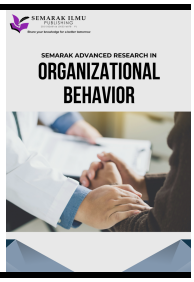
SEMARAK ILMU PUBLISHING

Semarak Advanced Research in Organizational Behaviour

Journal homepage:
https://semarakilmu.my/index.php/sarob/index
ISSN: 3030-6264



Extracting Absenteeism at Work using Random 2 Satisfiability Modified Reverse Analysis

Inaz Nazifa Dzulkifli[1], Mohd Shareduwan Mohd Kasihmuddin[1,*], Yueling Guo[2], Nur Ezlin Zamri[3], Nurul Atiqah Romli[1]

[1] School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM, Penang, Malaysia
[2] School of Science, Hunan Institute of Technology, 421002 Hengyang, China
[3] Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM, Serdang, Selangor, Malaysia

ARTICLE INFO

Article history:
Received 20 November 2024
Received in revised form 8 December 2024
Accepted 15 December 2024
Available online 31 December 2024

Keywords:
Absenteeism at work; modified evolutionary algorithm; knowledge extraction; discrete Hopfield neural network; random 2 Satisfiability modified reverse analysis; performance metrics

ABSTRACT

Absenteeism in the workplace can be voluntary or involuntary which significantly impacts organization productivity, employee morale, and operational costs. An artificial neural network will be employed to extract insights on absenteeism helping the employers to understand and mitigate the issue. There are limited attempts that propose neural network model for knowledge extraction in the human resources domain, particularly research revolving around absenteeism of employees. In this paper, a modified evolutionary based algorithm named Modified Niche Genetic Algorithm has been proposed to enhance the training phase of the Discrete Hopfield Neural Network with Random 2 Satisfiability Modified Reverse Analysis method. The performance of the Modified Niche Genetic Algorithm with the Random 2 Satisfiability Modified Reverse Analysis is evaluated using various performance metrics. In light of the findings, the proposed model demonstrated superior performance compared to the traditional classification model with 73% more accuracy in doing knowledge extraction for the human resources field. The contributions in the proposed model provide a robust platform and enhance the capabilities of the classification model.

1. Introduction

Through the lens of Artificial Intelligence (AI), Artificial Neural Network (ANN) serve as the backbone of computational machine learning systems that are capable of handling large data-intensive environments and offer insight in relationship discovery [1-3]. This is because ANN is inspired by the structure and human brain function. A study by Ye *et al.,* [4] shows that ANN is utilized to perform classification tasks by finding the patterns and trends within large dimensional datasets. The core component of ANN is the interconnected nodes in the network known as artificial neurons. These artificial neurons are structured into two or more layers which significantly have a layer known as input and output which depicts the process of the information through the network. The synaptic

* *Corresponding author.*
E-mail address: shareduwan@usm.my

https://doi.org/10.37934/sarob.4.1.1330a

weights represent the connections between these neurons and determine the importance of each neuron connection in the neural network. Noteworthy, the synaptic weights are produced during the training operations of ANN, and stored in storage feature of ANN to be retrieved or recalled during the retrieval phase. Additionally, the retrieval phase consists of a non-linearity property where the interactions between initial neuron states and stored synaptic weights are evaluated in the phase. Subsequently, the complex relationship between the input and output is simplified using the activation function. For instance, work by Rusdi *et al.,* [5] applied Hyperbolic Tangent Activation Function (HTAF) to normalize continuous output values into more interpretable binary values. ANN can be trained on data of any dimensional and can be adapted to any changes in order to improve performance over time. ANN requires fewer manual data handling and sturdy against data anomalies compared to traditional statistical methods. A well-trained ANN can efficiently handle complex tasks with minimal human intervention. Due to these reasons, ANN has become a versatile tool for solving various problems.

The use of rules in governing neuron connections within a Discrete Hopfield Neural Network (DHNN) is crucial due to the black box property. Abdullah [6] states that these rules ensure that neurons are accurately represented by logical constructs. If the logical conditions are not implemented effectively, it can lead to issues like poor synaptic weight management and overfitting. This typically results in the network converging to a singular type of solution as supported by Mansor *et al*., [7]. Over the years, various Satisfiability (SAT) logic structures, such as Horn Satisfiability (HornSAT) by Sathasivam [8], 2 Satisfiability (2SAT) by Kasihmuddin *et al.,* [9], and 3 Satisfiability (3SAT) by Mansor *et al.,* [7] have been embedded in DHNN to address these challenges. However, each has its drawbacks, such as redundancy in HornSAT leading to ineffective neuron connections and the easy approximation of synaptic weights in systematic SAT structures which diminishes the functionality of the neurons. In response to the limitations observed with systematic SAT structures, an unsystematic SAT known as Random 2 Satisfiability (RAN2SAT) proposed by Sathasivam *et al.,* [10] allowed for greater variability in synaptic weight characteristics by randomly generating clauses with positive and negated variables. Despite its potential, RAN2SAT struggled with generating specific combinations of variables required for certain experiments, limiting its ability to explore neuron behaviours comprehensively. This inability to produce desired variable combinations highlights a broader oversight in the field where the impact of negated variables on the logical conditions governing DHNN is often underestimated, possibly leading to faults or errors in SAT applications as reviewed in existing work by Saribatur *et al.,* [11]. Given the ongoing challenges in effective synaptic weight management and neuron representation, there is a pressing need for a new approach in SAT logic for DHNN. Proposing an unsystematic SAT rule that dynamically manages the distribution of negated variables could significantly enhance the network generalizability and pattern recognition capabilities. Such a dynamic structure not only improves the storage quality of synaptic weights but also increases the logical condition power to represent and uncover new patterns. The existing work by Zamri *et al.,* [12] shows this approach necessitates a thoughtful design of SAT logic that controls the impact of negated variables to ensure more robust and accurate neuron behaviour in DHNN.

Logic mining is an innovative optimization strategy developed to extract patterns from real-life datasets by using logical formulations. Unlike traditional black box models, logic mining utilizes the classification capabilities of the Satisfiability logical condition in the DHNN model to enhance knowledge representation. This is achieved by converting the final neuron states into a logical condition known as induced logic. This approach was first introduced by Sathasivam and Abdullah [13] through a model that combined DHNN with HornSAT using a Reverse Analysis (RA) approach. This RA method prioritized neuron states based on their global energy profiles before transforming them into induced logic. However, this method had several drawbacks: it was unable to process

continuous or non-categorical data due to the lack of data cleaning and produced an excessive number of overlapping induced logics without evaluating their effectiveness. Additionally, the use of HornSAT compromised the interpretability of the induced logic due to redundant variables. In attempts to refine this approach, Kho *et al.,* [14] and Jamaludin *et al.,* [15] shifted to using 2SAT as the logical condition in DHNN focusing on accuracy as a performance metric. These modifications enabled the extraction of good induced logic which generalizes the pattern of the dataset. Nevertheless, these models often overlooked data patterns associated with failures or negative outcomes, leading to a biased classification model that predominantly recognized true positives. This non-holistic approach excluded important patterns linked to true negatives. To address these challenges and reduce the risk of overfitting associated with systematic SAT formulations, particularly in higher orders of clause denoted as *k*, the logic mining model needs to be further developed to handle non-Horn and unsystematic SAT rules. This improved model should also manage continuous data and identify a singular optimal induced logic that accurately represents both negative and positive dataset outcomes.

