MULTI-OBJECTIVE BINARY CLONAL SELECTION ALGORITHM IN THE RETRIEVAL PHASE OF DISCRETE HOPFIELD NEURAL NETWORK WITH WEIGHTED SYSTEMATIC SATISFIABILITY

NURUL ATIQAH BINTI ROMLI

UNIVERSITI SAINS MALAYSIA

2024

MULTI-OBJECTIVE BINARY CLONAL SELECTION ALGORITHM IN THE RETRIEVAL PHASE OF DISCRETE HOPFIELD NEURAL NETWORK WITH WEIGHTED SYSTEMATIC SATISFIABILITY

by

NURUL ATIQAH BINTI ROMLI

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

September 2024

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious and the Most Merciful, I am grateful for Allah's guidance and strength, which enabled me to successfully complete this doctoral journey and comprehend this new knowledge.

There have been many ups and downs throughout this journey. At one point during my first year of doctoral studies, I planned to end this journey altogether. However, that intention had been stopped by my supervisor, Dr. Mohd Shareduwan Mohd Kasihmuddin. I am deeply thankful to Dr. Shareduwan for his unwavering support, profound knowledge, and insightful comments throughout this endeavour. I feel incredibly fortunate to have been guided by Dr. Shareduwan due to his patience, understanding, and positive encouragement, which consistently lifted my spirits during moments of demotivation. Honestly, completing this journey has been a struggle, and I realized that it would not have been possible without his constant support. Furthermore, I greatly appreciate Dr. Shareduwan's advice and wisdom, which I have found applicable not only to my studies but also to life in general.

I am deeply grateful to my parents, Romli Kassim and Siti Eshah Abd Samad, for their unwavering love, encouragement, and understanding throughout the highs and lows of my doctoral studies. Their continuous financial support has been invaluable during this journey. I am also indebted to my siblings, Nurul Nadiah Romli and Muhammad Amir Romli, who offered their assistance during challenging times. Special thanks go to my best friend and great motivator, Nur Amira Mohamad Jamil, for her unwavering encouragement and support in pushing me to complete this research. Their steadfast belief in my abilities and unwavering support provided the emotional strength needed to persevere through challenges and celebrate achievements.

I cannot thank enough Dr. Mohd Asyraf Mansor, Dr. Nur Fariha Syaqina Zulkepli, Dr. Nur Ezlin Zamri, Siti Zulaikha Mohd Jamaludin, Alyaa Alway, Siti Syatirah Sidik, Dr. Syed Anayet Karim, Amierah Abdul Malik, Chen Ju, Jiang Xiaofeng, and Guo Yueling for their invaluable knowledge sharing, encouraging comments, and helpful suggestions that significantly contributed to the completion of my research and enhanced my performance. Completing this research would not have been possible without the unwavering support of all members of the Artificial Intelligence Research Development Group (AIRDG). I extend my sincere appreciation to all AIRDG members for their substantial contributions to this endeavour. Additionally, I would like to express my gratitude to Nur ‘Afifah Rusdi, Nur Shazneem Roslan, Gaeithry Manoharam, and Suad Abdeen for their collaborative efforts throughout this research process.

Above all, I am deeply grateful to Prof. Dr. Hailiza Kamarulhaili, Dean of the School of Mathematical Sciences, as well as the academic and administrative staff at Universiti Sains Malaysia (USM) and the Institute of Postgraduate Studies (IPS), for creating a supportive academic environment and providing resources that greatly facilitated my research endeavours. I am truly 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

P_{2SAT}	2 Satisfiability logic
$A\beta$	Accumulation of β-amyloid
S_i	Activated neuron update
$\wedge$	AND
β	Antibody or B-cell in bCSA
v_i^*	Attracted voters
$S_{benchmark}$	Benchmark state of neuron
P_i^{best}	Best logic
P_{kSAT}^{best}	Best logic of $kSAT$
$+ / - / =$	Better / Worst / Similar
$C^{(k)}$	Clauses with any k-order
δ_C	Clone rate
δ_C	Cloning rate
$\left(\Gamma_i\right)_{i\in I}$	Collection of open sets
$E_{P_{2SAT}}$	Cost function of 2SAT
$E_{P_{rS\,2\mathrm{SAT}}}$	Cost function of rS2SAT
f_i	Current fitness of the neurons
d	Degree of negativity
P_d	Dependent attribute
β_d	Diversity affinity of B-cell
F_d	Diversity fitness
R_{GD}	Diversity global ratio
D_i	Diversity property of B-cell
ω	Diversity rate

D_r	Diversity ratio
N_{party}	Divided solution spaced
$\in$	Element of
X_{ij}^{i}	Entries in raw data
$\mathbb{R}^d$	Euclidean space
$H_{P_{2SAT}}^{\min}$	Expected global minimum energy of DHNN for 2SAT
FN	False Negative
FP	False Positive
$H_{P_{2SAT}}^{final}$	Final energy of DHNN for 2SAT
S_i^f	Final neuron state
S_i^f	Final neuron state
β_f	Fitness affinity of B-cell
f_{H_j}	Fitness function (eligibility value) for the candidate
f_{LP}	Fitness function of the logical rule
F_i	Fitness property of B-cell
g	Gene
G_f	Global ratio
H_j	Highest eligibility value
$\tanh(h_i)$	Hyperbolic tangent activation function
L_{ij}	Inconsistency of logical rule
M_i	Individuals consist of candidate and voters
$S_i^{induced}$	Induced literal
P_{ACC}^{I}	Induced logic for accuracy
P_{MCC}^{I}	Induced logic for Matthews correlation coefficient
P_{NPV}^{I}	Induced logic for negative predictive value
P_{SNS}^{I}	Induced logic for sensitivity
P_{SPC}^{I}	Induced logic for specificity

P_{kSAT}^{i}	Induced logic of $kSAT$
v_i^J	Influenced voter
$P_j^{initial}$	Initial learning logic
MAE_L	Learning mean absolute error
SSE_L	Learning sum squared error
$\leq$	Less than equal
h_i	Local field
P	Logical Permutation
$H_{P_{2SAT}}$	Lyapunov energy function for 2SAT
$H_{P_{rS2SAT}}$	Lyapunov energy function of DHNN for rS2SAT
MCC	Matthews correlation coefficient
$N_{P_i^{initial}}$	Maximum number of initial learning logic
MAE_{GD}	Mean absolute error of diversity global
$\bar{A}_i$	Mean of the attribute
$>$	More than
δ_M	Mutation rate
δ_M	Mutation rate
$\mathbb{N}$	Natural numbers
$\neg$	Negation
$\neg A_i$	Negative literal (False)
NPV	Negative Predictive Value
NC	Neuron Combination
V_i^{new}	New individual
$\neq$	Not equal to
H_o	Null hypothesis
y	Number of 2SAT clauses
I	Number of attributes in the dataset

