S-TYPE RANDOM K SATISFIABILITY LOGIC IN DISCRETE HOPFIELD NEURAL NETWORK

SUAD ABDEEN HELALY OSMAN

UNIVERSITI SAINS MALAYSIA

2025

S-TYPE RANDOM K SATISFIABILITY LOGIC IN DISCRETE HOPFIELD NEURAL NETWORK

by

SUAD ABDEEN HELALY OSMAN

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

May 2025

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful.

I am grateful for Allah's guidance and strength, which allowed me to complete this doctoral journey successfully and gain a deeper understanding of this new knowledge. I would like to extend my sincere thanks to my supervisor, mentor, and role model, Assoc. Prof. Dr. Mohd Shareduwan. I also want to express my deep gratitude to my co-supervisors, Assoc. Prof. Dr. Mohd Asyraf Mansor and Dr. Muhammad Fadhil Marsani. A special thanks to Dr. Siti Zulaikha for her ongoing guidance and support.

My deepest appreciation goes to my Mother, my role model in life. Words cannot capture how thankful I am for her. She was the first to believe in me and supported me throughout the pursuit of my dream. My heartfelt prayers and gratitude also go to my Father, who left before I could share this accomplishment with him. I am also profoundly grateful to my siblings Saad, Mohamed, Manal, Maram, Hala, and Hilal for their support.

To my PhD journey friends, Nurul Atiqah Romli, Nur Ezlin Zamri, Gaeithry Manoharam, Nur Afifah Rusdi and Nurshazneem Roslan, who became like family to me during my time in Malaysia, my second home. They guided and supported me throughout my PhD journey.

I would like to express my profound gratitude to Assoc. Prof. Dr. Ahmad Lutfi Amri Ramli, Dean of the School of Mathematical Sciences, along with the academic and administrative staff at Universiti Sains Malaysia (USM) and the Institute of Postgraduate Studies (IPS). I am sincerely thankful for the opportunities, guidance, and invaluable experiences that have shaped my academic journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Θ_{2SAT}	2 Satisfiability logic
ς_i	Actual outcome attributes in logic mining
$\widehat{o}$	*AIC* maximum likelihood estimator
o	*AIC* model parameters
Avg R	Average rank
b_m^e	Coefficients value in *EN* regression equation for $y*$
b_{pi}	Coefficients value in *MICE* regression equation for $\widehat{P}_i$
b_{ij} ,	Coefficients value in *MICE* regression equation for $g(x_{p_i})$
$\wedge$	Conjunction (AND)
cos()	Cosine function
$E_{\Theta_{\delta 2SAT}}$	Cost function of the $\delta 2SAT$
b, $b*$	Counter
CR	Crossover Rate
t	Current iteration in HBSCA
t^c	DEA current iteration
h_{ij}^*	DEA mutant individual solution
$m*$	DEA population size
j^r	DEA random number
R	DEA scaling factor
x_{ij}^*	DEA target individual solution
u_{ij}	DEA trial individual solution
I_i	Dependent variable in F-test
$g(x_{p_i})$	Dependent variable in *MICE* regression equation
ρ_q	Desired negative proportion
ε	DHNN threshold

$\vee$	Disjunction (OR)
$d\left(x_{tj}, x_{bj}\right)$	Distance between x_{tj} and x_{bj}
L_i	EA candidate
$w_{V_i^j}$	EA eligibility distance coefficient
f_{L_j}	EA eligibility of the candidate
$f_{V_i^j}$	EA eligibility of the voter
$f_{\delta 2SAT}$	EA fitness function in $\delta 2SAT$
V_i^*	EA impacted voters
$S_{V_i^j}$	EA updating state of each voter
$S_{V_i^*}$	EA updating state of impacted voter
V_i	EA voters
L^e	*EN* function
τ_1^e	*EN* parameter
τ_2^e	*EN* parameter
E_i	*EN* significant attributes
e_{p_i}	Error term in *MICE* regression equation for $\widehat{P}_i$
e_{ij}	Error term in *MICE* regression equation for $g(x_{p_i})$
$\widehat{I}_i$	Estimator of dependent variable I_i in *F*-test
S_i^c	Final neuron state
$T_x^{(1)}$	First order logic
$W_i^{(1)}$	First order synaptic weight
f_D	Fitness for diversified solution
f_Z	Fitness for global solution
$\Theta_{\delta 2SAT}$	General formula of $\delta 2SAT$
R_{gd}	Global Diversity Ratio
R_g	Global Minima Ratio
A_i^*	HBSCA crossing aspirant

A_i^b	HBSCA highest grade aspirant
A_i	HBSCA initial aspirants
A_i^r	HBSCA random aspirants
$\tanh\left(h_i\right)$	Hyperbolic Tangent Activation Function
α_i	Intercept variable in *MICE* regression equations
KW-test	Kruskal-Wallis test
ϕ	Learning iteration in *DHNN-δ2SAT*
Θ	Logic on SAT
H_{Θ}	Lyapunov energy function in DHNN
U-test	Mann-Whitney U test
d	Margin of Error
x_{ij}	*MICE* independent attribute
P_i	*MICE* placeholder
$\hat{P}_i$	*MICE* placeholder estimation
$H_{\Theta_{\delta 2SAT}}^{\min}$	Minimum energy function in *DHNN-δ2SAT*
$H_{\Theta_{\delta 2SAT}}$	Minimum Energy value in *DHNN-δ2SAT*
x_{p_i}	Missing value imputed by *MICE*
F	Multi-objective function
$\neg$	Negation (NOT)
σ^n	Negative advertisement in EA
$\neg r$	Negative literal
ρ	Negative literal proportion
S_i	Neuron state
d_i^a	Nodes after optimized
d_i^b	Nodes before optimized
H_0	Null hypothesis
λ_2	Number clauses in *δ2SAT*

e	Number of $(d_i^b, d_i^a) = (-1, -1)$
f	Number of $(d_i^b, d_i^a) = (1, 1)$
g	Number of $(d_i^b, d_i^a) = (-1, 1)$
h	Number of $(d_i^b, d_i^a) = (1, -1)$
y_2	Number of clauses in $\delta 2SAT$ second order logic
x_2	Number of clauses in $\delta 2SAT$ the first order logic
π^*	Number of different pairs in Wilcoxon Rank Sum Test
ω	Number of DU-DHNN units
A_y	Number of elements in population
ξ_{2SAT}	Number of full negativities second order clauses
ge	Number of generations in HBSCA
G_S	Number of global minimum energy
gr	Number of groups Kruskal-Wallis test
$N_{V_i^*}$	Number of impacted voters by negative advertisement in EA
N_{S_j}	Number of influenced voters by positive advertisement in EA
υ	Number of learning in $DHNN\text{-}\delta 2SAT$
L	Number of learning in $\delta 2SAT$ probability logic phase
y_1	Number of literals in $\delta 2SAT$ second order logic
x_1	Number of literals in $\delta 2SAT$ the first -order logic
λ_1	Number of literals/ neurons in $\delta 2SAT$
L_S	Number of local minimum energy
β_1	Number of negative literals in first order logic
β_2	Number of negative literals in second order logic
η	Number of neuron combination
η^*	Number of non-repetitive $\delta 2SAT$ combination
H	Number of optimal induced logics
k^*	Number of parameters in the F-test model

