The number of mutations can be determined by Equation (10).

$$\left(\begin{array}{l} \text{Number of} \\ \text{Mutation (Nb)} \end{array} \right) = \left(\frac{1}{\text{Number of variable}} \right)_{(10)}$$

$$(\text{affinity } N_i) + (1 - \text{affinity } N_i)(0.01)$$

Next, the affinity of the new generation of B-cells will be calculated. Consequently, if a particular of B-cells fails to attain the required affinity, Step 1 until Step 6 will be repeated until $f_{affinity} = NC$ where NC is the number of clauses.

OUTPUT

The best interpretation with $f_{affinity} = NC$ to be stored in HNN memory.

V. EXPERIMENTAL SETUP

RDASHNN-3SAT hybrid models will be simulated in Dev C++ Version 5.11 as a platform and carried out on Windows 10.1, Intel Core i7, 3.0 GHz processor with 8GB of RAM.

In this study, we will use the Autistic Spectrum Disorder (ASD) obtained from UCI machine learning repository, contributed by Thabtah et al. (2018). In every execution, 60% of the data instances are entrenched as the training data and the remaining 40% of the data points are utilized as testing data. The implementations are:

Step 1

Transform 3-SAT clauses into Boolean algebra form and assign the neuron to each of variable. Select 422 data points as the training data and 282 as the testing data of ASD data set. Convert all binary data set to bipolar form where 0 transforms as -1 and 1 remains as 1.

Step 2

Assign the bipolar value to the neuron state and segregate the collection of k neurons per clause $\{c_1, c_2 \dots c_n\}$ that lead to $P_{training} = 1$.

Step 3

Generate the best logic P_{Best} as an ideal representation of the ASD data set. Enumerate the variables of the best logic with neurons. Find the cost function of the optimum logic, $E_{P_{Best}}$.

Step 4

Check the clause satisfaction of $E_{P_{Best}} = 0$. The RDASHNN-3SATES will be trained by ES and the RDASHNN-3SATCSA will be trained by CSA. The data pattern will be stored in Hopfield CAM.

Step 5

Derive a cost function the best logic, $E_{P_{Best}}$ by considering

$$X = \frac{1}{2}(1 + S_X) \quad \text{and} \quad \bar{X} = \frac{1}{2}(1 - S_X).$$

Henceforth, the states of the neurons can be classified as $S_X = 1$ (True) and $S_X = -1$ (False).

Step 6

Derive the best cost function to be compared with the energy function. The synaptic weights of P_{Best} are computed by Wan Abdullah method.

Step 7

The states of the neurons are randomized. The network undergoes sequences of network relaxation. Calculate the corresponding local field $h_i(t)$ of the state by using Equation (4).

Step 8

Induce all possible 3-SAT logic from the neuron states to retrieve the induced logic, P_i^B . By employing test dataset, examine all the induced logic P_i^B by comparing the outcome of the induced logic with the actual outcome.

Step 9

Compute the corresponding learning and testing RMSE, MAE and SSE of the network including the computational time.

The control experiment in this research is the number of trials per execution is set to 1000. In order to avoid bias, the simulations are conducted by using similar computer.

VI. RESULTS AND DISCUSSIONS

A. Training Error

In this paper, the root mean square error (RMSE), mean absolute error (MAE) and sum of squared error (SSE) for the training phase deployed by RDASHNN models are recorded from $NN=9$ until $NN=90$.

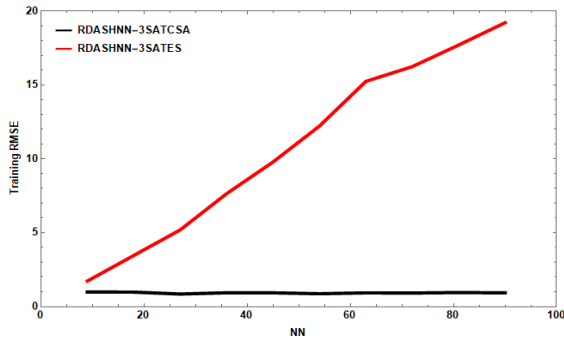


Figure 1. Training RMSE of RDASHNN models.