N_β	Number of B-cells
q	Number of best logic
A	Number of instances in the dataset
hc	Number of hypercubes
n	Number of neurons
k	Number of orders
N	Number of selected attributes
NT	Number of trials
N_{S_j}	Number of voters influenced by the candidate
$\vee$	OR
hp	Overlapping percentage
P_J	Parties
$f^{-1}(\Gamma_i)$	Points in the preimage
σ^P	Positive advertisement rate
A_i	Positive literal (True)
$P_i^{induced}$	Possible induced logic
p	*p*-value
$P_{RAN2SAT}$	Random 2 Satisfiability logic
N_{pop}	Random initialized population
$R(i)$	Random integer
r	Ratio of negative literals
R	Relaxation rate
R_s	Rogers-Tanimoto index
$RMSE_{GD}$	Root mean square error of diversity global
F_i	Scoring mechanism between the benchmark state and the final neuron state
δ_S	Selection rate
δ_S	Selection rate

SNS	Sensitivity
K_m	Sets of cluster
J_I	Similarity coefficient
γ	Similarity rate
SPC	Specificity
Δr	Step size ratio of negative literals
S_i^{sf}	Superior final neuron state
$W_i^{(1)}$	Synaptic weight for the first-order
$W_{ij}^{(2)}$	Synaptic weight for the second-order
W_{ij}	Synaptic weight from unit i to j
$\beta_{affinity}$	The affinity value of each B-cell
C_{r_i}	The chromosome
v_i^c	The coalition voter
CR	The crossover probability
$f_{\max}$	The desired fitness value
$\omega_{v_i^J}$	The eligibility distance coefficient
$\omega_{v_i^*}$	The eligibility distance coefficient of attracted voters
$\omega_{v_i^c}$	The eligibility distance coefficient of the coalition voter
$f_{v_i^*}$	The eligibility of the candidate of attracted voters
$f_{v_i^c}$	The eligibility of the coalition voter
$n\left(C_i^{(k)}\right)$	The frequency of the clauses
R_β	The maturation level of the B-cell
$n\left(P_j^{initial}\right)$	The maximum number of initial learning logic
α	The multiplication of number of trials and neuron combination
σ^N	The negative advertisement rate
N_D	The number of chromosomes selected
N_j	The number of individual

$N_{v_i^*}$	The number of influenced voters
$S_{v_i^*}$	The number of neuron states of attracted voters
P_{learn}	The outcome of the learning data
P_{test}	The outcome of the testing data
V_i	The possible solution to the problem
$\lfloor \ \rfloor$	The rounding down with the floor function
R_{β_d}	The scoring mechanism of each B-cell with respect to diversity
R_{β_f}	The scoring mechanism of each B-cell with respect to fitness
$C_{r_i}^*$	The selected chromosome
σ	The selection rate of the chromosomes
S_j	The state of neuron j
$*$	The value of ROI is not able to be evaluated due to division by zero
$f_{v_i^J}$	The voter
Tol	Tolerance value
λ	Total number of desired negative literals
n_{r2SAT}	Total number of literals exists in r2SAT
n_{2SAT}	Total number of literals in 2SAT
$n_{RAN2SAT}$	Total number of literals in RAN2SAT
N_v	Total number of negative literals exists in r2SAT
TV	Total number of neuron variations
$n(K)$	Total number of superior final neuron state that achieved global solution
$n(Q)$	Total number of superior final neuron state that achieved both global and diversity solution
κ	Total weight of negative literal
TN	True Negative
TP	True Positive
$g(h_i)$	Updated neuron states
Γ_i	Weight for each literal

η_i	Weight of negative literal
P_{r2SAT}	Weighted Random 2 Satisfiability logic
$P_{rS\,2\mathrm{SAT}}$	Weighted Systematic 2 Satisfiability logical rule

LIST OF ABBREVIATIONS

$\delta 2SAT$	S-Type Random 2 Satisfiability
2HESRA	Major 2 Satisfiability Hybrid Exhaustive Serach Reverse Analysis
2SAT	2 Satisfiability
2SATRA	2 Satisfiability based Reverse Analysis Method
3SAT	3 Satisfiability
ABC	Artificial Bee Colony
ABCJ	Artificial Bee Colony Algorithm by Jia et al.
ABCK	Artificial Bee Colony Algorithm by Kasihmuddin et al.
ABCS	Artificial Bee Colony Algorithm by Sidik et al.
ACC	Accuracy
AD	Alzheimer’s Disease
ADNI	Alzheimer’s Disease Neuroimaging Initiative
AERA	Amazon Employee Resources Access
AI	Artificial Intelligence
ANN	Artificial Neural Network
BBHA	Binary Black Hole Algorithm
bCSA	Binary Clonal Selection Algorithm
BHA	Black Hole Algorithm
bWOA-S	S-shaped Wolf Optimization Algorithm
bWOA-V	V-shaped Wolf Optimization Algorithm
CAA	Cerebral amyloid angiopathy
CAM	Content Addressable Memory
CNF	Conjunctive Normal Form
CRAN2SAT	Conditional Random 2 Random

CSA	Clonal Selection Algorithm
DEA	Differential Evolution Algorithm
DHNN	Discrete Hopfield Neural Network
DHNN- $\delta 2SAT$	$\delta 2SAT$ in DHNN
DHNN- *r*2SAT	*r*2SAT in DHNN
DHNN-2SATGA	GA as learning algorithm in 2 Satisfiability in DHNN
DHNN3-SATCSA	CSA as learning algorithm in 3 Satisfiability in DHNN
DHNN-HORNSAT	Horn Satisfiability in DHNN
DHNN-MAX*k*SAT	Maximum *k* Satisfiability in DHNN
DHNNRAN*k*SAT	RAN*k*SAT in DHNN
DHNN-SAT	SAT embedded in DHNN
E2SATRA	Energy based 2 Satisfiability Reverse-based Analysis Method
EA	Election Algorithm
EDA	Estimation Distribution Algortihm
E*k*SATRA	Energy *k* Satisfiability Reverse Analysis
ES	Exhaustive Search
GA	Genetic Algorithm
GBD	Global Burden of Disease
GHz	Gigahertz
GWO	Grey Wolf Optimization
HEA	Hybrid Election Algorithm
HES	Hybrid Exhaustive Search
HI	Human Intelligence
HNN2SAT-ABC	ABC as learning algorithm in 2 Satisfiability in HNN
HNN2SAT-ES	ES as learning algorithm in 2 Satisfiability in HNN
HORNSAT	Horn Satisfiability
HTAF	Hyperbolic Tangent Activation Function

IPS	Institute Postgraduate Studies
*k*SATRA	*k* Satisfiability based Reverse Analysis Method
LOAD	Late-onset Alzheimer’s Disease
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAX-*k*SAT	Maximum *k* Satisfiability
MCI	Mild Cognitive Impairment
MFT	Mean-Field Theory
MFTDHNN-2SAT	MFT in the DHNN-2SAT
MHNN	EDA in the DHNN
MRI	Magnetic Resonance Imaging
NFT	Neurofibrillary Tangles
P2SATRA	Permutation 2 Satisfiability Reverse Analysis
PD	Parkinson’s Disease
PET	Positron Emission Tomography
PSO	Particle Swarm Optimization
*r*2SAT	Weighted Random *k* Satisfiability
RA	Reverse Analysis
RAM	Random Access Memory
RAN2SAT	Random 2 Satisfiability
RAN*k*SAT	Random *k* Satisfiability
RANMAX2SAT	Random Maximum 2 Satisfiability
RK4	Runge Kutta 4
RMSE	Root Mean Square Error
ROI	Ratio of improvement
*r*S2SAT	Weighted Systematic 2 Satisfiability
*r*S2SATRA	Weighted Systematic 2 Satisfiability Reverse Analysis