N_{Party}	Number of parties in EA
Ω	Number of populations
N_{Pop}	Number of populations in EA
N_a	Number of populations in HBSCA
m	Number of Predictor variables in *EN*
p^e	Number of response variables in *EN*
λ_{2SAT}	Number of second order clauses in a specific string of logic
M_a	Number of second order logic optimized
φ	Number of trials in *DHNN-δ2SAT*
Δ	Number of variables shifting
$r_j^{\max\|u_l\|}$	Optimal attribute *l* selected in *δ2SATHRAα*
$r_j^{\max\|u_m\|}$	Optimal attribute *m* selected in *δ2SATHRAα*
$r_j^{\max\|u_n\|}$	Optimal attribute *n* selected in *δ2SATHRAα*
X_{best}	Optimal solution
d_i	Optimized nodes in HBSCA
$\Theta_{\varsigma_i}^{i}$	Outcome in NPLs induced logic
e_1	Parameter determines the position of movement direction
$\delta_\gamma 2SAT_\rho$	Parameter formula of *δ2SAT*
Υ	Parameter represent the probability of second order $p(y)$
γ_r	Pearson Correlation Coefficient
X_i^t	Position of the current solution in *i-th* dimension in SCA
P_i^t	Position of the destination point in *i-th* dimension in SCA
σ^P	Positive advertisement rate in EA
r	Positive literal
ϑ	Potential permutations
ρ_0	Pre-defined proportion range
x_m^e	Predictor variable in *EN* regression equation for *y**

P	Probability function
$p(x)$	Probability of first order logic
$p(y)$	Probability of second order logic
$\Theta_{RAN2SAT}$	Random 2 Satisfiability logic
rand()	Random number function
z	Random number generated in logic phase
e_2	Random number in interval [0,1]
e_3	Random numbers in interval [0, 2π]
e_4	Random numbers in interval [0,1]
R_G	Ratio of global minimum energy
ϖ	Ratio of Similarity
F^*	Represent the *FPR* in *AUC* formula
T^*	Represent the *TPR* in *AUC* formula
y^*	Response variable in *EN* regression equation for y^*
$\Theta_{S2SATRA}$	S2SATRA logic
n	Sample size
c	SCA constant number
T^c	SCA maximum number iteration
Np	SCA population size
X_i^{t+1}	SCA update agents
$T_y^{(2)}$	Second order logic in $\delta 2SAT$
$W_{ij}^{(2)}$	Second order synaptic weight
T	Set of clauses in $\delta 2SAT$
$\mathbb{N}$	Set of Natural number
τ_i	Set of variables in $\delta 2SAT$
a_i	Shapiro-Wilk test constants
SW	Shapiro-Wilk test value

Θ_{3SAT}^{best}	SHoRA best logic
Θ_{3SAT}^{learn}	SHoRA learning logic
Θ_{3SAT}	SHoRA logic by 3 Satisfiability
sig()	Sigmoidal transfer function
α	Significance Level
sin()	Sine function
$n*$	Size of dimension in DEA
n_i	Size sample in the *i-th* group in Kruskal-Wallis
W_S^*	Smallest of absolute values of the sum of x_i in Wilcoxon test
Sokal	Sokal And Michener Index
σ	Standard Deviation
KW	Statistic value of Kruskal-Wallis test
W	Statistic value of Wilcoxon test
U_1, U_2	Statistic values Mann-Whitney U test
R_i	Sum of ranks Kruskal-Wallis
R_1, R_2	Sum of the ranks in sample 1,2 in Mann-Whitney U test
τ_i^s	Summation of *TP* and *TN*
S_i^b	Superlative final neuron state
W_{ij}	Synaptic weight between *i* and *j*
W_{ii}	Synaptic weight of neuron *i*
$\bar{X}$	The arithmetic mean
$T_i^{(k)}$	The clause *i-th* with *k* variables in $\delta 2SAT$
$h_i(t)$	The local field in DHNN
$\tilde{X}$	The Median
u_l	The minimized *l* value obtained by Pearson correlation coefficient
u_m	The minimized *m* value obtained by Pearson correlation coefficient
u_n	The minimized *n* value obtained by Pearson correlation coefficient

Z	The upper $\alpha/2$ point of the normal distribution
$Tolr$	Tolerance value in *DHNN-δ*2*SAT*
S_d	Total number of diversified solutions
S_{gd}	Total number of global and diversified solutions
β_i	Total number of negative literals existing in *δ2SAT*
N_r	Total number of negative literals in *r*2SAT
N_{KW}	Total number of samples in Kruskal-Wallis
ξ	Total of negative literal in *δ*2*SAT*
$S_i(t)$	Update neuron state
x_i^n	Value in a completed data for an attribute with missing data
x_i	Value in the set of data
y_i	Value in the set of data
X^e	Vector of predictor variable in *EN* regression equation for *y**
Y^*	Vector of response variable in *EN* regression equation for *y**
*r*SAT	Weighted Random *k* Satisfiability
W-test	Wilcoxon Rank Sum test
$\delta_1 2SAT_\rho$	*δ*2*SAT* type 1
$\delta_2 2SAT_\rho$	*δ*2*SAT* type 2
$\Theta^b_{\delta 2SAT}$	*δ*2*SATHRAα* best logic
$\Theta^i_{\delta 2SAT}$	*δ*2*SATHRAα* induced logic
$S_i^{induced}$	*δ*2*SATHRAα* induced variable
Θ_{test}	*δ*2*SATHRAα* test logic
$\Theta^{top}_{\delta 2SAT}$	*δ*2*SATHRAα* top logic
Θ_{train}	*δ*2*SATHRAα* train logic

LIST OF ABBREVIATIONS

2SAT	2 Satisfiability
2SATRA	2 Satisfiability based Reverse Analysis
3SAT	3 Satisfiability
3SATRA	3 Satisfiability based RA Method
ACC	Accuracy
AIC	Akaike Information Criterion
AUC	Area Under the Receiver Operating Characteristic Curve
ABCJ	Artificial Bee Colony Algorithm by Jia *et al.* (2014)
ABCK	Artificial Bee Colony Algorithm by Kasihmuddin *et al*. (2017b)
ABCS	Artificial Bee Colony Algorithm by Sidik *et al*. (2022)
AI	Artificial Intelligence
ANN	Artificial Neural Network
BNM	Bank Negara Malaysia
BHA	Black Hole Algorithm
CSA	Clonal Selection Algorithm
CNF	Conjunctive Normal Form
CAM	Content Addressable Memory
CV	Cross-validation
DO	Decision Operator
DEA	Differential Evolution Algorithm
DHNN	Discrete Hopfield Neural Network
DNF	Disjunction Normal Form
DU-DHNN	Dynamic Unit Discrete Hopfield Neural Network
EN	Elastic Net regularization
EA	Election Algorithm
E2SATRA	Energy based 2 Satisfiability based Reverse Analysis Method
ES	Exhaustive Search
FN	False negative
FP	False positive
FPR	False Positive Rate
FNAE	Full Negativity Absolute Error