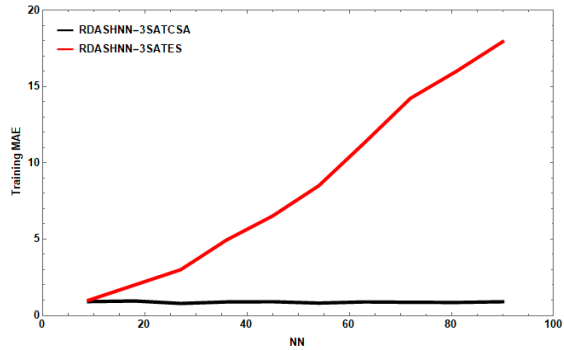


Figure 2. Training MAE of RDASHNN models.

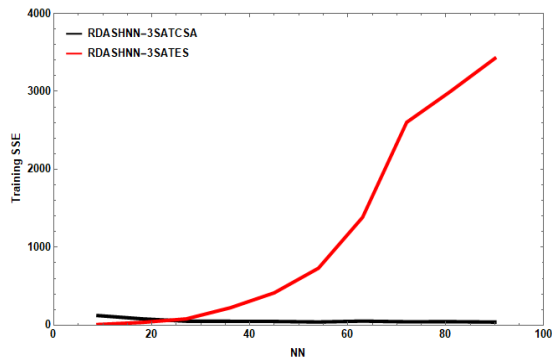


Figure 3. Training SSE of RDASHNN models.

Figure 1 depicts the RMSE recorded by RDAHNN models during training phase in generating the best logic to represent the ASD dataset. From the simulations, RDASHNN-3SATCSA has consistently smaller RMSE compared to RDASHNN-3SATES. The deviation of error in RDASHNN-3SATCSA is less significant because of less iterations needed to complete the training phase. On the contrary, RDASHNN-3SATES requires more iteration before generating the best logic resulting in massive error accumulations. In this aspect, the mechanism of ES which is the “generate and enumerate” procedures will consume time to achieve the convergence. Figure 2 demonstrates that the magnitude of MAE for RDASHNN-3SATCSA are

significantly lesser than the other counterpart even though the number of neurons (NN) have been increased. Figure 3 portrays the value of SSE recorded by both models. The RDASHNN-3SATES recorded the highest accumulated errors especially during training phase. The proposed approach, RDASHNN-3SATCSA is less prone towards error due to somatic hypermutation that reduce iterations during training to generate the best logic.

B. Testing Error

The testing error for RDASHNN-3SATES and RDASHNN-3SATCSA were consistently similar for each of $NC=1$ until $NC=8$.

Table 1. Testing Error for RDASHNN models.

Models	Testing RMSE	Testing MAE	Testing SSE
RDASHNN-3SATES	6.2526	0.3723	11025.00
RDASHNN-3SATCSA	2.7988	0.1667	2209.00

Table 1 manifests the accuracy of our proposed 3-Satisfiability data mining paradigm, 3-SATERAP incorporated with RDASHNN-3SATCSA and the conventional method, RDASHNN-3SATES in generating the best logic for ASD data set. The training errors recorded by RDASHNN-3SATCSA were consistently smaller that ease the retrieval during testing phase. The testing RMSE, testing MAE and testing SSE are superior performance indicator to indicate the accuracy and sensitivity of out proposed approach. It shown that RDASHNN-3SATCSA outperformed RDASHNN-3SATES in generating the best logic from ASD data set. The results obtained by RDASHNN-3SATES were still satisfactory, but the higher RMSE, MAE and SSE recorded during training indicated that the network produce a moderate logic.

C. Computation Time

The robustness of the models in logic mining can be determined by analysing the computational time (CPU time) taken by the models during training and testing

phase of RDASHNN models under different complexities.

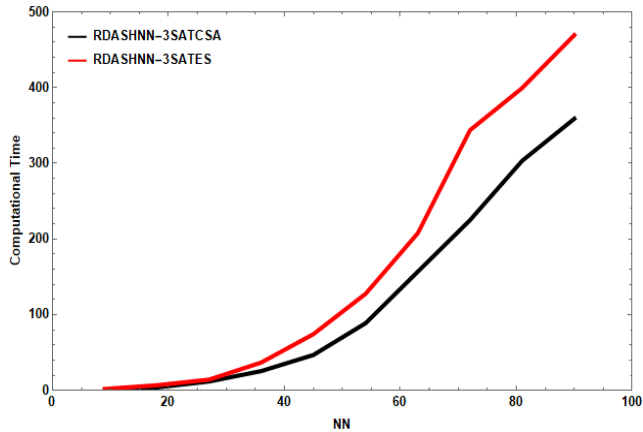


Figure 4. Computational Time of RDASHNN models.