S2SATRA	Supervised 2 Satisfiability Reverse-based Analysis Method
SAT	Satisfiability
SCA	Sine Cosine Algorithm
TDA	Topological Data Analysis
TSP	Travelling Salesman Problem
U2SAT	Weighted Systematic 2 Satisfiability Modified
U2SATRA	Weighted Systematic 2 Satisfiability Modified Reverse Analysis
UCI repository	UC Irvine Machine Learning repository
USM	Universiti Sains Malaysia
WA	Wan Abdullah
YRAN2SAT	Y-Type Random 2 Satisfiability

LIST OF APPENDICES

ALGORITMA MULTI-OBJEKTIF PILIHAN KLON PERDUAAN DALAM FASA PEROLEHAN SEMULA RANGKAIAN DISKRET NEURAL HOPFIELD DENGAN SATISFIABILITI PEMBERAT SISTEMATIK

ABSTRAK

Kestabilan rangkaian diskret neural Hopfield bergantung kepada kebolehan rangkaian untuk mengawal hubungan neuron yang menyebabkan beberapa isu untuk timbul seperti taburan rawak literal positif dan negatif dan keadaan neuron akhir berpadanan. Oleh itu, tesis ini mencadangkan peraturan logik Satisfiabiliti sistematik yang baharu iaitu 2 Satisfiabiliti Sistematik Berpemberat yang menggunakan ciri berpemberat untuk mengawal taburan literal negatif. Logik yang dicadangkan telah dimasukkan ke dalam rangkaian diskret neural Hopfield dan pengoptimuman fungsi multi-objektif telah diambil kira untuk mengesan keadaan neuron akhir yang unggul. Algoritma Pilihan Klon Perduaan telah dicadangkan untuk memastikan penjanaan optimum keadaan neuron akhir yang unggul. Algoritma yang dicadangkan di dalam fasa perolehan semula telah menunjukkan prestasi yang optimum berbanding dengan algoritma asas. Peraturan logik dan algoritma yang baharu dicadangkan akan menjadi komponen-komponen dalam model perlombongan logik iaitu Analisis Berbalik Yang Terubahsuai Dengan 2 Satisfiabiliti Sistematik Berpemberat. Model perlombongan logik yang dicadangkan mampu memperolehkan semula logik teraruh terbaik yang mewakili corak optimum bagi set data. Berdasarkan penemuan, model perlombongan logik yang dicadangkan mengatasi perlombongan logik asas yang lain bagi semua metrik prestasi yang digunakan dalam set data repositori. Model perlombongan logik yang dicadangkan telah teruji dalam set data hidup nyata daripada *Alzheimer's Disease Neuroimaging Initiative* dan telah menunjukkan prestasi yang unggul.

MULTI-OBJECTIVE BINARY CLONAL SELECTION ALGORITHM IN THE RETRIEVAL PHASE OF DISCRETE HOPFIELD NEURAL NETWORK WITH WEIGHTED SYSTEMATIC SATISFIABILITY

ABSTRACT

The stability of the Discrete Hopfield Neural Network is dependent on the ability of the network to govern the neuron connections that caused several issues to arise, such as random distribution of positive and negative literals and overfitting final neuron states. Therefore, this thesis proposes a new systematic Satisfiability logical rule namely Weighted Systematic 2 Satisfiability that uses a weighted feature to control the distribution of the negative literals. The proposed logic embedded into Discrete Hopfield Neural Network and considered the optimization of multi-objective function in the retrieval phase to locate superior final neuron states. A Binary Clonal Selection Algorithm is being proposed to ensure optimal generation of the superior final neuron states. The proposed algorithm in the retrieval phase showed optimal performance as compared to the baseline algorithms. The newly proposed logical rule and the algorithm will be the components in the logic mining model namely Weighted Systematic 2 Satisfiability Modified Reverse Analysis. The proposed logic mining model is able to retrieve best induced logic that represents the optimal patterns of the dataset. Based on the findings, the proposed logic mining model outperformed other baseline logic mining models for all the performance metrics used in the repository dataset. The proposed logic mining model was tested on a real-life dataset from the Alzheimer's Disease Neuroimaging Initiative, and it showed superior performance.

CHAPTER 1

INTRODUCTION

1.1 Overview

Over the past decade, the rapid advancements of Artificial Intelligence (AI) have revolutionized numerous aspects of our daily lives, from how we interact with technology to how industries operate and innovate. The term “AI” is frequently applied to developing systems endowed with intellectual processes similar to human characteristics, such as the ability to reason, discover meaning, generalize, or learn from past experience. Within the field of AI, Artificial Neural Network (ANN) play a significant role in modelling complex relationships and patterns, making them an integral part of AI technology. In this chapter, the introduction on ANN is briefly explained which align with the context of this thesis. Next, the suggested approach to address the current challenges in ANN is critically explained. With thorough justification, this thesis highlighted the problem statements in creating optimal AI system. Following to that, the research objectives, research questions, and supported methodologies involved in this thesis are addressed.

1.2 The Fundamentals of Artificial Intelligence

Intelligence is defined as the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly, and learn from experience. Intelligence can be reviewed from psychology and computer science perspectives, which correspond to human and artificial, respectively. Though there are many disagreement on what human intelligence (HI) is, there is broad agreement that HI is a psychological construct (Gignac & Szodorai, 2024). Psychological construct is an abstract, unobservable, hypothetical entity inferred from postulated thoughts and observable behaviours, representing patterns of psychologically related phenomena. In plain language, a psychological construct is a concept developed to describe a specific aspect of the mind or behaviour that is not directly observable but is inferred from patterns in thoughts, feelings, and actions. In addition to intelligence, examples of well-established psychological constructs include anxiety, self-esteem (Pyszczynski *et al.*, 2004), and motivation (Touré-Tillery & Fishbach, 2014). AI is not a psychological construct, as it does not originate from the same underlying human cognitive or emotional processes. Instead, AI may be considered as a computational construct, as it is inferred from the outcomes of simulated aspects of human thought and decision-making, which are facilitated by data processing, machine learning techniques, and algorithmic principles (Prasad *et al.*, 2023). Additionally, AI has evolved through computer science and engineering advancements, marked by human-initiated intervention, intellectual effort, and purposeful innovation. In recent years, AI has been making strides in transforming the way to understand intelligence. Particularly, AI is composed of many approaches that allow computer programs to address specific input and provide value-added output. These approaches refer to the methods, algorithms, and data science methods to perform tasks that traditionally require human beings.