GPSAT	Generalized Probabilistic Satisfiability by Bona *et al*. (2015)
GenPSAT	Generalized Probabilistic Satisfiability by Caleiro *et al*. (2017)
GA	Genetic Algorithm
HNN	Hopfield Neural Network
HornSAT	Horn Satisfiability
HBSCA	Hybrid Binary Sine Cosine Algorithm
HHNNs	Hyperbolic Hopfield Neural Networks
HTAF	Hyperbolic Tangent Activation Function
*k*CNF	*k* Conjunctive Normal Form
*k*SAT	*k* Satisfiability
k-NN	*k*-Nearest Neighbors Imputation
LOOCV	Leave-one-out cross-validation
A2SATRA	Log-Linear in 2 Satisfiability based Reverse Analysis
L2SATRA	Log-Linear with 2 Satisfiability based Reverse Analysis
LCO	Loose Crossover Operator
MAJ2SAT	Major 2 Satisfiability
MCC	Matthews Correlation Coefficient
MAE	Mean Absolute Error
MICE	Multiple Imputation by Chained Equations
NPV	Negative Predictive Value
NPLs	Non-Performing Loans
P2SATRA	Permutation 2 Satisfiability Reverse Analysis
PRE	Precision
PSAT	Probabilistic Satisfiability Problem
PON	Probability of Negativity
RAN2SAT	Random 2 Satisfiability
RAN3SAT	Random 3 Satisfiability
RAN*k*SAT	Random *k* Satisfiability
RA	Reverse Analysis
RMSE	Root Mean Square Error
SAT	Satisfiability
DHNN-SAT	Satisfiability in Discrete Hopfield Neural Network
SCA	Sine Cosine Algorithm
SPE	Specificity

$\delta 2SAT$	S-Type Random 2 Satisfiability
$\delta 2SATHRA\alpha$	S-Type Random 2 Satisfiability based Reverse Analysis
$DHNN\text{-}\delta 2SAT$	S-Type Random 2 Satisfiability in Discrete Hopfield Neural Network
$\delta kSAT$	S-Type Random k Satisfiability
S2SATRA	Supervised 2 Satisfiability based Reverse Analysis Method
SHoRA	Supervised Higher Order Reverse Analysis Method
TEO	Trigonometric Enhanced Operator
TN	True negatives
TPR	True Positive Rate
TP	True positives
UCI	University of California Irvine
WA	Wan Abdullah Method
WOA	Whale Optimization Algorithm
YRAN2SAT	Y-Type Random 2 Satisfiability

LIST OF APPENDICES

K SATISFIABILITI RAWAK LOGIK JENIS-S DALAM RANGKAIAN DISKRET NEURAL HOPFIELD

ABSTRAK

Dalam perwakilan logik satisfiabiliti semasa dalam rangkaian neural buatan, kebolehpercayaan adalah penting untuk memastikan kebolehtafsiran model kotak hitam. Walau bagaimanapun, isu dalam logik sedia ada telah mengabaikan taburan kebarangkalian set data, yang menyebabkan ketidakpastian. Dalam tesis ini, peraturan logik satisfiabiliti baharu yang dinamakan 2 Satisfiabiliti Rawak Jenis-S dicadangkan dengan taburan kebarangkalian bagi menentukan bilangan klausa peringkat logik dan literal negatif. Logik yang dicadangkan mempunyai fasa logik kebarangkalian yang mengagihkan literal negatif dan memastikan penjanaan struktur yang optimum, yang mencapai klausa terendah dan ralat negatif 0.41. Seterusnya, Algoritma Hibrid Kosinus Sinus Perduaan dengan fungsi multi objektif dicadangkan di dalam fasa dapatan semula untuk mendapatkan keadaan neuron akhir berkualiti tinggi. Algoritma yang dicadangkan memaparkan indeks keserupaan terendah iaitu 0.02 berbanding algoritma sedia ada. Fasa prapemprosesan berselia baharu telah dicadangkan untuk membantu dalam menyediakan set data dengan kaedah sedia ada yang berbeza dan menunjukkan prestasi yang memberansangkan. Akhir sekali, model perlombongan logik baharu telah dicadangkan iaitu analisis berbalik berasaskan 2 Satisfiabiliti Rawak Jenis-S dengan kaedah Rangkaian Diskret Neural Hopfield Unit-Dinamik Baharu dan berkesan meliputi keseluruhan ruang penyelesaian untuk mengekstrak semua optimum logik teraruh berbanding kaedah sedia ada. Model perlombongan logik yang dicadangkan telah digunakan untuk menganalisis set data pinjaman tidak berbayar dan mempamerkan prestasi cemerlang yang mencapai ketepatan 100%.

S-TYPE RANDOM K SATISFIABILITY LOGIC IN DISCRETE HOPFIELD NEURAL NETWORK

ABSTRACT

In the current satisfiability logic representation in artificial neural networks, reliability is crucial for the interpretability of black box models. However, the issue in existing logic is the neglect the probability distribution of the dataset that caused uncertainty. In this thesis, a new satisfiability logical rule namely S-Type Random 2 Satisfiability is proposed with probability distribution to determine the number of order logic clauses and negative literals. The proposed logic includes a probability logic phase that distributes negative literals and ensures the generation of an optimal structure, which achieved the lowest clauses and negative error of 0.41. Subsequently, the Hybrid Binary Sine Cosine Algorithm with a multi-objective function is proposed in the retrieval phase to obtain a high-quality final neuron state. The proposed algorithm achieved the lowest similarity index of 0.02 when compared to existing algorithms. The new supervised preprocessing phase was proposed to assist in preparing the datasets with different existing methods and demonstrate superior performance. Finally, a new logic mining model was proposed, named S-Type Random 2 Satisfiability based reverse analysis with the New Dynamic Unit Discrete Hopfield Neural Network method and effectively covers the entire solution space to extract all optimal induced logic as compared to existing models. The proposed logic mining model was applied to analyze the non-performing loans dataset and exhibited outstanding performance which achieves an accuracy of 100%.

CHAPTER 1

INTRODUCTION

1.1 Overview

Researchers across various scientific fields have aimed to develop machines or autonomous systems that demonstrate high levels of intelligence and certainty. Such systems enhance decision-making and the ability to analyze vast amounts of data. Consequently, these systems are valuable in numerous real-world applications for classification and prediction tasks. In this chapter, the introduction to Artificial Intelligence is briefly explained, followed by an overview of Artificial Neural Networks, which addresses the main question of this thesis. The discussion then focuses on the rationale for using a Discrete Hopfield Neural Network to manage uncertainty, highlighting the problem statements associated with developing a reliable intelligent system. This is followed by an outline of the research objectives and questions, supported by methodologies and an exploration of the research scope.