In another development, Jamaludin *et al.,* [16] improved the logic mining model by implementing the energy analysis when extracting the induced logic. Energy-based 2 Satisfiability Reverse Analysis Method (E2SATRA) is the work by Jamaludin *et al.,* [16] which focuses on the E-recruitment dataset of the insurance agency in Malaysia. E2SATRA implemented the energy operator in the retrieval phase which only extracts induced logic from the final solution that achieves global minimum energy. The primary goal is to identify features that promote to positive recruitment outcomes using the E2SATRA model. However, the work by Jamaludin *et al.,* [16] has several notable gaps and inconsistencies. The proposed E2SATRA does not explicitly describe a robust feature selection method, which might lead to the presence of redundant or unrelated features that can negatively affect the performance of the model. There are also concerns regarding the generalizability and risk of overfitting in the energy-based logic mining method. The rigid logical conditions and limited search space for optimal solutions may make the model prone to overfitting. Additionally, the energy-based approach lacks a detailed explanation of the training phase optimization, particularly in achieving global minima in neuron states. Also, the evaluation metrics and statistical analysis also show gaps in the energy-based paper. The evaluation metrics might not be as comprehensive and lacking detailed statistical analysis to validate the model performance across various datasets. This contrasts with the approach by Zamri *et al.,* [17] that introduces the Jaccard Feature Selection Method (JFSM) to guarantee features with high similarity to the dataset outcomes are selected. However, JFSM only considers the presence of a scoring mechanism but completely disregards the absences. JFSM can be less intuitive in fields of the absence feature since JFSM completely disregards this case. Notably, the issues with implementing logical conditions into DHNN persist and further study needs to be carried out to overcome the stated issues.

Although various approaches have been explored in the existing studies, there are significant areas that have yet to be thoroughly explored such as controlling the distribution of negative variables to enhance the effectiveness of the synaptic weight management. Next, the consideration of negative features in the feature selection method, and the inclusion of positive and negative outcome in finding the best logic. Given the listed challenges, this paper introduces a novel logic mining model that incorporates several contributions in the preliminary and training phases of the model. Unsupervised feature selection method is introduced in the preliminary phase. Then, multiple selection best logic is proposed from the innovative computation of best logic with a multi-objective training approach. Therefore, this paper main contributions are outlined as follows

1. To propose Weighted Random 2 Satisfiability logic in Discrete Hopfield Neural Network as neuron representation. A weighted feature namely ratio of negated variables, *r* is introduced in the proposed logical condition to govern the number of negated variables.
2. To propose a logic mining model namely Modified Weighted Random 2 Satisfiability Reverse Analysis in doing classification task for Absenteeism at Work dataset. In order to assure optimal training phase, a Modified Niched Genetic Algorithm is added to the training phase of the proposed logic mining model.
3. To assess the performance of Modified Weighted Random 2 Satisfiability Reverse Analysis by considering the evaluation of confusion matrix classifications based on the retrieved final induced logic to the actual outcome of the Absenteeism at Work dataset. The optimal induced logic is based on the highest retrieval frequency from all *K*-folds cross validation.

2. Methodology

The proposed methodology to provide knowledge extraction to the absenteeism at a workplace will be covered in this section. The methodology will discuss the formulation of Weighted Random 2 Satisfiability (*r*2SAT), the general concepts of *r*2SAT in DHNN (DHNN-*r*2SAT), and the Modified Weighted Random 2 Satisfiability Reverse Analysis.

2.1 The Formulation of Weighted Random 2 Satisfiability

Weighted Random 2 Satisfiability (*r*2SAT) is an unsystematic SAT that consists of first and second-order SAT. Notably, the element in each clause includes *k* variables where *k* = 1, 2. The overall formulation of *r*2SAT is expressed in Eq. (1) as follows

$$\Gamma_{r2SAT} = \wedge^{u}_{i=1} J_i^{(k=1)} \wedge^{v}_{i=1} J_i^{(k=2)}, \tag{1}$$

where $J_i^{(k=1)}$ and $J_i^{(k=2)}$ are the clauses from first and second-order in *r*2SAT, respectively. The total clauses in *r*2SAT are denoted as *c* such that *c* = *u* + *v*. The variables in each $J_i^{(k=1,2)}$ is shown in Eq. (2) as follows

$$J_i^{(k=1,2)} = \begin{cases} \left(A_i \vee B_i\right), & k = 2 \\ C_i, & k = 1 \end{cases}. \tag{2}$$

Every variable can be either negative $\left(\neg A_i\right)$ or positive $\left(A_i\right)$. The combined variables in Γ_{r2SAT} can be calculated via Eq. (3) as follows

$$\lambda_{r2SAT} = u + 2v. \tag{3}$$

The second key characteristics of Γ_{r2SAT} is the allocation of negated variables is determined dynamically in the logical condition. The amount of negated variables generated is governed by fixed ration negated variables, *r* as described in paper by Zamri *et al.,* [12]. Then, Genetic Algorithm (GA) will be used in the logic phase to minimize Eq. (4) as follows

$$\min\left[f_{LP}\right]=\left|\kappa-N_{v}\right|. \tag{4}$$

The minimization of Eq. (4) is important to attain desired fitness $f_{LP}=0$. Eq. (5) shows one of the examples of Γ_{r2SAT} that have been computed.

$$\Gamma_{r2SAT}=\left(A_1\vee B_1\right)\wedge\left(\neg A_2\wedge B_2\right)\wedge\left(\neg A_3\wedge B_3\right)\wedge\left(A_4\vee B_4\right)\wedge C_1\wedge C_2. \tag{5}$$

Hence, the computed Γ_{r2SAT} will correspond to the neurons in the process of the DHNN.

2.2 The Process of Weighted Random 2 Satisfiability in Discrete Hopfield Neural Network

Delgado-Currín *et al.,* [18] state that the structure of DHNN is easy to understand where the output is directly related to the input with no hidden layers involved. DHNN-*r*2SAT have two main phases, which are the training phase and the retrieval phase. Wan Abdullah (WA) method is utilized in the training phase to train Γ_{r2SAT} to guarantee optimal management of synaptic weight. Then, an associative memory feature in DHNN known as Content Addressable Memory (CAM) is used to store the produced synaptic weights before being retrieved in the retrieval phase of DHNN [19]. The synaptic weight will be used in the computation of the local field to utilize the final states. The objective function in the DHNN-*r*2SAT is formulated in Eq. (6) and Eq. (7) as follows

$$\Phi_{\Gamma_{r2SAT}}=\frac{1}{4}\sum_{i=1}^{u}\left(\prod_{j=1}^{2}\mathrm{Z}_{ij}\right)+\frac{1}{2}\sum_{i=1}^{v}\left(\prod_{j=1}^{1}\mathrm{Z}_{ij}\right), \tag{6}$$

$$\mathrm{Z}_{ij}=\begin{cases}\left(1-S_{A_i}\right) & \text{if } \neg A_i\\ \left(1+S_{A_i}\right) & \text{if } A_i\end{cases}. \tag{7}$$

Successfully minimizing the objective function or $\min\left[\Phi_{\Gamma_{r2SAT}}\right]$ will guarantee optimal synaptic weights in DHNN. Note that, every clause in the Γ_{r2SAT} is satisfied when $\Phi_{\Gamma_{r2SAT}}=0$ (zero objective function) while $\Phi_{\Gamma_{r2SAT}}\neq 0$ depicts the number of unsatisfied clauses. During the retrieval phase, the process of evaluating the final state happens in the retrieval phase of DHNN-*r*2SAT. The final state is retrieved using the local field $\left(\chi_i\right)$ computation as shown in Eq. (8)

$$\chi_i\left(t\right)=\sum_{j=i,i\neq j}^{\lambda_{r2SAT}}W_{ij}^{(2)}S_j+W_i^{(1)}, \tag{8}$$

where $W_{ij}^{(k)}$ is the synaptic weight of the DHNN-*r*2SAT based on the orders in each clause involving neurons *i* and *j*. Then, the values of χ_i is normalized into a binary state using Hyperbolic Tangent activation function (HTAF). The energy profile of the final states is evaluated using Eq. (9) as follows

$$H_{\Gamma_{r2SAT}} = -\frac{1}{2}\sum_{i=1,i\neq j}^{N}\sum_{j=1,i\neq j}^{N} W_{ij}^{(2)} S_i S_j - \sum_{i=1}^{N} W_i^{(1)} S_j \,. \tag{9}$$

The final states retrieved will be impacted by the optimality of the synaptic weights. All the fundamental components of DHNN-*r*2SAT have been covered and will serve as the foundation for the proposed logic mining model.