In simple terms, it has been suggested that AI is what one does when one does not know what to do. This highlights the importance of novelty to the entity when encountering and solving intellectual problems, a crucial component of valid intelligence testing in humans. Therefore, it is imperative to allow AI to be truly intelligent than humans which resulting to the following perspective,

"*Scale up the intelligence of Artificial Intelligence.*"

In literal terms, scale up means the process of increasing the size or scope of something. Scaling up commonly involves expanding operations or growing a venture to reach a larger audience or market. In this context, scaling up the intelligence of AI refers to improving the solutions retrieved by the network by adding an optimizer to make the AI model smarter than humans. This process enhances the capability of AI in producing optimal solutions and makes AI systems more robust and interpretable. By contributing to the advancements in AI, this helps pave the way for more intelligent and capable AI systems that can address complex challenges. One of the potential prospect of AI is through ANN. ANN are inspired by the way of biological nervous systems process information and gives appropriate feedbacks. Being a fundamental subset of AI, ANN is a computational model designed to simulate the way the human brain works in analyzing and processing information. ANN has a wide application that had been proposed in many areas in gathering knowledge by revealing patterns and relationships among data during learning. One of the variant in ANN is known as Discrete Hopfield Neural Network (DHNN). Throughout the years, DHNN has widely been applied to solve optimization problem. However, DHNN are limited in interpreting the output of the network to the user. Therefore, the following section will discussed in detail the challenges occurs in the existing DHNN.

1.3 Existing Challenges in Discrete Hopfield Neural Network

DHNN is one of the earliest ANN proposed by Hopfield and Tank (1985) to provide potential solution for Travelling Salesman Problem (TSP) through connectionist model. Generally, DHNN consists of fully interconnected neurons with input and output neurons without hidden layers. Each neuron is connected through synaptic weight, where synaptic weight measures the connection between the neurons.

The synaptic weights are the building block of Content Addressable Memory (CAM). Notably, the reason why CAM is called that way is because DHNN has the capability to access the memory address and search the closest match for the required pattern (Wen *et al.*, 2009). Initially, each neuron fires and updates iteratively until the final neuron state converges towards the optimal solution. According to Bruck and Sanz (1988), the network can converge to a minimum from any initial neuron state which makes DHNN became attractive to various applications. Thus, each neuron update will enhance the synaptic weight among neurons until the optimal CAM is constructed. In addition, the quality of the neuron state in DHNN can be measured in terms of energy function which is always being minimized by the network.

One of the perspective to improve the quality of the neuron state in DHNN is by implementing symbolic rule via logical rule. In 1992, Abdullah propose a method of expressing special type logic namely Horn Satisfiability (HORNSAT) via DHNN. The main motivation behind the proposed work is to create a method of doing logical inference through minimization of logical inconsistencies on a symmetric neural network. In recent years, the development of logical rules has been proposed by many researchers, such as Jiang *et al.* (2024), Roslan *et al.* (2024), and Alway *et al.* (2022). However, the logical rules proposed focused more on higher-order logical rule. Higher-order logical rule is more complex compared to the lower-order. The complexity often leads to difficulty in decision-making, where the networks struggle to determine whether the statements are true or false. This can reduce the reliability and effectiveness of the system designed to reason and make decisions based on logical rules. Therefore, the lower-order logical rule is more efficient than the higher-order logical rule because it is less complex, and the network can ensure reliability and predictability.

Another perspective of DHNN that can be improved is by making the network more transparent. Like any other variants of ANN, the internal operation of the network is unknown due to the nature of the black box mechanism. Thus, any changes in the neuron connections that lead to the final output of the network cannot be tracked or detected. In reality, the neuron states in DHNN converges from an initial state into a final neuron state without knowing exactly which direction of the neuron signal. This showed that the existing DHNN is unable to detect which part of the network is successfully optimized and which of the network that is wrongly optimized. One of the approaches to improve the modelling of DHNN is by considering the neuro-modelling technique using a sort of formal language via symbolic logic. According to Abdullah (1992), the role of logic in DHNN is to act as the governing rule in representing the neurons. In this context, the set of solutions that minimize the energy function in the network is equal to the set of truth assignments that satisfied the Satisfiability (SAT) logical rule. Therefore, the implementation of SAT into DHNN provided better understanding of the hidden units in the network. Due to these assumptions, several researchers choose for other neuro-symbolic models of SAT in DHNN (DHNN-SAT) by considering other possible satisfiable SAT structures such as 2 Satisfiability (2SAT) and 3 Satisfiability (3SAT). Ideally, the black box property of DHNN is further refined when considering the dynamic structure of satisfiable SAT as neuron representation in the network. This allows DHNN to have better neuron navigations with interpretable input and output for full transparency.

Another improvement that can be highlighted in the existing DHNN is the issue of diversity. Particularly when dealing with large-scale data, the performance of DHNN deteriorates as an excessive amount of learning data can cause overfitting. Several researchers, such as Karim *et al.* (2022) and Alway *et al.* (2023) focused on improving the learning phase in order to overcome the overfitting issue. Consequently, this will increase the risk of retrieving overfitting final neuron states. This is because the production of final neuron states only relies on the random initial state in the local field and activation function. Hence, there is a high possibility that the same initial state is executed for the local field. Even though the final neuron state produces an optimal solution, there is no variety in the solution. This shows that with large-scale data, the network can only produce one pattern of solution. Ideally, the diversity of the final neuron state can be expanded by inserting a particular type of objective function that is able to offer a variety of solutions produced in the retrieval phase. This allows DHNN to produce a diversified final neuron state without relying on the local field and activation function.

Despite having numerous applicability, conventional DHNN is prone to storage capacity issues. This means that the number of stored memory patterns is severely limited because the network learns the data from one point of view. Most of the researchers (Alway *et al.*, 2020; Kho *et al.*, 2020; Jamaludin *et al.*, 2023) focus on learning data that have positive outcomes in finding the best logic to represent the data. This will affect the performance of the model in learning the behaviour of the data because most of the real-world datasets often have imbalanced distributions, where one or more classes of outcomes may be underrepresented or overrepresented, which results in an uneven distribution of outcomes. This can lead to an ineffective model in classification, as minority classes may be overlooked due to having fewer data to learn from. This has been described in Ramyachitra and Manikandan (2014), where imbalanced datasets can affect the classification performance. Imbalance datasets can deteriorate the performance of the network because of the difficulty in learning and evaluating the data, which indicates the need to consider both positive and negative outcomes of the data. This indirectly increases the ability of the network to find optimal induced logic because the network stored the pattern of the data in both perspectives of outcome.

Despite powerful feature extraction capabilities, DHNN is sensitive to input data quality and preprocessing. This means that the performance of DHNN heavily relies on the characteristics of the input data. Therefore, Kho *et al.* (2020) option for data dimensionality reduction to only bipolar inputs. Conjointly, binary inputs require several modifications in the units of the network. For example, the evaluation of the final energy by the existing Lyapunov energy function would always converge to zero. Consequently, the distinction of global and local minimum solutions of the final output cannot be detected. Building a model is one thing, but understanding the data that goes into the model is another. Ideally, DHNN can perform effective classification tasks when suitable data pre-processing is applied. This allows DHNN to produce a significant relationship between the attributes and the target variable. Based on all observations, this resulting to a broader philosophical standpoint, whereby:

"*The advancement of existing components in the DHNN with respect to diversity.*"

In layman terms, DHNN can be deemed as one of the highly intelligent computational models when suitable mechanisms are added to the DHNN. One of the ways to make DHNN to be an intelligent computational model is by improving the existing components in the DHNN. This approach can increase the potential of DHNN as a robust optimization model that can be applied to other exciting applications. In this thesis, the improvements of the existing DHNN are based on various perspectives, such as the structure of the SAT logical rule, optimization in the retrieval phase, and logic mining. The discussion of each perspective will be covered in the problem statements section.