1.2 The Fundamental of Artificial Neural Network

Artificial intelligence (AI) is a broad field focused on creating intelligent machines. AI research explores how the human brain thinks, learns, decides, and solves problems. It aims to address complex challenges in a more human-like manner but with far greater speed than a human could achieve. In the field of AI, Artificial Neural Network (ANN) models play an important role in simulating the functions of the human brain through mathematical computations (Chen *et al.*, 2021). ANN processes data in discrete units called "neurons", which exchange information through "connections" that mimic neural synapses. These neurons in the input layer handle the data scientists provide, while the output layer computes the estimated "answer" by the model. This architecture enables ANN to accurately represent and understand a wider range of information from various datasets, making it an integral part of AI systems. Due to its significance, flexibility, and efficiency, ANN is widely used to model complex systems across various sectors. These sectors include image processing (Shah *et al*., 2023), spatial information processing (Goel *et al*., 2023), education (Waheed *et al*., 2023), bank loans (Ahmed *et al*., 2023), and medicine (Ahmad *et al*., 2022). A properly trained ANN is considered an effective method for managing complex optimization tasks, as it can interpret data with minimal or no human supervision. However, to maximize its effectiveness, there is a need for ANN components allowing humans to understand the causes of errors and uncertainty better. Therefore, this research is driven by a core motivation:

Inspiring intelligent Artificial Neural Networks.

In this context, "inspire" refers to influencing or motivating new ideas that lead to advancements. To inspire someone means to encourage creativity. Here, inspiring intelligent ANN suggests introducing new components or developments within ANN. This drives ANN to evolve, improve, or innovate in functionality and problem-solving capabilities. Through inspiration, the network can produce optimally interpreted outcomes with high accuracy supporting the expansion of these models across various fields. Therefore, inspiring ANN by incorporating additional mechanisms or other AI techniques aims to improve knowledge representation and enhance ANN capabilities. This approach also introduces new features to advance AI research. In this context, researchers in mathematical fields have developed various techniques within ANN models. These techniques offer feasible solutions for problems, including optimization challenges. One of these significant and simple ANN is the Discrete Hopfield Neural Network (DHNN) (Wen *et al*., 2009). The DHNN attracts attention due to its simple structure, which consists of input and output layers with no hidden layers. The fundamental disadvantage of DHNN models is their uncertainty due to their "black box" nature, which is expected of all types of ANN (Juan *et al*., 2023). Consequently, the next section will discuss the current issue with the existing DHNN.

1.3 The Current challenges with Discrete Hopfield Neural Network

DHNN is a significant type of ANN formulated by Hopfield (1982) and further emphasized by Wen et al. (2009), utilizes a learning model based on associative features, and operates as a single-layer recurrent network with interconnected neurons, where each output neuron is feedback into every input. DHNN has several useful properties, including learning capability, fault tolerance, and the memory feature of Content Addressable Memory (CAM). Nevertheless, the main issue of DHNN is the uncertainty associated with the “black box” nature. A black box refers to a model that operates as a hidden system where the internal workings are not easily accessible or interpretable. Each new computation on the network results in a different weight matrix, which differs from the initial weights. This uncertainty indicates that the model is unable to determine whether parts are successfully or incorrectly optimized. As a result, any changes in neuron connections that contribute to the final output of the network cannot be tracked or recognized, which greatly limits their application, particularly in decision-making tasks. According to Gawlikowski *et al*. (2023), the uncertainty in ANN commonly arises from two sources: first, the data aleatoric (data uncertainty) due to insufficient or noisy measurements of variables, and second, epistemic (model uncertainty) arising from a lack of knowledge of the neural network model. Therefore, a successful approach to minimize the uncertainty of the DHNN model involves incorporating a neuro-modeling technique that utilizes a formal language based on symbolic logic. Abdullah (1992) proposed incorporating the concept of Satisfiability (SAT) into DHNN, where SAT is defined as a symbolic representation of neuron connections within the model. From this perspective, several researchers have introduced different SAT structures in DHNN by considering systematic and non-systematic SAT (Romli *et al*., 2024; Roslan *et al*., 2024). However, the logical rules presented experience uncertainty due to the random structure, which complicates solution interpretation. Therefore, incorporating a probabilistic SAT structure with an added neuron representation layer helps DHNN reduce black box uncertainty by enabling clearer mapping of neurons and better understanding of inputs and outputs. Another issue with the existing DHNN model is the poor quality of final neuron states retrieved by the network. This can be supported by the findings of Zamri *et al*. (2022a), which highlighted that the DHNN tends to produce repetitive final neuron states instead of generating diverse ones compatible with dataset needs. Therefore, enhancing the retrieval capabilities of the network is essential to improve the quality of the final neuron. The second type of uncertainty in DHNN concerns the input data, which is crucial for effective learning in describing the behaviour of the dataset by Reverse Analysis. However, the formulation of SAT does not consider data reliability, which can affect the quality of the output. Overlooking this aspect can result in poor dataset characteristics and reduce the interpretability of the data due to the increased uncertainty. Therefore, ensuring access to reliable datasets is essential for establishing an appropriate logical framework and enhancing knowledge extraction by DHNN. Another inefficiency in conventional DHNN lies in the limited search space, which is not adequate for the real-world datasets needed. This limitation negatively impacts the performance of the classification task model and can lead to uncertainty in classification. Therefore, enhancing the storage capabilities of DHNN is crucial for effectively storing dataset patterns. Considering all observations, this leads to a more comprehensive philosophical perspective in which:

"*The inspiration of the intelligence capabilities of the Discrete Hopfield Neural Network through probabilistic components to minimize uncertainty*".

Probability is a mathematical approach that can address uncertainty. In layman's terms, incorporating probabilistic elements into the DHNN helps minimize uncertainty, which makes it a more robust computational model. This probabilistic structure is crucial for reducing uncertainty, especially since no single model is ideally suited for all datasets. Probabilistic thinking leads to mitigating the uncertainty created by random SAT logic structures in DHNN by understanding the distribution of the dataset. Moreover, using a strategy for organizing SAT in DHNN makes more informed decision-making by focusing on various SAT structures that reflect a greater degree of certainty. Additionally, optimizing the DHNN model in the retrieval phase contributes to increasing the reliability of the model. Furthermore, adequately preparing for noisy datasets contributes to reducing uncertainty during logic mining classification. Expanding the search space of DHNN to adapt to the dataset appropriately will reduce uncertainty in the classification task and expand to include more applications. This thesis aims to improve the current DHNN by incorporating probabilistic perspectives. These improvements comprise a new systematic SAT structure, logic organization layer, optimizing the retrieval phase, preparing data, and implementing dynamic units for logic mining. The examination of each improvement aspect will be discussed in the next section of the problem statements.

1.4 Problem Statements

Managing uncertainty in DHNN classification models has been a key challenge due to the fundamental nature of the model and dataset. The logical rules, which consist of rigid and random structures do not align with the distribution of the dataset. Additionally, limited strategies for generating an ideal logic structure and an ineffective retrieval phase further complicate the process. Moreover, the lack of preprocessing and limited search space lead to inefficiencies and constraints in logic mining.