Figure 4 manifests the computational time recorded by RDASHNN-3SATES and RDASHNN-3SATCSA during comprehensive simulation by manipulating different number of neurons (NN) and complexities. In this investigation, the computational time were basically less than 500 seconds. Based on Figure 4, RDASHNN-3SATCSA generates and computes the best logic faster as compared to RDASHNN-3SATES. The capability of CSA to reduce the iterations during training and retrieve the best logic for a minimal testing error is due to the somatic hypermutation operator. Henceforth, RDASHNN-3SATES requires a substantial amount of time due to neuron oscillations during testing phase resulting in more iterations. Thus, more time are required to generate the best logic and complete one execution of training and testing phase.

VII. CONCLUSION

In this paper, we have investigated the performance of 3-Satisfiability data mining technique namely 3-SATERAP with RDASHNN-3ATCSA and RDASHNN-3SATES in training and testing the autistic spectrum disorder data set. The proposed method, RDASHNN-3SATCSA outperforms the other counterpart, RDASHNN-3SATES in term of training errors, testing errors and computational time. In addition, our proposed method can be applied in screening more vigilant diseases such as seizures and heart diseases.

VIII. ACKNOWLEDGEMENT

This research is supported by School of Distance Education, Universiti Sains Malaysia and Short-Term Research Grant (304/PJJAUH/6315263) by Universiti Sains Malaysia.

IX. REFERENCES

- Abdullah, W.A.T.W. (1993). The logic of neural networks. *Physics Letters A*, 176(3-4),202-206.
- Ahmadlou, M., Adeli, H. and Adeli, A. (2010). Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of autistic spectrum disorder. *Journal of Clinical Neurophysiology*, 27(5),328-333.
- Aiman, U. and Asrar, N. (2015). Genetic algorithm-based solution to SAT-3 problem. *Journal of Computer Sciences and Applications*, 3(2),33-39.
- Brueggemann, T. and Kern, W. (2004). An improved deterministic local search algorithm for 3-SAT. *Theoretical Computer Science*, 329(1-3),303-313.
- Cabrera, E. and Sossa, H. (2018). Generating exponentially stable states for a Hopfield Neural Network. *Neurocomputing*, 275,358-365.
- Dasgupta, D., Yu, S. and Nino, F. (2011). Recent advances in artificial immune systems: models and applications. *Applied Soft Computing*, 11(2),1574-1587.
- Fu, H., Xu, Y., Wu, G., Jia, H., Zhang, W. and Hu, R. (2018). An Improved Adaptive Genetic Algorithm for Solving 3-SAT Problems Based on Effective Restart and Greedy Strategy. *Int. J. Comput. Intell. Syst.*, 11(1),402-413.
- Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8),2554-2558.
- Kasihmuddin, M.S.M., Mansor, M.A. and Sathasivam, S. (2018). Discrete Hopfield Neural Network in Restricted Maximum k-Satisfiability Logic Programming. *Sains Malaysiana*, 47(6),1327-1335.
- Layeb, A., 2012. A clonal selection algorithm based tabu search for satisfiability problems. *Journal of Advances in Information Technology*, 3(2), pp.138-146.
- Mansor, M.A., Kasihmuddin, M.S.M. and Sathasivam, S., 2017. Artificial Immune System Paradigm in the Hopfield Network for 3-Satisfiability Problem. *Pertanika Journal of Science & Technology*, 25(4), pp.1173-1188.
- Mathias, A.C. and Rech, P.C., 2012. Hopfield neural network: the hyperbolic tangent and the piecewise-linear activation functions. *Neural Networks*, 34, pp.42-45.
- Pinkas, G., 1991. Symmetric neural networks and propositional logic satisfiability. *Neural Computation*, 3(2), pp.282-291.
- Sathasivam, S., 2010. Upgrading logic programming in Hopfield network. *Sains Malaysiana*, 39(1), pp.115-118.
- Sathasivam, S. and Abdullah, W.A.T.W., 2011. Logic mining in neural network: reverse analysis method. *Computing*, 91(2),119-133.
- Thabtah, F., 2018. Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Informatics for Health and Social Care*, pp.1-20.
- Thabtah, F., Kamalov, F. and Rajab, K., 2018. A new computational intelligence approach to detect autistic features for autism screening. *International journal of medical informatics*, 117, pp.112-124.
- Ulutas, B.H. and Kulturel-Konak, S., 2011. A review of clonal selection algorithm and its applications. *Artificial Intelligence Review*, 36(2), pp.117-138.