1.4 Problem Statements

Problem Statement 1

In practice, effective symbolic rule is essential to model the connection of the neurons in DHNN. Humans use this form of knowledge representation to make decisions according to various conditions. As an example, Satisfiability (SAT) formulation expressed in Conjunctive Normal Form (CNF) embodied decisions in the form of AND-OR rules. As an important type of cognitive intelligence, each structural component in SAT must be rightly represented before being encoded as symbolic rules into any computational system. An existing work by Kasihmuddin *et al.* (2017a) and Sathasivam *et al.* (2020a) first employed systematic and non-systematic SAT, known as 2 Satisfiability (2SAT) and Random 2 Satisfiability (RAN2SAT), respectively. Both of the works disregarded the distribution of positive and negative literals throughout respective SAT logical rules where the distribution is set randomly. When the SAT operates in DHNN, the network is unable to offer different variants of the logical rule and the composition of negative or positive literals is set at random. Therefore, it is quite impossible to understand the processing units of the neuro-symbolic model comprehensively. In another development of the non-systematic SAT logical rule, existing work by Zamri *et al.* (2022a) proposed Weighted Random *k* Satisfiability (*r*2SAT), while Roslan *et al.* (2024) proposed conditional Random 2 Satisfiability (CRAN2SAT). Both of the works by Zamri *et al.* (2022a) and Roslan *et al.* (2024) control the distribution of the negative literals in the logical rule. On the other hand, Guo *et al.* (2022) proposed another class of SAT that combines both features of systematic and non-systematic logical rule known as Y-Type Random 2 Satisfiability (YRAN2SAT). Even though *r*2SAT and CRAN2SAT have specific features to control the distribution of the negative literals in the logical rule, all the development of the non-systematic SAT possesses first-order logic that will degrade the quality of the logical rule as neuron representation. This is because first-order logic is unable to offer diversified solutions since there is only one exact states that will satisfy the clause. Even by acknowledging the importance of negative literals, the overfitting solutions cannot be avoided. Therefore, it is imperative to propose a systematic logic that consists of an additional feature to control the distribution of negative literals. The proposed logic must not contain first-order logic that possibly increases the overfitting issue. The only way to investigate the potential of the proposed logical rule as a neuron representation is by embedding the proposed logic into a neural network medium.

Problem Statement 2

Logical rule is needed to govern the neuron connection in the network due to the black box property of DHNN. The structural components of the logical rule are exceptionally important to ensure that the neuron is best represented. Ineffective implementation of the logical rule in DHNN highly contributed to the poor synaptic weight management and overfitting output whereby the network only converge to only one type of solution. The earliest SAT to be embedded in DHNN (DHNN-SAT) models was Maximum *k* Satisfiability (MAX-*k*SAT) in DHNN (DHNN-MAX*k*SAT) by Kasihmuddin *et al.* (2018). However, the structure of MAX-*k*SAT contains redundant literals which leads to ineffective synaptic weight. When redundant variable exist in the logical structure, the effect of the synaptic weight will be cancelled out which results in non-zero cost function. Thus, this will affect other processing units in the DHNN due to the nature of the logical rule is unsatisfiable. In the context of satisfiable logical rule in DHNN, existing works such as *r*2SAT in DHNN (DHNN-*r*2SAT) by Zamri *et al.* (2022a) and S-Type Random 2 Satisfiability ($\delta 2SAT$) in

DHNN (DHNN-$\delta 2SAT$) by Abdeen *et al.* (2023) had an additional processes to generate the logical rule with desired number of negative literals. As a consequence, the practicality of adding an additional layer in DHNN is conflicted by high computational time. Therefore, it is imperative to proposed an additional feature to control the desire number of negative literals in the proposed logical rule without the needs of an additional layer. The additional feature must have the ability to produce logical structure with any number of negative literals with unbiased distribution. Successful implementation of SAT into DHNN can be verified with high number of global minimum solutions. However, due to the feedback property of the DHNN, any improvements during the retrieval phase is commonly disregarded.

<u>Problem Statement 3</u>

The dynamic behaviours of DHNN model are strongly dependent on its network structure. The conventional DHNN has a high chance of being trapped in local minimum when the number of neuron increased due to lack of interpretability and variation. Over the time, several researchers have incorporated learning algorithm into the DHNN to increase the searching capability of the DHNN. This can be seen in the existing work by Zamri *et al.* (2020) employed Clonal Selection Algorithm (CSA) as the learning algorithm in the 3 Satisfiability (3SAT) in DHNN model which is known as DHNN3-SATCSA. The proposed DHNN3-SATCSA was compared to the conventional Exhaustive Search (ES) and showed acceptable results with 100% global minimum solutions retrieved. Unfortunately, the optimization in the learning phase disregards the solutions in terms of diversity. Hence, the network tends to produce overfitting solutions. As the number of neurons increase, the performance of the network degrades due to the existence of local minimum solutions. Subsequently, Karim *et al.* (2022) and Alway *et al.* (2023) attempted to address the issue of overfitting by introducing the multi-objective function in the learning phase of the DHNN model. The proposed multi-objective function successfully produce multiple strings of neuron states that posses both high fitness and diversity. Both of the works attempted to produce diversified solutions by allowing multiple units of CAM in the learning phase of DHNN. However, the number of CAM are restricted to only five. Hence, the network has a tendency to retrieve repetitive final neuron states. The expansion of final neuron states with more variations and at the same time achieve global minimum solution is crucial to make DHNN as a highly intelligent optimization model. In this context, multi-objective function should be initiated in the retrieval phase of DHNN to identify final neuron states that lead to high fitness, diversity and dissimilarity. The overall performance of the model must be assessed accordingly once all the components in the proposed model is established. The proposed model should be tested with real-life dataset like any other developing data mining models.