2.3 The Modified Weighted Random 2 Satisfiability Reverse Analysis

The Modified Weighted Random 2 Satisfiability Reverse Analysis (*r*2SATMRA) is the proposed logic mining model that consists of three main domains. First, the feature selection method by the Simple Matching Selection Method (SMSM). Second, all the fundamental components in the DHNN-*r*2SAT and the multi-objective training approach by Modified Niche Genetic Algorithm (MNGA). Primarily, the phases in *r*2SATMRA are the preliminary, logic, training, and retrieval phases as displayed in Figure 1 below. Then, the following subsections will discuss the explanations for each phase.

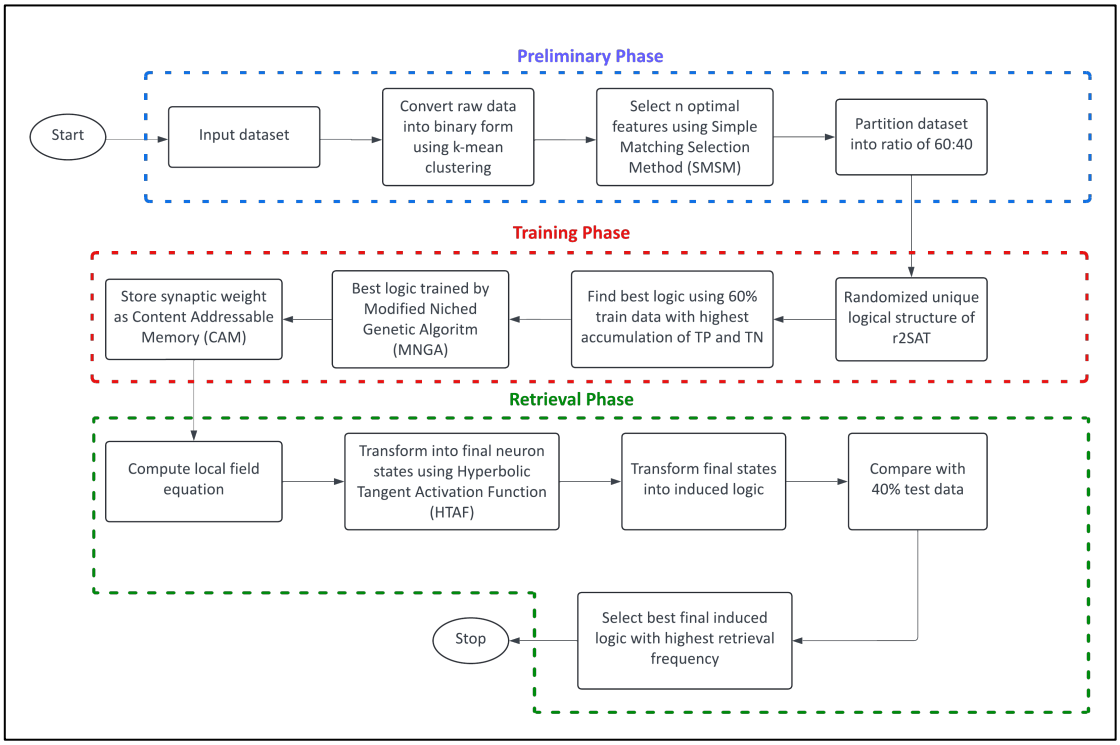


Fig. 1. Brief implementation and process of four phases involved in *r*2SATMRA

*2.3.1 Preliminary phase in r*2SATMRA

Before applying the Absenteeism at Work dataset to logic mining, a preliminary phase was carried out to enhance data quality and ensure the suitability of the dataset for the specific classification task. The data cleaning process of the dataset is the focus of this phase. This phase consists of three steps which are data preparation, the proposed SMSM, and data portioning. Therefore, the outline of each step will be explained as follows.

Step of data preparation: The raw entries (e_i) are transformed into binary form via the *k*-means clustering method as supported by Rusdi *et al.,* [5]. This method is important to represent the values in the neurons analyzed by DHNN.
Step of SMSM: SMSM is used as a feature (E_i) selection method to eliminate unimportant features that have less similarity with the class of the dataset. The formulation of Simple Matching (SM) expressed in Eq. (10) as follows

$$SM = \frac{\sum_{i=1}^{n} \vartheta_{a_i} + \sum_{i=1}^{n} \vartheta_{d_i}}{\sum_{i=1}^{n} \vartheta_{a_i} + \sum_{i=1}^{n} \vartheta_{b_i} \sum_{i=1}^{n} \vartheta_{c_i}}, \tag{10}$$

where *n* represents the total entries in the dataset.

According to Yu *et al.,* [20], many feature selection methods neglect the distribution differences of entries with the classes. Thus, the SMSM is introduced to distinguish significant features using the SM formula. Notably, the distribution of entries that leads to the positive class of the dataset will give high values of SMSM which are preferable. Figure 2 below depicts the values of SMSM for all features in simulation data of absenteeism at work.

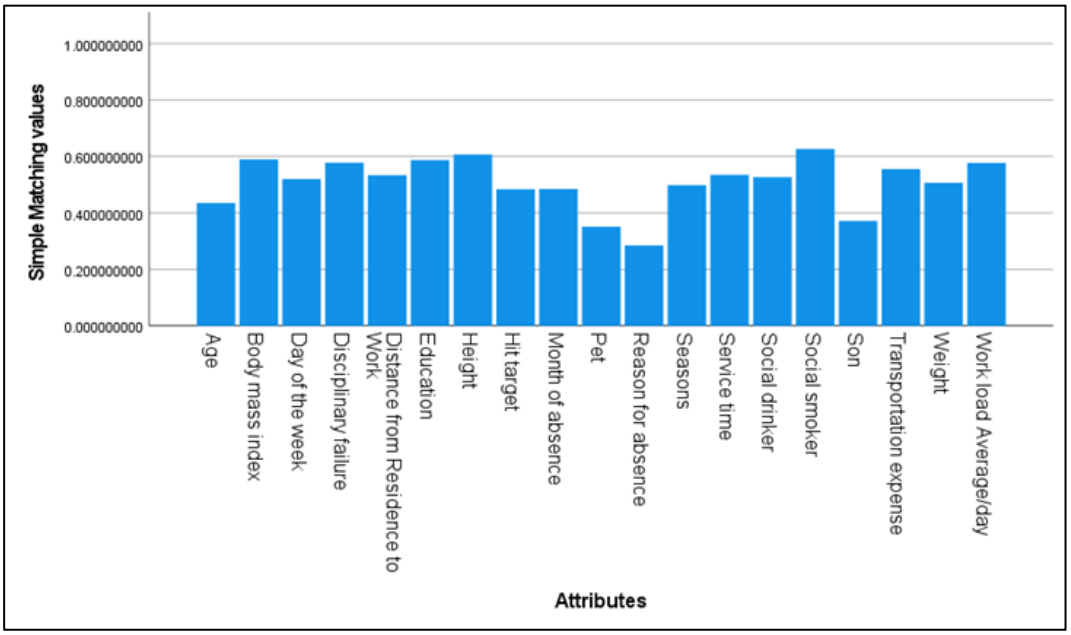


Fig. 2. Values of SMSM for all features in absenteeism at work

Step of data portioning: The dataset was divided for training and retrieval with the ratio of $60:40$ represented as $\Gamma_{\text{train}} : \Gamma_{\text{test}} = 60:40$ [21]. This ratio is used in good agreement with most of the existing logic mining models. Furthermore, *K*-fold cross validation was applied in the proposed *r*2SATMRA. Hence, the result for all metrics is derived from the average of all folds.