Problem Statement 1

Logical relations formulated as SAT can effectively bridge the gap between different data types and their associated uncertainties in ANN. This is achieved by providing a structured framework for representing complex relationships. Given the crucial role of SAT formulation, the SAT structures implemented in DHNN have been significantly modified to find optimal solutions for various problems. In this context, Abdullah (1992) introduced Horn Satisfiability (HornSAT) as the first SAT implemented in DHNN. However, HornSAT includes redundant variables. These redundant variables will lead to ineffective synaptic weight and result in zero effect on the retrieved final neuron state. This finding motivated the emergence of a new era of research with different perspectives. Kasihmuddin *et al*. (2017b) introduced the systematic 2 Satisfiability, and Mansor *et al*. (2017) proposed a high order of Satisfiability, namely 3 Satisfiability. Despite the successful implementation of systematic logic, the limitation in these logics lies in the restricted literals within the clauses. Subsequently, Sathasivam *et al*. (2020b) attempted to address this issue by introducing the Random 2 Satisfiability. This logic presents a dynamic logical framework that extends the range of synaptic weights. Unfortunately, the Random 2

Satisfiability faces uncertainty associated with the random selection of first and second order logic, which leads to poorly defined logical formulations. This affects the ability to understand the behavior of the dataset, which increases the uncertainty in DHNN. From a different perspective on SAT, Henderson *et al*. (2020) proposed the concept of Probabilistic Sentence Satisfiability (PSAT). This concept aims to generate an equation that expresses the relationship between variables and conditional probabilities. Despite this unique perspective aiming to incorporate probability into the SAT, this study lacked structural variation. In particular, this study failed to introduce diverse clauses and negative literals into the PSAT framework. As a result, the PSAT proves ineffective for addressing diverse real-life datasets because of the rigid structure where literals can be repetitive. Hence, it is essential to formulate a non-systematic probabilistic SAT with a probability dynamic distribution of the clauses and the negative literals. The proposed SAT must accurately choose the correct number of negative literals within the clauses based on the probability distribution of the data. In order to generate non-systematic logic correctly and distribute negative literals within clauses, an additional strategy is needed. This strategy enhances the structures of logical rules to represent the neuron connections in DHNN.

Problem Statement 2

Each structural component of SAT must be correctly represented before being encoded as symbolic rules in any computational system. An ineffective symbolic system can lead to overfitting, where the output converges to only one pattern of solution. Zamri *et al*. (2022a) introduced Weighted Random 2 Satisfiability with a proposed logic phase to produce non-systematic logic based on the ratio of negative literals. In this phase, the ratio distribution is assigned in the logic phase by employing a Genetic Algorithm to avoid bias and increase phase effectiveness. The findings showed that having a dynamic distribution of negative literals increases the generated of global minimum energy with different final neuron states. Although the logic phase successfully determines the distribution of the negative literals, the weighting scheme is limited by its dependency on a random selection system. This increases uncertainty in the DHNN model, making the contribution of different ratios to diverse final neuron states unclear. The neglect of the probability distribution of the dataset within the SAT structure contributes to model uncertainty. Consequently, the recently introduced non-systematic logic fails to address and manage uncertainty adequately. Roslan *et al*. (2024) introduced a fascinating perspective on the representation of negative literals in DHNN. The proposed non-systematic SAT named Conditional Random k Satisfiability is designed to reduce the influence of positive literals in the second order clause. This approach successfully enhanced neuron variation within DHNN solution spaces. Although the strategy has been effective in managing literal states, it still has limitations. One issue is that the distribution of the number of clauses and negative literals is chosen randomly due to the absence of a clear system for organization. Therefore, an organized layer or phase is needed to generate accurate SAT logical rules before implementation in any intelligent system. This layer mechanism should be compatible with calculating the probability from real datasets and assigning a number of clauses and negative literals. However, due to the feedback feature in DHNN, optimization made during the retrieval phase is often overlooked.

<u>Problem statement 3</u>

The retrieval phase in DHNN is responsible for generating final neuron states that achieve global minimum energy. Several existing works have been proposed to enhance the retrieval phase of DHNN in locating more global minimum energy states. However, this strategy did not consider the quality of the states. Consequently, there is a high chance of achieving global minima that are overfitted. For example, Kasihmuddin *et al*. (2019) proposed an Estimation of Distribution Algorithm to optimize the retrieval phase of the proposed Mutation Hopfield Neural Network. The proposed method showed the capability to explore more global minimum energy. However, the strategy still does not adequately address the diversification of the optimized final neuron state in terms of negativity. Sathasivam *et al*. (2020a) applied Mean Field Theory in the retrieval phase of DHNN. Mean Field Theory focuses on one network spin while replacing all other spins with an average local field utility to reduce neurons stuck in local minimum values. Although this technique increased the number of global minimum energy states, it did not consider diversity fitness or the similarity index ratio in the final neuron state. The proposed SAT in DHNN should incorporate an optimization algorithm with multi-objective function across the retrieval phase to improve final neuron states. Enhancing the retrieval phase can reduce model uncertainty. Key considerations include using multi-objective function to achieve global solutions, diversifying neuron states in terms of negativity, and minimizing the similarity index. However, overfitting or underfitting issues may still occur even after implementing the ideal network. External factors such as noisy datasets and unwanted variables make a robust approach necessary to preparing real-life datasets.

<u>Problem statement 4</u>

Data preprocessing to handle noisy datasets and unwanted variables is an essential aspect in reducing input uncertainty. This step also provides accurate data interpretation of logic mining in the DHNN. Kasihmuddin *et al*. (2022) proposed the Supervised 2 Satisfiability based Reverse Analysis method by using Pearson correlation as a features selection. This method is characterized as a simplistic and primitive approach. Despite the success of this strategy, this study does not provide a clear method to address missing data, which led to data uncertainty issue. Later, an attempt was made by Jamaludin *et al*. (2022) that implement log-linear analysis as a feature selection with 2 Satisfiability based Reverse Analyses logic. Log-linear considers all variables equally without distinguishing between dependent and independent variables. Thus, there is a chance of choosing the dependent variable as a significant variable. In this sense, the SAT in DHNN will learn the incorrect logical structure, which increases the chance of obtaining inaccurately induced logic. Thus, to improve the outcome of the logic mining model, an initial phase must be involved to clean and impute missing data to avoid overfitting. Additionally, feature selection techniques should be employed to reduce the impact of insignificant variables in the model. This approach can assist SAT in DHNN in producing optimal induced logic. However, after preparing the dataset and establishing all the components in the proposed model, increasing the search space becomes necessary. This helps reduce uncertainty in the classification task. The success of the proposed model relies on the performance achieved in the context of logic mining.

Problem statement 5

Unlike traditional ANN black box models, logic mining uses the ability of classification based on SAT in the DHNN model to reduce uncertainty, which enhances knowledge extraction. Sathasivam and Abdullah (2011) proposed the first logic mining, namely the Reverse Analysis incorporating HornSAT and DHNN. The proposed logic mining is used to extract logical rules from datasets, which effectively identifies and analyses patterns within the data. However, Reverse Analysis has several weaknesses. Primarily, this model has been proposed without incorporating data preprocessing steps or feature selection methods. Additionally, the model cannot generate a single induced logic potential for the datasets, which increases uncertainty. This weakness prevents the proposed logic mining approach from creating induced logic capable of performing classification tasks effectively. Using the systematic logic in DHNN, Kho *et al*. (2020) proposed a new technique based on the Reverse Analysis method for extracting useful information from datasets in the form of 2 Satisfiability. Later, Kasihmuddin *et al*. (2022) improved this work by proposing a new feature selection. Both works successfully extracted the best induced logic and addressed the problem of interpretability in data mining. However, these studies only focus on data entries associated with 1 and ignore data entries related to -1, which leads to the construction of incomplete classification models. Consequently, the induced logic retrieved becomes biased towards only true positive classifications. In a different perspective, Zamri *et al*. (2024) proposed a non-systematic logic mining model with specific units of CAM. This approach enhances the search space of DHNN and aids in locating an optimal induced logic. The results showed that the non-systematic model outperformed the systematic model by significantly improving the search space. Two main issues arise in the study. Firstly, the study considered a predefined or restricted number of CAMs. Secondly, the study only focuses on obtaining single induced logic. These issues may lead to increased uncertainty in the model since the model cannot capture all dataset patterns. The enhanced logic mining model should be capable of representing non-systematic SAT logical rules. Additionally, this model must effectively prepare the dataset. Moreover, the DHNN should adapt dynamically to expend the search space and store patterns in dynamic CAMs. These capabilities are needed to meet the requirements of the dataset and generate all potential induced logic, which contributes to diverse and reliable decision-making.