Problem Statement 4

Logic mining is a subset of data mining where the information from the datasets were extracting in the form of logical rule. Instead of becoming the black box model that delivers the final outcome, logic mining gives better representation by translating the final neuron state into logical rule. Sathasivam and Abdullah (2011) is the first work that introduces logic mining namely Reverse Analysis (RA) which incorporates DHNN and HORNSAT. The proposed RA only transformed the final neuron states that achieves global energy profile into the induced logic. However, the proposed RA prone to several weaknesses. First, the proposed RA has no capability to process continuous data because there is no data filtering mechanism before the data was converted into logical rule. Second, the proposed RA has no capability to create one induced logic because all the obtained induced logic is considered as the potential behaviour to the datasets. By this standard, RA will produce thousands of unimportant induced logic that potentially creating overlapping induced logic. Third, there is no attempt to validate the performance of the induced logic. There is no proper validation metric to measure the quality of the induced logic. These weaknesses reduce the performance of the proposed RA as a classification model. Recent attempts by Kho *et al.* (2020) and Zamri *et al.* (2020) were initiated to improve the existing RA by capitalizing 2SAT and 3SAT, respectively as the logical rule in DHNN. Both works are able to extract the best induced logic that represents the behaviour of the dataset by considering accuracy value as the validation metric. Unfortunately, these works only focused on data entries or patterns that is associated to success or 1. This will cause a creation of non-holistic classification model when dealing with imbalanced datasets because another pattern of the data entries are removed and disregarded. Consequently, the retrieved induced logic are biased to only true positive classification. Therefore, the improved logic mining model must have the ability to handle continuous raw entries and obtain single best induced logic that is able to represent both negative and positive outcomes of the dataset. The improved logic mining model should assess the impact of different performance metrics as best logic to the quality of the retrieved final induced logic. The validation of the induced logic based on different performance metrics gives different insights or undiscovered patterns of the dataset.

Problem Statement 5

The production of the best logic is crucial as it demonstrates the optimal connections between neurons with the correct synaptic weight. The best logic will influence the quality of the retrieved induced logic because best logic learned the overall pattern of the dataset. Particularly for DHNN, optimal best logic allows the network to learn the data effectively. This will increase the possibility of retrieving best induced logic in the retrieval phase. Generally, majority of the existing logic mining models such as Energy based *k* Satisfiability Reverse Analysis (E*k*SATRA) by Jamaludin *et al.* (2020), supervised 2 Satisfiability Reverse Analysis (S2SATRA) by Kasihmuddin *et al.* (2022), and Permutation 2 Satisfiability Reverse Analysis (P2SATRA) by Jamaludin *et al.* (2023) improved the existing logic mining by adding an additional components in the model. The additional components added to these works are energy based selection for the retrieved induced logic in E*k*SATRA, statistical-based attribute selection method in S2SATRA, and permutation of retrieved induced logic in P2SATRA. Even though all the mentioned works improved the logic mining to increase the performance of the induced logic produced, the works still not able to retain the highest accuracy. This is because there is no additional components or improvements were applied in the learning phase of these models, which led to a suboptimal learning phase. Note that, the pattern of the data is learned thoroughly in the learning phase of the logic mining. Therefore, it is imperative to discover multiple best logic that considers both true and false classification of learning data. Each best logic focuses on the highest value of performance evaluation with respect to the learning data. The performance evaluation considered in the best logic will affect the performance of the retrieved induced logic. The extracted best induced logic provides insightful knowledge of the dataset because extensive possibilities of the best logic were considered.

1.5 Research Questions

In order to clarify the consistency between the problem statements given in Section 1.4 and research objectives of this thesis, this section lay out several important research questions to outline various aspects of the proposed study. Notably, these research questions are grounded on the following question,

"*How can the Discrete Hopfield Neural Network be scaled up to increase the solution space and achieve higher computational intelligence*?"

Therefore, the research questions involved in this thesis are listed as follows:

(a) What is the structural component that can be implemented in the formulation of systematic logic to ensure the dynamic distribution of negative literals?

(b) What is the alternative formulation and structure of the systematic logic that can control the distribution of negative literals and effectively represent the connection of the neuron in the Discrete Hopfield Neural Network?

(c) What are the additional objective functions that can be considered in the current retrieval phase of the Discrete Hopfield Neural Network to ensure the network has the ability in generating diverse final solutions?

(d) In the context of logic mining, how can the best logic be formulated based on the confusion matrix to ensure the best logic effectively extracts information of the dataset?

(e) What approach can be developed in the current logic mining to produce a set of induced logic that has the ability to extract optimal patterns of the dataset and perform classification tasks?

1.6 Research Objectives

This thesis is addressed by modelling Discrete Hopfield Neural Network using systematic logical rule. The proposed model will be implemented with nature-inspired algorithm that has the ability to minimize the cost function of the model. In this context, the proposed model must be computationally stable and has the capability to maximize the final solution. Therefore, the objective of this thesis are as follows:

(a) To propose a variant of systematic logical rule namely Weighted Systematic 2 Satisfiability. The proposed logical rule comprises of different number of negative literals. The number of negative literals generated is controlled by a weighted feature namely the ratio of negative literals.

(b) To propose the implementation of Weighted Systematic 2 Satisfiability into Discrete Hopfield Neural Network as the symbolic rule. The minimization of the cost function corresponds to the satisfiable property of all clauses in the logical rule. Maximum number of satisfied clauses indicates optimal learning phase of the network.

(c) To propose a multi-objective function during the retrieval phase of the Discrete Hopfield Neural Network to retrieve superior final neuron states that possessed high fitness, diversity, and dissimilarity coefficient value. A new retrieval algorithm namely the Binary Clonal Selection Algorithm will locate these superior final neuron states using several global and local search operators. Each string of superior final neuron states will depict optimal energy profile of the network.

(d) To propose a new computation in generating the best logic during the learning phase of logic mining. The best logic will be extracted by considering the highest value of the selected performance metrics. These metrics are measured based on the classification between the actual outcome of the learning dataset to the logical outcome of the initial learning logic. The extracted best logic will represent the majority patterns of the dataset with respect to the selected performance metrics.

(e) To propose a logic mining namely Weighted Systematic 2 Satisfiability Modified Reverse Analysis in executing classification tasks for various real-life datasets. Multiple best logic will configurate respective units of Discrete Hopfield Neural Network. The proposed logic mining will extract best induced logic that represent the overall behaviour of the analysed dataset.

1.7 Methodology and Limitations

Methodology and Limitation 1

Recently, the number of SAT formulations have increased dramatically. Therefore, it is imperative to choose the optimal SAT representation that can effectively govern the ANN. In terms of new the SAT formulation, this thesis formulates the Weighted Systematic 2 Satisfiability by capitalizing the systematic SAT structure with 2 literals per clause. The basis of the proposed SAT formulation was inspired by the following works. First, an existing work by Krom (1970) expressed the general SAT formula in Conjunctive Normal Form (CNF). In this context, each clause in the proposed SAT will be connected with disjunction that contributed to the final outcome of the Weighted Systematic 2 Satisfiability. Additionally, the systematic manner of the proposed Weighted Systematic 2 Satisfiability took inspiration from the work by Kasihmuddin *et al.* (2017a) which proposes a fixed logical order which respect to the second-order. The existing work further suggested a feature of clausal weight to control the generation of non-monotonic clauses. In this case, a non−monotonic clause refers to the clause that contains all negative literals. Inspired from Zamri *et al.* (2022a), the proposed Weighted Systematic 2 Satisfiability is inclusive of a weighted feature in the form of ratio of negative literals or *r* to control the distribution of negative literals. The distribution of which literal to be negated is set at random. According to Lallouet *et al.* (2020), appropriate method should be introduced to systematically tune the desired logical rule prior being a representation for a specific problem. Inspired from an idea expressed by Lallouet *et al.* (2020), this thesis implemented ES as a searching technique to generate the correct logical rule of Weighted Systematic 2 Satisfiability with respect to the desired ratio of negative literals. The implementation of ES is important to ensure that the correct structure of Weighted Systematic 2 Satisfiability is generated. In order to guarantee that DHNN learns the correct logical rule of Weighted Systematic 2 Satisfiability, the performance of the learning error is evaluated.