*2.3.2 Logic phase in r*2SATMRA

Logic phase is the phase after the preliminary phase of *r*2SATMRA. During this phase, the logical Γ_{r2SAT} is generated with different distributions of *r* where $r=[0.1,0.9]$ with $\Delta r=0.1$. The values of $\lambda_{r2\mathrm{SAT}}$ is always equal to the combination of selected E_i and the class of the dataset. Notably, GA effectively optimized negated variables distribution in *r*2SATMRA. It is crucial to examine every possible structure of *r*2SAT to ensure that the proposed *r*2SATMRA does not overlook any potential of neuron connections, especially when the negative linkages are positioned to another variable. Worth mentioning that, unique Γ_{r2SAT} is generated to avoid redundant logical conditions. Note that, each structure of *r*2SAT for any *r* will undergo a different training phase of DHNN-*r*2SAT.

*2.3.3 Training Phase in r*2SATMRA

In the training phase of *r*2SATMRA, each generated Γ_{r2SAT} will embed Γ_{train} and the values of true negative (*TN*) and true positive (*TP*) is calculated. A work by Rusdi *et al*., [5] highlights that the selection of super logic, $\Gamma_{r2\mathrm{SAT}}^{b}$ is according to the highest total of *TP* and *TN*. This indicates that the proposed *r*2SATMRA can effectively train a significant structure of Γ_{r2SAT} that capable of representing patterns of the Γ_{train}. Next, the multi-objective training approach is optimized using MNGA. The role of MNGA is to optimize the logic to ensure the optimal synaptic weight values have various diversity of CAM when $\Gamma_{r2\mathrm{SAT}}^{b}$ is generated. In order to generate $\Gamma_{r2\mathrm{SAT}}^{b}$, the logical outcome of Γ_{r2SAT} is compared with the actual outcome from Γ_{train}. Figure 3 below shows the possible classification of the confusion matrix in the training phase of *r*2SATMRA.

		Actual Outcome	
		1	-1
Logical Outcome	$\Gamma_{r2SAT}=1$	*TP*	False Positive (*FP*)
	$\Gamma_{r2SAT}=-1$	False Negative (*FN*)	*TN*

Fig. 3. Classification of comparison between $\Gamma_{r2\mathrm{SAT}}$ with Γ_{train}

Notably, the outcome of Γ_{r2SAT} is positive (1) when each e_i of the Γ_{train} obtained full consistency. Even so, the outcome of Γ_{r2SAT} is classified as negative (-1) if any e_i is not satisfied. Then, the calculation of the training outcome for each e_i is outlined in Eq. (11) as follows

$$\Gamma_{r2\mathrm{SAT}}=\begin{cases}1, & \text{satisfied}\\ -1, & \text{otherwise}\end{cases}. \tag{11}$$

Note that, the confusion matrix in Figure 3 is significant in *r*2SATMRA since the *q* number of $\Gamma_{r2\mathrm{SAT}}^{b}$ are selected when compared to Γ_{train}. Eq. (12) formulates $\Gamma_{r2\mathrm{SAT}}^{b}$ with the highest total of *TP* and *TN.*

$$\max\left[TP+TN\right]. \tag{12}$$

Each $\Gamma_{r2\text{SAT}}^{b}$ must achieve two objective functions which are being trained using MNGA as represented in Eq. (13) with conditions expressed in Eq. (14) below.

$$\max\left[f_L, f_D\right], \tag{13}$$

such that

$$\begin{aligned} &\max\left[f_L\right] = m \\ &d_1 + d_2 + d_3 + \ldots + d_i = tol_D \end{aligned}', \tag{14}$$

where tol_D is assigned a value equal to *u* (100% diversity tolerance).

*2.3.4 Retrieval phase in r*2SATMRA

Retrieval phase is the final phase involved in *r*2SATMRA. The final states are generated using the local field formulated in Eq. (8). The recalled final states are changed into the logical SAT based on the pioneered RA method proposed by Sathasivam and Abdullah [13]. All the procedures involved in the retrieval phase of *r*2SATMRA are described in this subsection. The synaptic weight values in each Γ_d-CAM will be utilized in Eq. (8) and all possible induced logic $\left(\Gamma_{r2\text{SAT}}^{I}\right)$ correspond to the independent number of trials (*b*) is generated. Figure 4 below presents the confusion matrix for the possible outcome of comparing $\Gamma_{r2\text{SAT}}^{I}$ with the Γ_{test}.

		$\Gamma_{r2\text{SAT}}^{I}$ from Γ_{test}	
		1	-1
$\Gamma_{r2SAT}^{I}\left[\Gamma_{test}\right]$	$\Gamma_{r2SAT}=1$	*TP*	*FP*
	$\Gamma_{r2SAT}=-1$	*FN*	*TN*

Fig. 4. Classification of comparison between Γ_d with $\Gamma_{r2\text{SAT}}^{I}$

The best Γ_{r2SAT}^{I} will be determined from the entries according to the highest frequency obtained from the entries of Γ_{test}. The accuracy, *ACC* offers a broad view of the *r*2SATMRA process in producing output with high *TP* and *TN* which depicts the actual outcome of the dataset [22]. Noteworthy, the performance of *r*2SATMRA is further measured with other performance metrics approaches such as specificity (*SPE*), Matthew's Correlation Coefficient (*MCC*), and F-score (*F1*). The pseudocode of the proposed *r*2SATMRA that summarized all the phases is shown in Algorithm 1 below.

Algorithm 1: The pseudocode of the proposed *r*2SATMRA

1 Input dataset
2 Convert dataset into binary entries using k-means clustering method
3 Determine decision class of dataset, Γ_d
4 Select optimal features using SMSM
5 Split the dataset into Γ_{train} and Γ_{test}
6 For each $\Gamma_{r2\text{SAT}}$ in Γ_{train}

7 Generate $\Gamma_{r2\mathrm{SAT}}$ via GA
8 Optimize the produced $\Gamma_{r2\mathrm{SAT}}$ using binary optimization algorithm
9 Produce super logic, $\Gamma^{b}_{r2\mathrm{SAT}}$
10 Train super logic, $\Gamma^{b}_{r2\mathrm{SAT}}$ using MNGA in Eq. (13)
11 Store produced best logic units as CAM, Γ_{d}-CAM
12 Compute synaptic weights using local field equation in DHNN
14 Squash local field into binary neuron states using HTAF
15 Select final states using Eq. (8)
16 Transform final neuron states into induced logic, $\Gamma^{I}_{r2\mathrm{SAT}}$
17 Select best induced logic based on Eq. (12)
18 Output best induced logic

3. Experimental Setup

The implementation of the experimental setup that was carried out throughout this paper is discussed in this section including the data type, the device settings, the parameter settings, and the performance metrics formulation to assess the performance of *r*2SATMRA on the absenteeism at work dataset. In particular, the neurons in the DHNN model were assigned to binary form, $S_i \in \{-1,1\}$ for better compatibility compared to binary values. The experiment was conducted on a personal computer using Dev C++ (Version 4.9.2) with an Intel Core i5 processor to guarantee consistent settings and processing power. The simulation followed a supervised training approach with each algorithm iteration limited by a predefined maximum to fully evaluate MNGA. Table 1 lists the parameters settings *r*2SATMRA.