1.5 Research Questions

To ensure consistency between the problem statements provided in Section 1.4 and the research objectives of this thesis, this section presents significant research questions that outline various facets of the proposed study. These questions are around the following question,

"*How can the Discrete Hopfield Neural Network be inspired to reduce uncertainty and provide higher computational intelligence performance with dynamic search space?*"

Hence, the research questions relevant to this thesis have been listed as follows:

(a) What type of logical rule can be proposed to support probabilistic distribution of clauses and negative literals while enabling neuron representation in a Discrete Hopfield Neural Network?

(b) What logic generation approach can be proposed to probabilistically assign negative literals across clauses before training a Discrete Hopfield Neural Network?

(c) What kind of optimization mechanism can be proposed to enhance the retrieval phase by producing high-quality final neuron states in a Discrete Hopfield Neural Network?

(d) What preprocessing approach can be proposed to clean data, impute missing values, extract significant features, and partition datasets for logic mining classification tasks?

(e) What kind of logic mining model can be proposed to ensure the generation of all optimal induced logic that accurately describes dataset patterns and feature positions?

1.6 Research Objectives

This thesis aims to introduce non-systematic logic with probabilistic features that enhance the ability of DHNN to reduce uncertainty. A new phase will generate the correct probabilistic logic structure before entering the training phase of DHNN. In order to enhance the final neuron state, a population-based algorithm will be utilized in optimization. The proposed model will prepare real-life datasets and extract the empirical patterns for classification tasks. In this context, the proposed model must demonstrate higher computational intelligence performance with dynamic storage capabilities. Therefore, the objectives of this thesis are outlined as follows:

(a) To formulate a non-systematic logical rule, namely S-Type Random 2 Satisfiability, which utilized the probability distribution of the dataset. The probability distribution will be used to determine the number of first and second order logic. The Binomial population sample size will be used to assign the number of negative literals. The formulated S-Type Random 2 Satisfiability logical rule will be able to represent the neuron connection in the Discrete Hopfield Neural Network.

(b) To propose a new phase called the probability logic phase to generate correct logic and provide a dynamic distribution of negative literals in the logic structure. The new phase will distribute the negative literals assigned in first and second order logic based on the probability for each clause. The phase will ensure the S-Type Random 2 Satisfiability is correctly generated before being implemented in the Discrete Hopfield Neural Network.

(c) To propose a Hybrid Binary Sine Cosine Algorithm with multi-objective functions in the retrieval phase to optimize the final neuron state by achieving a global, diversified solution and reducing the similarity index. The proposed algorithm features a modified operator, namely Trigonometric Enhanced, to ensure a diverse neuron state. It also introduces two new additive operators, the Loose Crossover operator, to explore more diverse solutions and the Decision operator, to ensure retrieval of the optimal solution. The multi-objective functions are designed to ensure obtaining a high-quality final neuron state during the retrieval phase in the Discrete Hopfield Neural Network.

(d) To propose a new logic mining phase called the supervised preprocessing phase to prepare datasets used for classification tasks. This new phase aims to clean the dataset and impute missing values using the Multiple Imputation by Chained Equations to avoid biased estimates. Additionally, the Elastic Net regularization technique is applied to select the most important features, while the *k*-fold cross-validation technique is used for partitioning to minimize bias across the entire dataset.

(e) To construct a new logic mining model for classification tasks on various datasets, namely S-Type Random 2 Satisfiability based Reverse Analysis with the new Dynamic Unit Discrete Hopfield Neural Network method and newly shifting operator in the retrieval phase. The proposed model aims to cover the entire solution space completely by utilising dynamic network units. The dynamic mechanism enhances search space by creating dynamic units of CAM and producing all possible best and induced logic to represent every dataset pattern precisely. A proposed shifting operator will ensure all features are correctly positioned in each final induced logic. The final optimal induced logic represents the attributes that contributed to the outcome of the dataset.

1.7 Research Methodology and Scopes

Building a superior logic mining model with high accuracy in extracting information is the primary goal of this thesis. This model seeks to inspire the intelligence capabilities of DHNN through a probabilistic structure to reduce the element of uncertainty. The methodology adopted in this study is designed to address key challenges identified in the problem statement Hence, the methods selected in this thesis ensure that all objectives from Section 1.6 are fulfilled. This section introduces techniques and approaches for generating an advanced logic mining model for classification task. This includes a strategy for implementing the proposed logic in DHNN and a mechanism for selecting clauses and negative literal numbers. Additionally, the section outlines the algorithms with multi-objective function method during the retrieval phase. Finally, the approach for preprocessing datasets and a mechanism for adding new components to enhance logic mining are presented.

<u>Research Methodology (a)</u>

In recent studies of SAT, selecting an ideal SAT that aligns significantly with DHNN and accurately represents the dataset is essential. Therefore, Objective (a) formulates the S-Type Random 2 Satisfiability by utilizing a non-systematic SAT structure with first order and second order logic. The proposed S-Type Random 2 Satisfiability uses the probability distribution of the dataset. This structured approach is motivated by Sathasivam *et al*. (2020b), who proposed combining multiple logical orders within a single SAT formulation. First, the probability of distinctive attributes appearing in the dataset is calculated to control the composition probability of first and second order logic. Second, the Binomial population sample size will used to determine the appropriate number of negative literals. Notably, researchers tend to neglect negative literals due to indirectly mapped errors in a logical structure (Saribatur & Eiter, 2021). Furthermore, each clause will be connected with the Conjunctive Normal Form and literals by Disjunction Normal Form, which will contribute to the final formulation of the S-Type Random 2 Satisfiability. After formulating the proposed logic, each literal in the clause will represent the neuron and the logic structure will represent connection in DHNN.

Scopes of Methodology (a)

While this methodology can effectively be an optimal method to design a DHNN that is derived from the SAT logical rule. Certain restrictions must be considered to ensure the reproducibility of the objectives in this thesis. First, all clauses must contain neither repetitive nor redundant literals and must be represented in the Conjunctive Normal Form. The repetitive and redundant literals will affect the performance of the DHNN, which leads to a non-zero cost function. Second, the number of second order clauses must be more than the number of first order clauses. This is due to the first order gives one satisfied interpretation (Darmann & Döcker, 2021), which reduces the diversity of the final solution. Third-order logic is not included in the proposed model to avoid dependency issues within the neuron connection in DHNN. Third, the states of the neurons will be expressed using bipolar values of 1 and -1 that represent true and false (Kasihmuddin *et al*., 2022). Due to Lyapunov energy function dynamics is insufficient for representing binary information (1 and 0). When the states are zero, the network always has a minimum energy of zero, making it impossible to ascertain the exact minimum energy of the DHNN.