While this methodology can be an optimal approach to generate the proposed Weighted Systematic 2 Satisfiability, there are several limitations applied to ensure the reproducibility of the first objective in this thesis. First and foremost, ES is only obligated to generate Weighted Systematic 2 Satisfiability with respect to the ratio of negative literals. The reason why the proposed logic focused on the negative literals is to promote more emphasis on the role of negative literals as neuron representation. According to Saribatur and Eiter (2021), researchers commonly neglect negative literals because negative literals usually associated to fault or error to the goal of a SAT formula. By recognizing this gap, the role of negative literals is vital to improve the interpretation of the SAT as logical rule in DHNN. Indirectly, the negative literals in the proposed logical rule will make it possible for DHNN to obtain diverse final neuron states without additional modifications in the units of the network. Secondly, the ratio of negative literals is set at a certain range with a predefined step size. This approach is taken into account to avoid any repetitive distribution of the negative literals. Specific amount of negative literals in the proposed Weighted Systematic 2 Satisfiability will examine different complexity of the neuron connections in DHNN.

Methodology and Limitation 2

The proposed Weighted Systematic 2 Satisfiability with DHNN requires effective and efficient learning phase as the number of neurons increased. After formulating the proposed Weighted Systematic 2 Satisfiability, each variable in the clause will represents the neuron in DHNN. The strength of the connection between the neuron denoted as synaptic weight will be computed using Wan Abdullah method (Abdullah, 1992). This can be achieved by computing the cost function that is associated with the proposed Weighted Systematic 2 Satisfiability and was compared with the Lyapunov energy function. In this thesis, the proposed DHNN incorporates with Weighted Systematic 2 Satisfiability will be represented in two phases which are learning and retrieval phase. To assess the compatibility of the proposed network, two perspectives was introduced. First, the learning phase of DHNN is required to ensure at least one satisfied interpretation is obtained. Notably, Election Algorithm (EA) was implemented as the learning algorithm for finding the satisfied interpretation as supported by Bazuhair *et al.* (2021). By obtaining satisfied interpretation that leads to zero cost function, the synaptic weight can be obtained effectively and stored as content addressable memory. In this context, learning error will be introduced to measure the distance between the ideal neuron fitness with the current neuron fitness. Secondly, the quality of the final neuron states produced in the retrieval phase of DHNN is measured in terms of energy profile and neuron variation. The energy profile of the retrieved final neuron states can be measured by examining the difference between the global minimum energy and local minimum energy. Worth mentioning that, the neuron variation is evaluated by capitalizing the formulation of similarity index.

While this methodology can be an optimal approach to pattern reconstruction, there are a few limitations to ensure the reproducibility of the objective in this thesis. First, the variable in each clause does not contain redundant variable and must be expressed in CNF. Following to the WA method, if both structural components are considered, this will affect the performance of the learning phase in DHNN which leads to non−zero cost function. Additionally, the smallest number of literals that can be initiated is ten. This approach was taken into consideration to ensure that at least one negative literal exists in the proposed logical rule. Third, the neuron states will be represented in bipolar forms of 1 and -1. According to Stern and Shea-Brown (2020), the dynamics of ANN that relied on Lyapunov energy function are not suitable to represent information in the form of binary (1 and 0). This is due to the possible elimination of important coefficients in the energy function. In this case, when the states are zero, the minimum energy of the network will always be zero. Hence, the actual minimum energy of the network cannot be determined. Fourth, the neuron update in the retrieval phase of DHNN will go through the local field computation that operates based on the Ising spin model by Sherrington and Kirkpatrick (1975). Notably, the wrong updating rule will reduce the effect of the synaptic weight obtained from the learning phase of DHNN. Hence, this will be difficult to avoid local optimal solutions.

Methodology and Limitation 3

In general, the investigation of multi-objective function in the retrieval phase of DHNN-SAT models is relatively new. According to Cuéllar *et al.* (2009), multi-objective function give important advantages in optimization of ANN. This will force the search to return a set of optimal networks instead of a single one. In this thesis, the proposed multi-objective function during the retrieval phase of DHNN is responsible to produce superior final neuron states that possess high fitness, diversity, and dissimilarity. The proposed multi-objective function in DHNN with Weighted Systematic 2 Satisfiability requires an effective and efficient retrieval phase as the number of neurons increases. To solve this, this thesis proposed a new retrieval algorithm, namely Binary Clonal Selection Algorithm that will locate the superior final neuron states. High fitness corresponds to the final neuron state that achieves a global minimum solution. This particular objective function is essential to ensure correct synaptic weight values are used during the computation of the local field. Subsequently, high diversity is attained based on the highest distribution of negative neuron states. As for diversity, this objective function is to promote more variations of final neuron states in terms of negativity. Additionally, high dissimilarity coefficient value indicates a significant difference between the superior final neuron state and initial final neuron state. Initial final neuron state correspond to the final neuron state before the implementation of Binary Clonal Selection Algorithm. The multi-objective functions are crucial to improve quality of the final neuron states produce by DHNN. Final neuron state that achieved the multi-objective functions will be categorized as superior final neuron state. The production of superior final neuron state will be done by the proposed Binary Clonal Selection Algorithm using several global and local search operators. Notably, each superior final neuron state form induced logic in the logic mining model.

Despite this methodology can be an optimal approach to ensure the optimality of the retrieval phase of DHNN, there are several limitations to ensure the reproducibility of the objective in this thesis. Although multi-objective function commonly associated with Pareto optimality, this thesis will disregard this concept. This is because the main goal in the retrieval phase of DHNN is to obtain optimal final neuron state. If there is any trade-offs between fitness, diversity, and dissimilarity, the attained final neuron state can only satisfy one or the other objective function which is the contrary to the aim of the thesis objective. Secondly, the fitness function does not considering local minimum solution because local minimum solution will distrupt the diversity of the final neuron state (Karim *et al.*, 2021). Lastly, for the diversity function, the negativity in the final neuron states cannot be less than 50% of the total clauses because the impact of the negation cannot be seen (Karim *et al.*, 2022).

Methodology and Limitation 4

The main weakness in most existing logic mining models is the lack of consideration for both negative and positive instances in a dataset. Hence, it is difficult for the existing models to achieve good classification results that corresponds to high accuracy. In this thesis, an improved RA namely Weighted Systematic 2 Satisfiability Reverse Analysis with Binary Clonal Selection Algorithm was proposed to address this issue. First, the proposed logic mining model will formalize a pre-processing phase to execute the data cleaning, data preparation, and attribute selection process. The data cleaning process involved two steps, which are data imputation for missing values and clustering method by *k*-mean clustering to handle datasets with non-categorical or continuous entries and transform into bipolar forms of {1, -1}. After that, data preparation process proceeds with train test splitting and *k*-fold cross validation of all entries by a predefined ratio that is compatible with the existing works. Then, the pre-processing phase ends with random attribute selection. Secondly, the proposed logic mining model introduced a new method of entrenching the learning dataset into the proposed logical rule. Notably, the classification of binary confusion matrix are being evaluated based on the dependent attribute in the learning dataset to the outcome of the proposed logical rule after embedding learning dataset. Then, one selected performance metric will be calculated based on the binary confusion matrix classification. The logical rule with the highest value of selected performance metric will be renamed as the best logic. The selected best logic will be learned by the learning and retrieval phase of DHNN. This approach was taken into consideration to ensure that the generated best logic is able to represent the majority patterns of the dataset. Subsequently, induced logic with the highest accuracy will be considered as the best induced logic. The proposed logic mining model will be measured and investigated with 20 repository datasets retrieved from reputable databases. The proposed logic mining model will be compared with several existing logic mining models based on accuracy, sensitivity, specificity, negative predictive value, and Matthew's correlation coefficient metrics.