Table 1
Parameter settings involved in *r*2SATMRA

Parameter/Notation	Value/Remark
Range of *r*	[0.1, 0.9] [12]
Logic phase algorithm	GA [17]
Feature Selection	SMSM
Training threshold	100 [23]
Number of *p*-CAM	5 [24]
K-fold cross validation	5
Number of trials	100 [25]
$\Gamma_{train} : \Gamma_{test}$	60 : 40 [5]
Logical condition	Γ_{r2SAT} [12]
Selection super logic	Max [*TP* + *TN*] [5]

The absenteeism at work dataset is selected based on three criteria. First, the dataset should have at least 11 features or variables to accommodate the ten variables in Γ_{r2SAT} and ensure the inclusion of at least one negated variable within the logic. Second, a minimum of 100 entries is required to guarantee the statistical properties. The training capability of the logic mining model is poorly evaluated when analysing datasets with small entries. Third, e_i of the dataset will operate discretely using binary values (-1 and 1). As supported by Li *et al.,* [26], the dataset with real numbers, integers, and categorical feature types is preferable because it is easy to handle and prevents information loss. After that, the raw entries will be transformed into the binary form of 1 and -1 by *k*-mean clustering method which is displayed in Table 2 below. As mentioned above, this is because DHNN only compatible with binary representation of entries. The simulation dataset comprised ten features

which are selected from the top ten highest of *SM* values: i) Transportation expense; ii) Distance from residence to work; iii) Service time; iv) Workload Average/day; v) Disciplinary Failure; vi) Education levels; vii) Social drinker; viii) Social smoker; ix) Height; x) Body Mass Index (BMI) and Absenteeism hours as dependent features. The binary value is assigned based on the mean value where a value more than and equal to the mean value is denoted as -1. Meanwhile, a value less than the mean value is referred to as 1.

Table 2
Features of the absenteeism at work dataset with the range for the binary representation

Feature Name	Label	Binary Representation	Range
Transportation expense	x_1	1	> 221.33
		-1	≤ 221.33
Distance from Residence to Work	y_1	1	> 29.63
		-1	≤ 29.63
Service Time	x_2	1	> 12.55
		-1	≤ 12.55
Work Load Average/day	y_2	1	>271.49
		-1	≤ 271.49
Disciplinary Failure	x_3	1	Yes
		-1	No
Education	y_3	1	Postgraduate,Doctorate
		-1	Highschool, Graduate
Social Drinker	x_4	1	Yes
		-1	No
Social Smoker	y_4	1	Yes
		-1	No
Height	z_1	1	>172.11
		-1	≤ 172.11
Body Mass Index (BMI)	z_2	1	>26.68
		-1	≤ 26.68

The confusion matrix-based metrics that are considered to measure the performance of *r*2SATMRA are formulated in Eq. (15) – Eq. (18) as follows [27-29]

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}, \tag{15}$$

$$SPE = \frac{TN}{TN + FP}, \tag{16}$$

$$F1 = \frac{2\left[\left(\frac{TP}{TP + FP}\right) \times \left(\frac{TP}{TP + FN}\right)\right]}{\left(\frac{TP}{TP + FP}\right) + \left(\frac{TP}{TP + FN}\right)}, \tag{17}$$

$$MCC = \frac{\left(TP \times TN\right) - \left(FP \times FN\right)}{\sqrt{\left(TP + FP\right)\left(TP + FN\right)\left(TN + FP\right)\left(TN + FN\right)}}. \tag{18}$$

4. Results and Discussion

This section discusses the performance of the proposed *r*2SATMRA in doing absenteeism at work dataset when the feature selection method is applied with new computation of super logic. The first section explains the analysis of *r*2SATMRA in predicting the dataset. Meanwhile, the following section displays the performance of the traditional classification method in doing absenteeism at work dataset.

4.1 r2SATMRA-based Analysis in Predicting Absenteeism at Work

The performance of DHNN-*r*2SAT model will be discussed in this study. The proposed model DHNN-*r*2SAT with *r*2SATMRA will be used to extract induced logic across 5 folds and the absenteeism hours was chosen as an aim. Notably, the outcome of induced logic in *r*2SATMRA will be provided in the binary form where -1 signifies the absenteeism hours are below the mean value while 1 indicates the absenteeism hours are above the mean value. Four metrics have been analysed to evaluate the performance of the proposed *r*2SATMRA and the values for all performance metrics were represented in Figure 5 below.

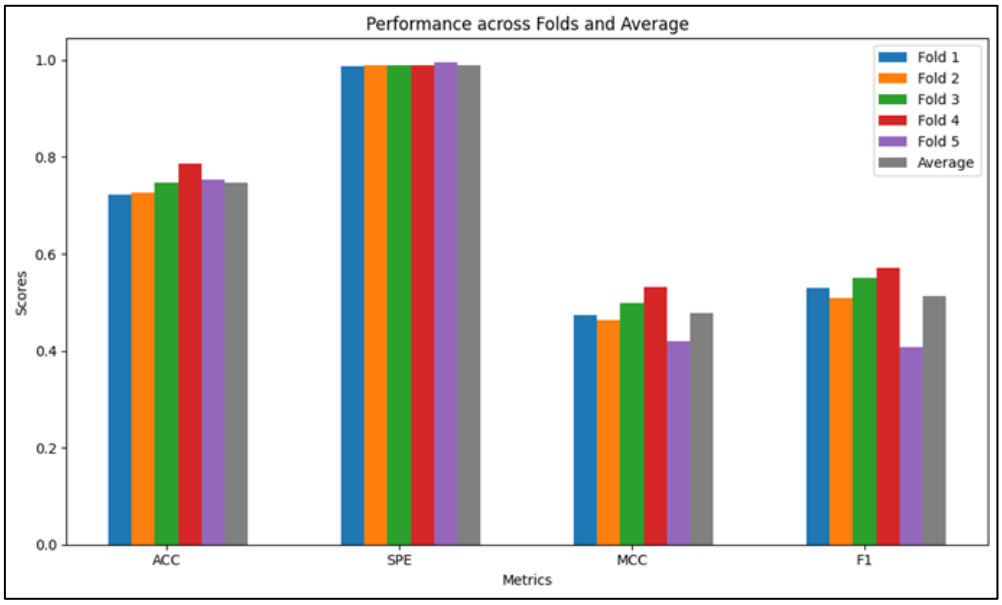


Fig. 5 Performance metrics across folds and average

The classification of the confusion matrix between Γ_d with $\Gamma_{r2\text{SAT}}^{I}$ for all 5 folds were displayed in Figure 6 below.