Research Methodology (b)

An effective method is necessary to generate the proposed S-Type Random 2 Satisfiability accurately and efficiently before training with DHNN. The probability logic phase first focuses on ensuring the correct formulation of the logic structure as based on the probability distribution. The phase ensures the probability distribution of the dataset adjusts with the increasing number of neurons. Second, the phase distributes the negative literals in each clause based on the probabilities assigned to each clause. The concept of an additional phase was inspired by Zamri *et al*. (2022a), who introduced a further phase to produce an accurate logical rule based on the desired proportion of negative literals. In this thesis, negative literals represent the distinctive attribute in the population from a binomial distribution. Two proposed performance metrics will be measured to ensure the DHNN trains the correct logical rule of S-Type Random 2 Satisfiability.

Scopes of Methodology (b)

Although this methodology can effectively create the S-Type Random 2 Satisfiability, certain restrictions are set to ensure the reproducibility of the second objective in this thesis. First, the proposed probability logic phase must produce S-Type Random *k* Satisfiability for $k = 1, 2$ according to the probability distribution of the dataset. The probability distribution determines the number of first and second order logic and negative literals within clauses. The two components mentioned will shape the behavior of the dataset in the DHNN. This thesis only considered the Binomial distribution population to exhibit the distinctive attributes due to the Binomial distribution, which considered variables with two outcomes (Probst *et al*., 2023). Second, the probability of clauses appearing is set within a specific range with a predefined step size to examine the efficacy of the number of clauses in the proposed logic. This approach considers the most significant differences range to avoid addressing repetitive logic structures. Third, to prevent the redundant distribution of the proportion for negative literals, the proportion is established within a specific range using a predefined step size. It is important to note that instead of manually adjusting the distribution of the literals in a SAT, it is better to use an organized technique to generate an accurate structure for the SAT (Zamri *et al*., 2022a).

Research Methodology (c)

To address the inefficiencies in the existing retrieval phase of DHNN, the thesis proposes a multi-objective function for the retrieval phase of DHNN to ensure that the final neuron states reach their optimal configurations. The Election Algorithm proposed by Sathasivam *et al*. (2020c) is incorporated into the training phase to aid in achieving the global minimum energy. Additionally, the enhanced population-based optimization algorithm, a Hybrid Binary Sine Cosine Algorithm with multi-objective function will be proposed during the retrieval phase. The proposed algorithm will ensure that the optimized final neuron states satisfy the requirements of the multi-objective function. Two additional operators will be implemented in this proposed algorithm. The first is the Loose Crossover Operator, which follows the Trigonometric Enhanced Operator (main operator). The second is the Decision Operator, which ensures the retrieval optimal solution. Furthermore, the proposed algorithm includes a special isolation mechanism to isolate first order clauses with optimal states during the optimization process. For a comprehensive assessment of the proposed algorithm, a comparative analysis will be conducted by using 10 established baseline algorithms that are commonly used for optimization purposes.

Scopes of Methodology (c)

While the proposed retrieval algorithm can be an effective strategy for achieving a high-quality final neuron state through the multi-objective function, several scopes must be addressed to ensure the reproducibility of the objective in this thesis. First, it is crucial to ensure that all the retrieved final neuron states achieve the global minimum energy before the states are optimized. Second, the existing retrieval algorithms must align consistently with the given objective functions. Third, the retrieval phase of DHNN aims to obtain the optimal final neuron states by achieving global solutions, diversifying the solution string, and maintaining the lowest similarity index. There are no trade-offs between these three objectives, which is why the multi-objective function in this thesis is not associated with Pareto optimality. Fourth, according to Parsajoo *et al*. (2022), tuning parameters affect the global search capability. Therefore, the parameter configurations for the proposed algorithm and the baseline algorithms will be based on successful tests conducted in previous studies. This approach helps prevent the final neuron states from being trapped in local minimum energy.

Research Methodology (d)

This thesis will establish a new supervised preprocessing phase to remove any inconsistencies or errors present in the dataset. This preprocessing phase will aid in correctly selecting the attributes, thereby enhancing the overall quality of the logic mining analysis (Alway *et al*., 2023). The phase involved three steps: First, cleaning the dataset consists of removing irrelevant and duplicated data, as well as imputing missing data using the Multiple Imputation by Chained Equations technique. This technique was chosen due to its ability to handle missing data by creating multiple imputations and providing more reliable results compared to simpler imputation methods. The performance of the imputation technique will be compared with two existing methods, *k*-Nearest Neighbour and Mean imputations after imputation assigns the outlier by box plot. Second, the Elastic Net regularization feature selection technique is employed to identify and select the most significant *n* attributes from the dataset. The performance of Elastic Net regularization will be compared with the correlation feature selection method introduced by Rusdi *et al*. (2023). Then, the dataset will transform into a bipolar representation. This step is crucial to ensure a suitable setup of the probability logic phase in Objective (b), creating a correct S-Type Random 2 Satisfiability compatible with DHNN. Third, the dataset will be split into train and test folds to avoid overfitting. The *k*-fold cross-validation approach will be used to split and will be compared with the *k*-fold leave-one-out cross-validation approach.

Scopes of Methodology (d)

Although this methodology can effectively prepare the datasets for data mining tasks, several scopes need to be addressed to ensure the reproducibility of the objectives in this thesis. First, the Multiple Imputation by Chained Equations technique will be used to handle the missing data since the missing data is classified as Missing at Random (Carreras *et al*., 2021). Second, this thesis employs the deletion of outlier data, which is assigned by using the box plot. This deletion is due to the small amount of such data present in the selected dataset. Third, categorical variables in the dataset used in this thesis will be transformed into discrete data by the label encoding method. Fourth, the parameter setup for all current techniques used in this phase will be based on successful testing conducted in previous studies. This ensures that the chosen parameters are validated and appropriate for the specific dataset and objectives of this research.

Research Methodology (e)

To address the limited number of CAMs issue, an improved Reverse Analysis method, namely S-Type Random k Satisfiability Reverse Analysis for $k=1,2$ with Dynamic Unit Discrete Hopfield Neural Network has been proposed. First, the probability logic phase in the proposed logic mining approach will create logic with optimal attributes from the dataset, depending on the data distribution, as outlined in Research Methodology (d). Second, the summation of True Positives and True Negatives is evaluated based on the dependent attribute in the training dataset. This approach considers both negative and positive instances in a dataset as proposed by Zamri *et al*. (2024). Third, all logical rules with the highest summation will be identified as the best logic. Subsequently, due to the Dynamic Unit DHNN, each best logic selected will be trained independently and in parallel by using the Election Algorithm as introduced in Research Methodology (c). Moreover, the Dynamic Unit DHNN ensures that each best logic corresponds to each distinct CAM. Fourth, each best logic will proceed in parallel through the independent retrieval phase. The proposed logic mining model will expand the search space and identify all high-quality induced logic enhanced by the proposed metaheuristics outlined in Objective (c). Finally, the new proposed shifting operator will be applied to determine the best arrangement for an attribute within the induced logic structure. Then, all induced logic with the highest summation will be regarded as the optimal induced logic. To evaluate the effectiveness and robustness of the proposed logic mining approach in this thesis, 20 repository datasets obtained from reputable databases will be used (Romli *et al.*, 2024).