While this methodology can be an optimal approach to extract information from real-life datasets, there are several limitations to ensure the reproducibility of the objective in this thesis. First and foremost, the use of *k*-mean clustering is not applicable when the data is categorical or discrete in nature. Direct transformation to bipolar forms will be executed. For fair comparison and reproducibility purposes, the similar train test splitting, *k*-fold cross validation ratio, and attribute selection are applied to all models and set as constants. Last but not least, repository datasets acquired for this thesis were mainly retrieved from University of California Irvine (UCI) and Kaggle Machine Learning Repository. Using these retrieved datasets, the proposed logic mining model can be compared with other existing logic mining models in the field of feature selection method.

<u>Methodology and Limitation 5</u>

Generally, the investigation of multiple best logic in the logic mining models is relatively new. Thus, it is important to appropriately formulate adaptive multiple best logic that can optimally learn the behaviour of the dataset. In this thesis, multiple best logic was proposed in the learning phase of the logic mining model to produce multiple best induced logic. The new logic mining model is an improved logic mining model from Objective 4 namely Weighted Systematic 2 Satisfiability Modified Reverse Analysis. In this approach, multiple best logic are proposed whereby each best logic is correspond to different performance metrics. Notably, the proposed multiple best logic will correspond to distinguish CAM of DHNN. In this context, the proposed logic mining model possessed multi-units of CAM that proffers multi-units of local field computation which leads to multiple best induced logic (Alway *et al.*, 2023). This approach will help the proposed logic mining model to widen the search space and locate non-overfitting induced logic. On the other hand, unsupervised attribute selection method known as topological data analysis was proposed in the pre-processing phase of the logic mining model. The main goal of this proposed method is to filter the attributes that have dissimilar behaviour to the dataset.

Despite this methodology can be an optimal approach to ensure the optimality of the logic mining model, there are several limitations to ensure the reproducibility of the objective in this thesis. Firstly, the selected performance metrics for respective best logic must be distinct and do not have the same insight. For instance, metric accuracy and F1 score share the same insight where both of the metrics represent the classification of true positive and negative outcome (Chicco & Jurman, 2020). Similar metrics will lead to redundant CAM, which is prone to repetitive production of final neuron states. Consequently, there will be no differences in the pattern of the induced logic (Alway *et al.*, 2023). In relation to the discussed problem statements, objectives, and methodology, Figure 1.1 describes the overall methodologies discussed in this thesis.

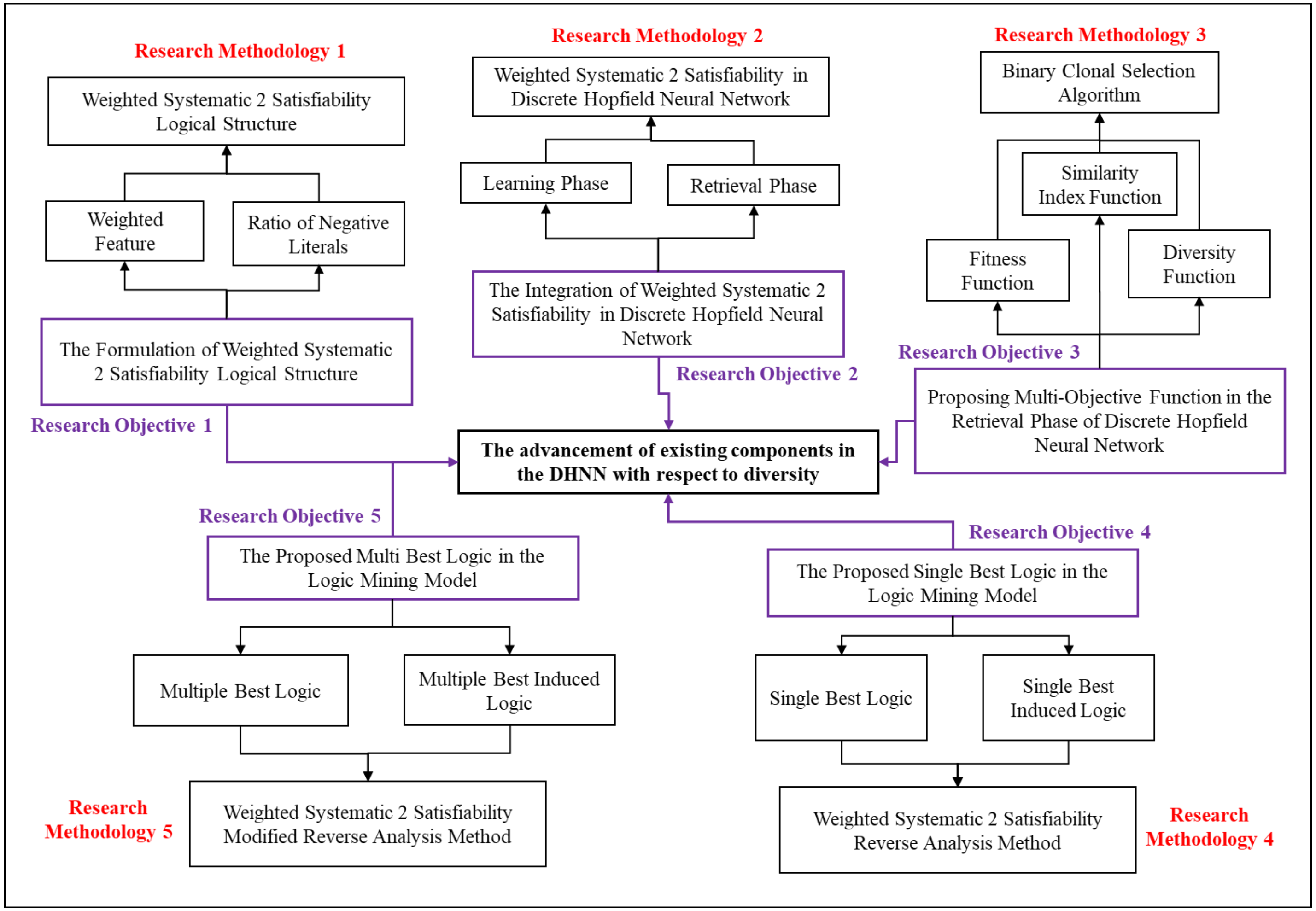

Figure 1.1 Flowchart of overall methodologies discussed in this thesis.