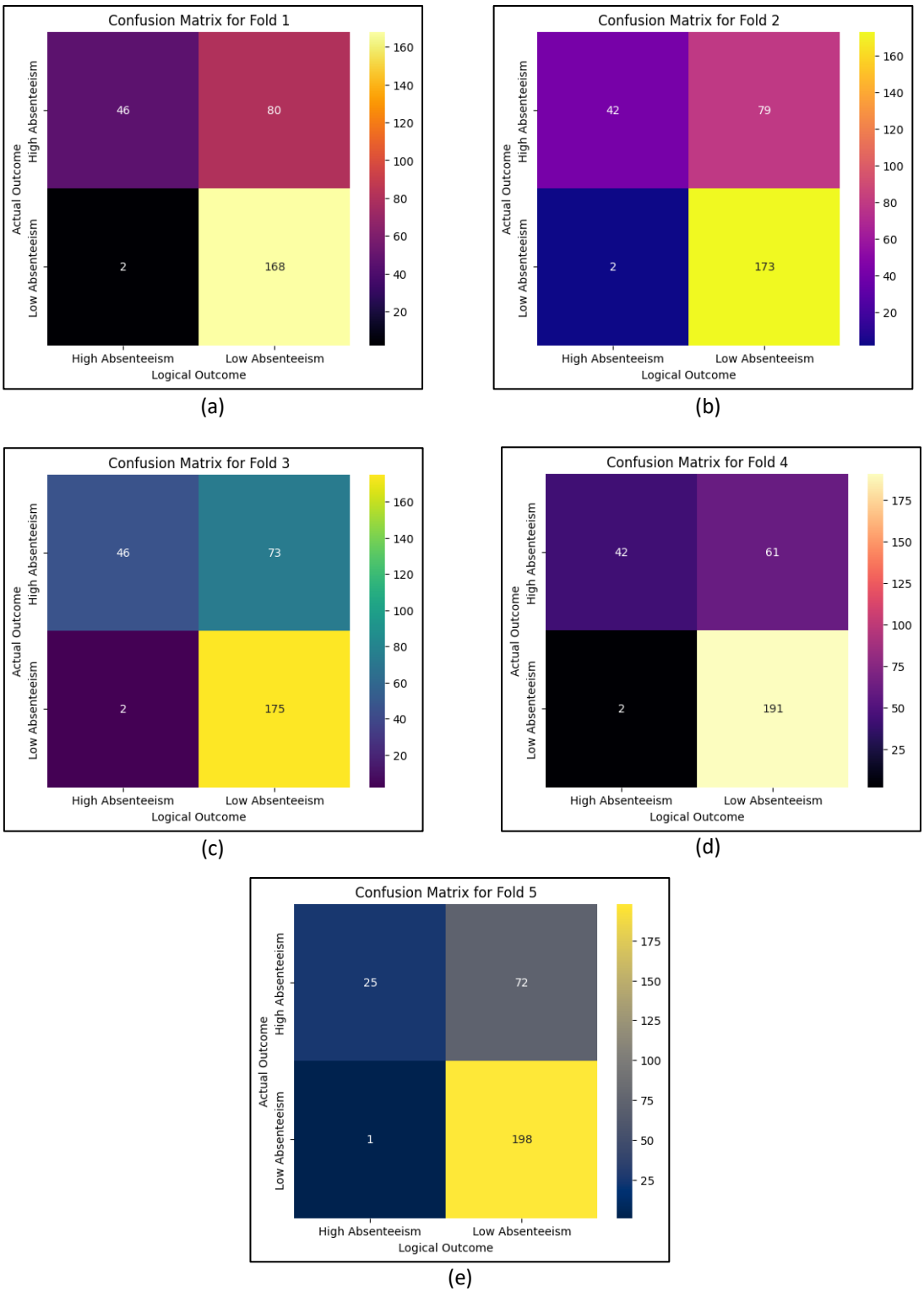


Fig. 6. Classification confusion matrix between Γ_d with $\Gamma^I_{r2\text{SAT}}$ (a) Confusion matrix for fold 1 (b) Confusion matrix for fold 2 (c) Confusion matrix for fold 3 (d) Confusion matrix for fold 4 (e) Confusion matrix for fold 5

The retrieved Γ^I_{r2SAT} for all 5-fold cross validation are presented in Eq. (19) – Eq. (23) as follows

$$\Gamma^{I1}_{r2\text{SAT}} = (x_1 \vee y_1) \wedge (x_2 \vee \neg y_2) \wedge (\neg x_3 \vee y_3) \wedge (x_4 \vee \neg y_4) \wedge z_1 \wedge \neg z_2, \tag{19}$$

$$\Gamma_{r2\mathrm{SAT}}^{I2}=\left(x_1\vee y_1\right)\wedge\left(x_2\vee\neg y_2\right)\wedge\left(\neg x_3\vee y_3\right)\wedge\left(x_4\vee\neg y_4\right)\wedge z_1\wedge\neg z_2, \tag{20}$$

$$\Gamma_{r2\mathrm{SAT}}^{I3}=\left(x_1\vee y_1\right)\wedge\left(x_2\vee y_2\right)\wedge\left(\neg x_3\vee y_3\right)\wedge\left(x_4\vee\neg y_4\right)\wedge z_1\wedge\neg z_2, \tag{21}$$

$$\Gamma_{r2\mathrm{SAT}}^{I4}=\left(x_1\vee y_1\right)\wedge\left(x_2\vee\neg y_2\right)\wedge\left(\neg x_3\vee y_3\right)\wedge\left(x_4\vee\neg y_4\right)\wedge z_1\wedge\neg z_2, \tag{22}$$

$$\Gamma_{r2\mathrm{SAT}}^{I5}=\left(x_1\vee\neg y_1\right)\wedge\left(\neg x_2\vee y_2\right)\wedge\left(\neg x_3\vee y_3\right)\wedge\left(x_4\vee\neg y_4\right)\wedge z_1\wedge z_2, \tag{23}$$

whereby the highest retrieval frequency of $\Gamma_{r2\mathrm{SAT}}^{I}$ can be seen through the formulation in Eq. (19). Thus, the induced logic from fold 1, $\Gamma_{r2\mathrm{SAT}}^{I1}$ is selected as the best final induced logic, $\Gamma_{\mathrm{best}}^{I}$. Based on Figure 5, the average value of *SPE* is the highest. This suggests that *SPE* results are more crucial in order to investigate features that contributed to the lower absenteeism hours. The high value of *SPE* indicates that the proposed *r*2SATMRA are able to accurately predict high value of *TN*. Hence, the entries that would make $\Gamma_{\mathrm{best}}^{I}$ unsatisfied or negative will provide more significant and insightful information for employers. As mentioned above, Eq. (19) is selected as the representative of $\Gamma_{\mathrm{best}}^{I}$. The possible entries that would produce negative outcome for $\Gamma_{\mathrm{best}}^{I}$ is $\left(-1,-1,-1,1,1,-1,1,1,1,1\right)$. The possible deductions of negative outcome (low absenteeism outcome) will be presented in Table 3 below.

Table 3
Deductions and recommendations for reducing absenteeism based on dataset analysis

Deduction	Explanation	Recommendations for Employers	Supporting Evidence
Impact of Transportation Expense $\left(x_1\right)$ and Commute Distance on Absenteeism $\left(y_1\right)$	Lower absenteeism hours (-1) when transportation expenses are under $221.33 USD/month $\left(x_1=-1\right)$and commute distance is less than 29.63 km $\left(y_1=-1\right)$. Reduced financial burdens and physical fatigue from commuting encourage better attendance.	a) Provide transportation subsidies or shuttle services. b) Offer remote or flexible work options for employees with long commutes.	A study by Shifrin *et al.,* [30] shows that a meta-analytic review involving 33 studies and over 90,000 participants highlight the positive influence of flexible work arrangements on employee health and absenteeism rates.
Service Time $\left(x_2\right)$ and Daily Workload $\left(y_2\right)$ Influence on Absenteeism	Absenteeism hours may be lower (-1) for employees with less than 12 months of service $\left(x_2=-1\right)$ and light workloads $\left(\neg y_2=-1\right)$. Newer employees are less stressed, more motivated, and eager to make a good impression.	a) Recognize and provide growth opportunities for long-tenured employees. b) Ensure workloads are manageable, provide flexibility, and offer resources to reduce job strain.	Mistry [31]: Growth opportunities improve motivation and reduce turnover. Bhattaru *et al.,* [32]: Balanced workloads are critical for preventing absenteeism caused by stress and heavy workloads.
Disciplinary Record $\left(x_3\right)$ and Education	Employees with no disciplinary issues $\left(\neg x_3=-1\right)$ and higher education $\left(y_3=1\right)$	a) Provide targeted training and mentoring programs.	The paper by BMC Health Services Research [33] states that the employees who receive clear guidance and

Level (y_3) on Absenteeism	tend to have lower absenteeism rates. They are more responsible and	b) Offer regular performance feedback to improve engagement and correct behavior constructively.	support through these programs are more likely to remain engaged and present at work, feeling more integrated into their roles and the organization as a whole.

This study implemented *r*2SATMRA to extract knowledge from the absenteeism at work dataset. The extracted knowledge or information will be used to identify relevant factors that contribute to absenteeism in the workplace. With *r*2SATMRA, the optimal induced logical condition is able to be obtained in the execution of this experiment and successfully describes the factors of absenteeism based on the dataset. Therefore, the companies can use these findings to explain to employers in order to improve their performance.