Scopes of Methodology (e)

Although this methodology can be an effective way to extract information from the datasets, numerous scopes need to be addressed to ensure the reproducibility of the objective in this thesis. First, the number of attributes selected is restricted to the number of neurons in the logic. Second, for a fair comparison, the chosen datasets provide similar values of the logic parameter to ensure a consistent structure of the proposed logic. Additionally, the same number of k-folds will be applied to all models. Third, all logic mining models employ the Wan Abdullah method to obtain synaptic weight due to the tendency of Hebbian learning to generate sub-optimal values and encourage neuron oscillations (Sathasivam *et al.*, 2020b). Fourth, the chosen performance metrics for the induced logic must be distinct and should not provide the same interpretation. By applying distinct performance metrics, the experimental results effectively demonstrate the overall performance of the proposed logic mining model compared to other baseline models Zamri *et al.* (2024). Finally, the datasets used for this thesis were obtained from the University of California Irvine Machine Learning Repository and the Kaggle community platform, which are popular resources for data scientists and machine learning enthusiasts. The testing datasets will be utilized to compare the performance of the proposed logic mining model with existing logic mining models. Figure 1.1 provides a detailed illustration of the overall objectives and methodologies discussed in this thesis.

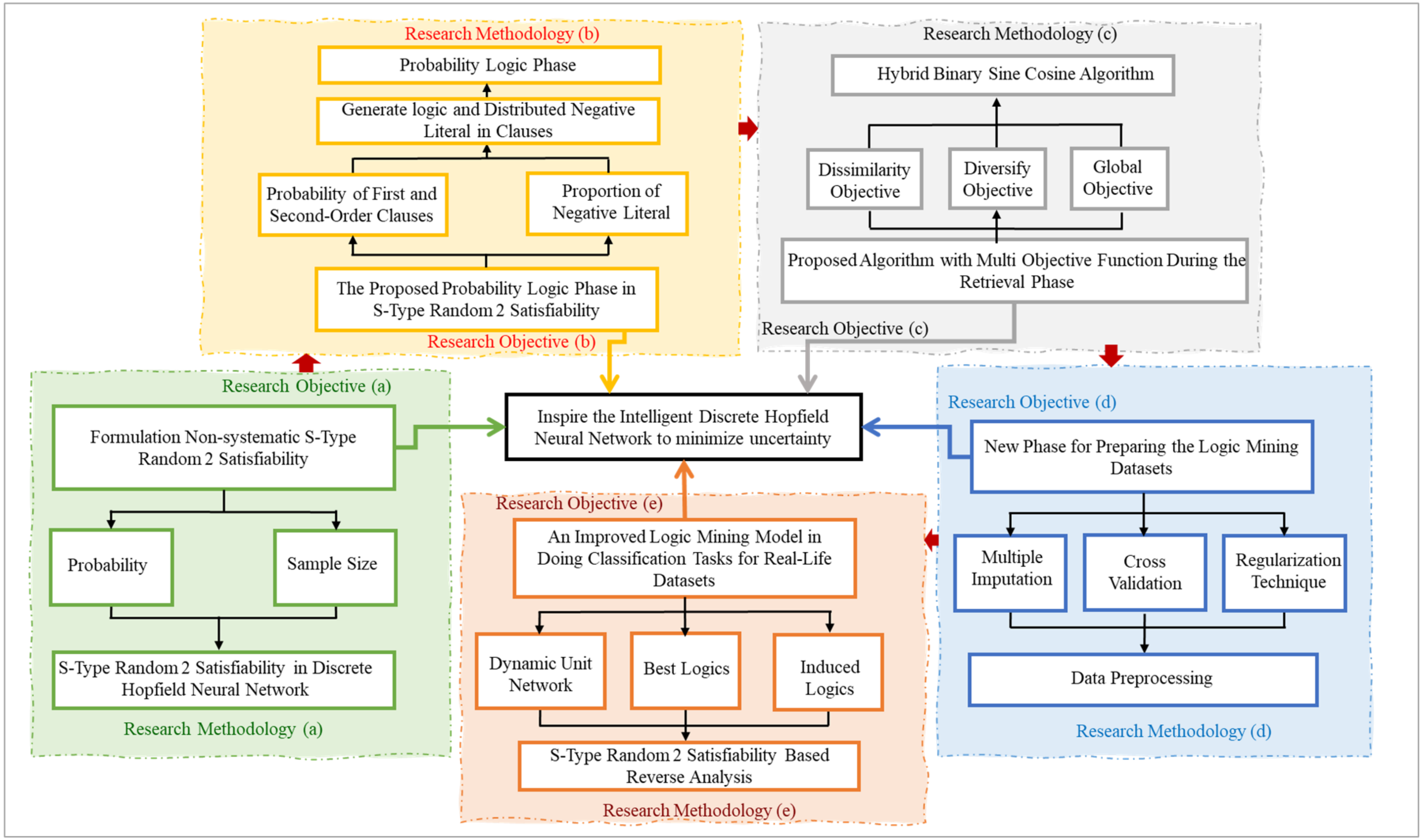

Figure 1.1 The thesis flowchart.

1.8 Organization of Thesis

The remainder of this thesis is organized as follows: In Chapter 2, a literature review of DHNN with various designs and applications will be explored. The significance of both systematic and non-systematic SAT applied in DHNN will be reviewed. Additionally, this chapter will explore various studies on the advancement and importance of Probabilistic Satisfiability representation. Furthermore, Sine Cosine Algorithms and their applications will be discussed. Finally, studies in DHNN logic mining will be thoroughly reviewed. Chapter 3 demonstrates the theoretical framework of Propositional Satisfiability in DHNN. It begins with a discussion on Boolean SAT, followed by the concept of DHNN as an intelligent model and the training and retrieval phases. Then, the optimize training phase by Election Algorithm. This chapter also reviews existing SAT logic in DHNN. Furthermore, the framework of the Sine Cosine Algorithm and the Differential Evolution Algorithm will be presented. Finally, existing logic mining models will be discussed. Chapter 4 introduces the concept of a new proposed non-systematic logic rule S-Type Random 2 Satisfiability with a new probability logic phase and implementation in the DHNN. This chapter explores establishing a hybrid network by embedding an improved metaheuristic algorithm with multi-objective function during the retrieval phase. Additionally, the preprocessing techniques used in this thesis will be introduced. Furthermore, the proposed Dynamic Unit DHNN with a logic mining model will be discussed in detail. Chapter 5 demonstrates the overall performance analysis of the proposed model using simulated datasets. Performance metrics are applied to evaluate across probability logic and retrieval phases. Chapter 6 illustrates the details of the preprocessing phase and the methods used to prepare repository datasets. The chapter also introduces the