4.2 Performance of SVM-based Analysis in Predicting Absenteeism at Work

The absenteeism patterns identification is not limited to any Artificial Intelligence models but also can be applied to any relevant machine learning models. One of the models that can be used is Support Vector Machine (SVM) [34]. In this subsection, the performance of SVM to predict patterns in the absenteeism dataset as compared to the *r*2SATMRA is displayed in Figure 7 below.

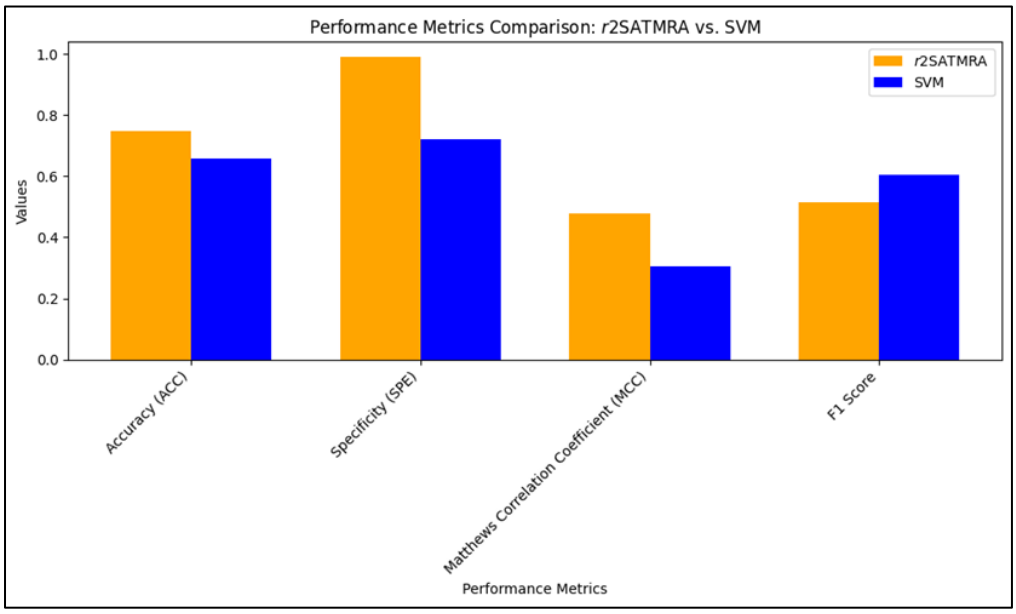


Fig. 7. The performance metrics comparison for *r*2SATMRA and SVM

In comparing the two models, *r*2SATMRA generally outperforms SVM in terms of *ACC*, *SPE,* and *MCC* showing its capability in producing more reliable and consistent classifications across different folds. The high specificity of *r*2SATMRA indicates its advantage in identifying employees with low absenteeism which may be useful for human resources. However, SVM has a slight edge in the F1 Score suggesting that SVM is better suited for cases where a balance between precision and recall in high absenteeism is required. This makes SVM useful for employers to identify at-risk employees for targeted interventions.

5. Conclusion

In summary, the three main objectives were accomplished throughout this paper. First, *r*2SAT logic successfully be formulated as neuron representation in DHNN. A weighted feature namely ratio of negated variables, *r* is introduced in the proposed logical condition to govern the number of negated variables. The proposed *r*2SAT will govern the neurons in the DHNN. The *r*2SATMRA has been successfully proposed in doing classification tasks for the absenteeism at work dataset. In order to assure optimal training phase, MNGA is added in the proposed *r*2SATMRA. The reverse analysis method is applied in the proposed *r*2SATMRA to retrieve the classification output in the form of the induced logic. Next, the performance of *r*2SATMRA has been accessed by considering the evaluation *TP*, *TN*, *FP*, and *FN* based on the extracted final induced logic to the actual outcome of the absenteeism at work dataset. The best final induced logic is selected based on the highest retrieval frequency from all *K*-folds cross validation. It is recommended for employees to make work balanced and manageable, clear expectations and resources provided to complete the task to reduce absenteeism hours among employees.

Despite the promising results in the proposed *r*2SATMRA, there are several limitations such as we only consider first and second-order logic in the proposed *r*2SAT logic. First and second-order logic have shown a good performance in representing the pattern of the dataset. In addition, the weighted feature introduced has play a significant role in improving the interpretation of SAT as logical rule in DHNN from the perspective of negative variables. Then, the binary representation of raw data has help in giving exact solution to either True or false since it is compatible with DHNN. For the future recommendation, this work can further be improved using different training metaheuristics algorithm such as Artificial Bee Colony and features selection method such as Jaccard Feature Selection Method. In addition, the proposed *r*2SATMRA can be applied to current issues such as analyzing TikTok engagement, advancements in STEM education, and ongoing impacts of COVID-19. Furthermore, the logic mining model ability to handle higher order logic and integrating alternative training algorithms suggest a wide scope of application in complex analytical scenarios.

Acknowledgement
This work was not funded by any grant.

References
[1] Deepika, V. "Applications of artificial intelligence techniques in polycystic ovarian syndrome diagnosis." *J. Adv. Res. Technol. Manag. Sci* 1, no. 3 (2019): 59-63.
[2] Morshidi, Azizan, Noor Syakirah Zakaria, Mohammad Ikhram Mohammad Ridzuan, Rizal Zamani Idris, Azueryn Annatassia Dania Aqeela, and Mohamad Shaukhi Mohd Radzi. "Artificial Intelligence and Islam: A Bibiliometric-Thematic Analysis and Future Research Direction." *Semarak International Journal of Machine Learning* 1, no. 1 (2024): 41-58. https://doi.org/10.37934/sijml.1.1.4158
[3] Ong, Siew Har, Sai Xin Ni, and Ho Li Vern. "Dimensions Affecting Consumer Acceptance towards Artificial Intelligence (AI) Service in the Food and Beverage Industry in Klang Valley." *Semarak International Journal of Machine Learning* 1, no. 1 (2024): 20-30. https://doi.org/10.37934/sijml.1.1.2030
[4] Ye, Jianhong, Zhiyong Zhao, Ehsan Ghafourian, AmirReza Tajally, Hamzah Ali Alkhazaleh, and Sangkeum Lee. "Optimizing the topology of convolutional neural network (CNN) and artificial neural network (ANN) for brain tumor diagnosis (BTD) through MRIs." *Heliyon* 10, no. 16 (2024). https://doi.org/10.1016/j.heliyon.2024.e35083
[5] Kasihmuddin, Mohd Shareduwan Mohd, Nurul Atiqah Romli, Gaeithry Manoharam, and Mohd Asyraf Mansor. "Multi-unit Discrete Hopfield Neural Network for higher order supervised learning through logic mining: Optimal performance design and attribute selection." *Journal of King Saud University-Computer and Information Sciences* 35, no. 5 (2023): 101554. https://doi.org/10.1016/j.jksuci.2023.101554
[6] Abdullah, Wan Ahmad Tajuddin Wan. "Logic programming on a neural network." *International journal of intelligent systems* 7, no. 6 (1992): 513-519.

[7] Mansor, M. A., M. S. M. Kasihmuddin, and S. Sathasivam. "Artificial Immune System Paradigm in the Hopfield Network for 3-Satisfiability Problem." *Pertanika Journal of Science & Technology* 25, no. 4 (2017).
[8] Sathasivam, Saratha. "Upgrading logic programming in Hopfield network." *Sains Malaysiana* 39, no. 1 (2010): 115-118.
[9] Kasihmuddin, Mohd Shareduwan Mohd, Mohd Mansor, and Saratha Sathasivam. "Robust artificial bee colony in the hopfield network for 2-satisfiability problem." *Pertanika Journal of Science & Technology* 25, no. 2 (2017).
[10] Sathasivam, Saratha, Mohd Asyraf Mansor, Mohd Shareduwan Mohd Kasihmuddin, and Hamza Abubakar. "Election algorithm for random k satisfiability in the Hopfield neural network." *Processes* 8, no. 5 (2020): 568.
[11] Saribatur, Zeynep G., and Thomas Eiter. "Omission-based abstraction for answer set programs." *Theory and Practice of Logic Programming* 21, no. 2 (2021): 145-195. https://doi.org/10.1017/S1471068420000095
[12] Zamri, Nur Ezlin, Siti Aishah Azhar, Mohd Asyraf Mansor, Alyaa Alway, and Mohd Shareduwan Mohd Kasihmuddin. "Weighted random k satisfiability for k= 1, 2 (r2SAT) in discrete Hopfield neural network." *Applied Soft Computing* 126 (2022): 109312. https://doi.org/10.1016/j.asoc.2022.109312
[13] Sathasivam, Saratha, and Wan Ahmad Tajuddin Wan Abdullah. "Logic mining in neural network: reverse analysis method." *Computing* 91 (2011): 119-133.
[14] Kho, Liew Ching, Mohd Shareduwan Mohd Kasihmuddin, Mohd Asyraf Mansor, and Saratha Sathasivam. "Logic mining in football matches." *Indones. J. Electr. Eng. Comput. Sci* 17 (2020): 1074-1083. http://dx.doi.org/10.11591/ijeecs.v17.i2.pp1074-1083
[15] Jamaludin, Siti Zulaikha Mohd, Nurul Atiqah Romli, Mohd Shareduwan Mohd Kasihmuddin, Aslina Baharum, Mohd Asyraf Mansor, and Muhammad Fadhil Marsani. "Novel logic mining incorporating log linear approach." *Journal of King Saud University-Computer and Information Sciences* 34, no. 10 (2022): 9011-9027. https://doi.org/10.1016/j.jksuci.2022.08.026
[16] Jamaludin, Siti Zulaikha Mohd, Mohd Shareduwan Mohd Kasihmuddin, Ahmad Izani Md Ismail, Mohd Asyraf Mansor, and Md Faisal Md Basir. "Energy based logic mining analysis with hopfield neural network for recruitment evaluation." *Entropy* 23, no. 1 (2020): 40. https://doi.org/10.3390/e23010040
[17] Zamri, Nur Ezlin, Mohd Asyraf Mansor, Mohd Shareduwan Mohd Kasihmuddin, Siti Syatirah Sidik, Alyaa Alway, Nurul Atiqah Romli, Yueling Guo, and Siti Zulaikha Mohd Jamaludin. "A modified reverse-based analysis logic mining model with Weighted Random 2 Satisfiability logic in Discrete Hopfield Neural Network and multi-objective training of Modified Niched Genetic Algorithm." *Expert Systems with Applications* 240 (2024): 122307. https://doi.org/10.1016/j.eswa.2023.122307
[18] Delgado-Currín, Raúl R., Williams R. Calderón-Muñoz, and J. C. Elicer-Cortés. "Artificial Neural Network Model for Estimating the Pelton Turbine Shaft Power of a Micro-Hydropower Plant under Different Operating Conditions." *Energies* 17, no. 14 (2024): 3597. https://doi.org/10.3390/en17143597
[19] Sivaganesan, S., and E. Udayakumar. "An event-based neural network architecture with content addressable memory." *International Journal of Embedded and Real-Time Communication Systems (IJERTCS)* 11, no. 1 (2020): 23-40. https://doi.org/10.4018/ijertcs.2020010102
[20] Yu, Qiao, Shu-juan Jiang, Rong-cun Wang, and Hong-yang Wang. "A feature selection approach based on a similarity measure for software defect prediction." *Frontiers of Information Technology & Electronic Engineering* 18, no. 11 (2017): 1744-1753. https://doi.org/10.1631/
[21] Jha, Kanchan, and Sriparna Saha. "Incorporation of multimodal multiobjective optimization in designing a filter based feature selection technique." *Applied Soft Computing* 98 (2021): 106823,. https://doi.org/10.1016/j.asoc.2020.106823 106823
[22] Mahesh, T. R., V. Vinoth Kumar, V. Vivek, K. M. Karthick Raghunath, and G. Sindhu Madhuri. "Early predictive model for breast cancer classification using blended ensemble learning." *International Journal of System Assurance Engineering and Management* 15, no. 1 (2024): 188-197. https://doi.org/10.1007/s13198-022-01696-0
[23] Mohd Kasihmuddin, Mohd Shareduwan, Mohd Asyraf Mansor, Md Faisal Md Basir, and Saratha Sathasivam. "Discrete mutation Hopfield neural network in propositional satisfiability." *Mathematics* 7, no. 11 (2019): 1133. https://doi.org/10.3390/math7111133
[24] Karim, Syed Anayet, Mohd Shareduwan Mohd Kasihmuddin, Saratha Sathasivam, Mohd Asyraf Mansor, Siti Zulaikha Mohd Jamaludin, and Md Rabiol Amin. "A novel multi-objective hybrid election algorithm for higher-order random satisfiability in discrete hopfield neural network." *Mathematics* 10, no. 12 (2022): 1963. https://doi.org/10.3390/math10121963
[25] Gao, Yuan, Mohd Shareduwan Mohd Kasihmuddin, Ju Chen, Chengfeng Zheng, Nurul Atiqah Romli, Mohd Asyraf Mansor, and Nur Ezlin Zamri. "Binary ant colony optimization algorithm in learning random satisfiability logic for discrete hopfield neural network." *Applied Soft Computing* 166 (2024): 112192. https://doi.org/10.1016/j.asoc.2024.112192

[26] Li, Xiangjun, Zijie Wu, Zhibin Zhao, Feng Ding, and Daojing He. "A mixed data clustering algorithm with noise-filtered distribution centroid and iterative weight adjustment strategy." *Information Sciences* 577 (2021): 697-721. https://doi.org/10.1016/j.ins.2021.07.039

[27] Singh, Namrata, and Pradeep Singh. "A hybrid ensemble-filter wrapper feature selection approach for medical data classification." *Chemometrics and Intelligent Laboratory Systems* 217 (2021): 104396.,https://doi.org/10.1016/j.chemolab.2021.104396 104396

[28] Chicco, Davide, and Giuseppe Jurman. "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation." *BMC genomics* 21 (2020): 1-13. https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-019-6413-7

[29] Luque, Amalia, Alejandro Carrasco, Alejandro Martín, and Ana de Las Heras. "The impact of class imbalance in classification performance metrics based on the binary confusion matrix." *Pattern Recognition* 91 (2019): 216-231. https://doi.org/10.1016/j.patcog.2019.02.023

[30] Shifrin, Nicole V., and Jesse S. Michel. "Flexible work arrangements and employee health: A meta-analytic review." *Work & Stress* 36, no. 1 (2022): 60-85. https://doi.org/10.1080/02678373.2021.1936287

[31] Mistry, H. 2021. "A Study on Employee Motivation with Reference to Selected Companies in and Around Vapi." *Scholarly Research Journal for Interdisciplinary Studies* 8, no. 65: 1332. https://dx.doi.org/10.21922/srjis.v8i65.1332

[32] Bhattaru, Sarathsimha, SrinivasaRao Kokkonda, and Raju Challa. "From Burnout to Balance: A Sustainability-Oriented Survey on Job Stress and Work-Life Integration." In *MATEC Web of Conferences*, vol. 392, p. 01040. EDP Sciences, 2024. https://doi.org/10.1051/matecconf/202439201040.

[33] BMC Health Services Research. 2023. "Exploring the Impact of Mentoring Functions on Job Satisfaction and Organizational Commitment of New Staff Nurses." *BMC Health Services Research*. https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-023-09957-5.

[34] Marchetti, Mara, Lorenzo Fongaro, Antonio Bulgheroni, Maria Wallenius, and Klaus Mayer. "Classification of uranium ore concentrates applying support vector machine to spectrophotometric and textural features." *Applied Geochemistry* 146 (2022): 105443. https://doi.org/10.1016/j.apgeochem.2022.105443.