Contributors of Myopia using Major First Order Satisfiability Reverse Analysis

Nurul Huda Ahmad Rusli[1, a], Nurin Kamalin Mohd Fadil[1, b], Nurin Hazwani Hamidi[1, c], Farisya Husna Mansor[1, d], Muhammad Ameer Azammudin[1, e] and Nur Ezlin Zamri[1, f*]

[1]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, Serdang, Selangor 43400 UPM, Malaysia

[a]214520@student.upm.edu.my, [b]214516@student.upm.edu.my, [c]214615@student.upm.edu.my, [d]215191@student.upm.edu.my, [e]200238@student.upm.edu.my, [f]ezlinzamri@upm.edu.my

Abstract. Myopia is a growing concern for vision health, particularly among children and adolescents. The growing prevalence of myopia calls for a systematic approach to identify the key contributing factors behind its progression. This study presents an artificial intelligence with logic mining technique to identify the major contributors to myopia using the Major First Order Satisfiability Reverse Analysis (MFOSRA) model. The model incorporates several core components, including a Discrete Hopfield Neural Network, Exhaustive Search Algorithm as the training algorithm, Major First Order Satisfiability as the logical rule structure, a feature selection method based on the Pearson Correlation and a Reverse Analysis approach. The experiments were carried out using a health dataset related to myopia that included genetic, environmental and lifestyle indicators. The MFOSRA model is capable of extracting optimal logic patterns from the dataset in the form of interpretable rules. The induced logic highlights key contributors to myopia such as screen exposure time, lack of outdoor activities and family history. The results show that the proposed method providing high level of accuracy (91.13%). This demonstrates that the logic mining approach can extract optimum patterns from a dataset in the form of best induced logic. The

A. Akhyar et al. (eds.), *Proceedings of the International Conference on Research, Innovation, Sustainability, and Educations (IRISECON 2026)*, Atlantis Advances in Applied Sciences 5,
https://doi.org/10.2991/978-94-6239-745-3_29

resulting logic consists of composite measures of near-work activities, spherical equivalent refraction, parental myopia status, anterior chamber depth, the year the subject entered the study, time spent in sports activities and reading for pleasure. These findings indicate that this approach valuable for health researchers, patient guardians and medical professionals.

Keywords: Data Mining, Satisfiability, Reverse Analysis, Artificial Intelligence, Neural Networks.

1 Introduction

Myopia commonly known as nearsightedness is a refractive error in which distant objects appear blurred while close objects remain clear due to the eye being unable to correctly focus light on the retina [1]. It is the most common ocular disorder worldwide and the leading causes of visual impairment especially among children and young adults. The prevalence of myopia has been rising globally with projections estimating that half of the world population may be myopic by 2050 unless effective preventive strategies are implemented [1]. Myopia typically results from excessive axial elongation of the eyeball which causes light to focus in front of the retina rather than directly on it and leading to blurred distance vision [2].

The development and progression of myopia are influenced by a complex interplay of genetic and environmental factors. Key environmental contributors include prolonged near work such as reading and screen use, and reduced time spent outdoors which affects retinal signaling pathways involved in eye growth regulation [3]. The pathophysiology involves disruption of the emmetropization process which is a feedback mechanism that normally guides the eye to grow to the correct length for clear vision [4]. Visual signals particularly retinal image quality and defocus, regulate ocular growth through biochemical and structural changes in the sclera and choroid leading to axial elongation when dysregulated. Understanding these mechanisms is critical for developing interventions to control myopia progression and reduce associated risks such as retinal detachment, glaucoma and cataracts [5].

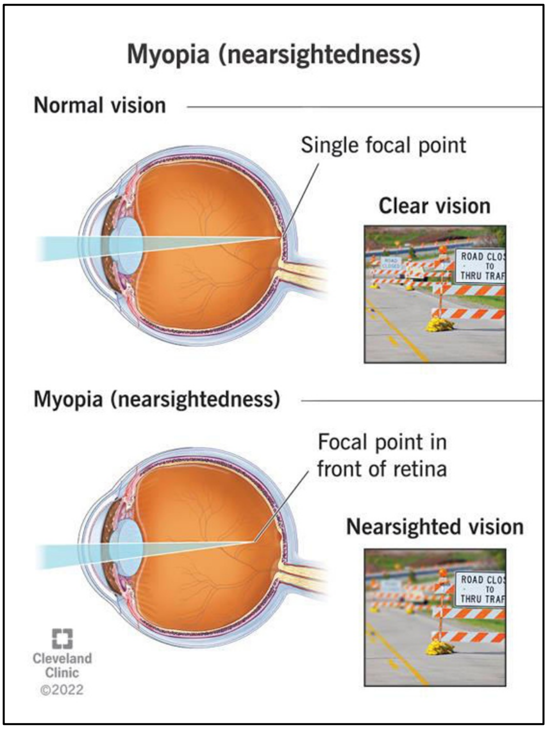

Fig. 1: Focus that happens in front of the retina of eye instead of at the retina results in myopia (nearsightedness)

Logic mining is a discrete classification technique that enables computational models to analyze complex datasets, identify underlying patterns and support decision-making processes. It is particularly useful for uncovering relationships in multifactorial problems such as the contributors to myopia where genetic and environmental factors interact in complex ways. Artificial Neural Networks (ANN) are computational models inspired by the architecture of the human brain. The brain consists of interconnected neurons that process information, recognize patterns and learn from data [6]. Similarly, ANN comprises networks of artificial neurons (nodes) connected by weighted links called synaptic weights. These weights determine the strength and influence of the connections between neurons and are stored in the associative memory of the network [6]. During the retrieval phase, synaptic weights are recalled to interpret new data. ANN excel at solving complex optimization problems and can extract meaningful insights from data with minimal human intervention making them powerful tools in logic mining across various domains. A prominent variant of ANN used in logic mining is the Discrete Hopfield

Neural Network (DHNN) introduced by Hopfield and Tank. DHNN consist of fully interconnected neurons without hidden layers where each neuron operates in a bipolar state (±1) [7]. The synaptic weights in DHNN form the basis of Content Addressable Memory (CAM) which enables the network to retrieve stored patterns by searching for the closest match to an input pattern [7]. This associative memory capability allows DHNN to perform efficient pattern recognition and optimization. In the logic mining process, the behavior of DHNN can be guided by implementing logical rules during the logic phase. These rules help the network model complex logical relationships within the data and facilitating reverse analysis to extract interpretable patterns [8, 9]. By applying Major First Order Satisfiability Reverse Analysis within this framework, it becomes possible to systematically identify and analyze the key contributors to myopia and revealing the interactions among various risk factors such as screen exposure, outdoor activity, and genetic predisposition. The significance of this model lies in its ability to combine the pattern recognition power of neural networks with rule based logic inference enabling explainable artificial intelligence for domains that require transparency and interpretability. In the context of myopia research, this hybrid framework allows researchers to go beyond black-box prediction and instead identify specific contributing variables and the logical interactions. This makes DHNN based logic mining a valuable tool for clinical decision support and public health strategy development.

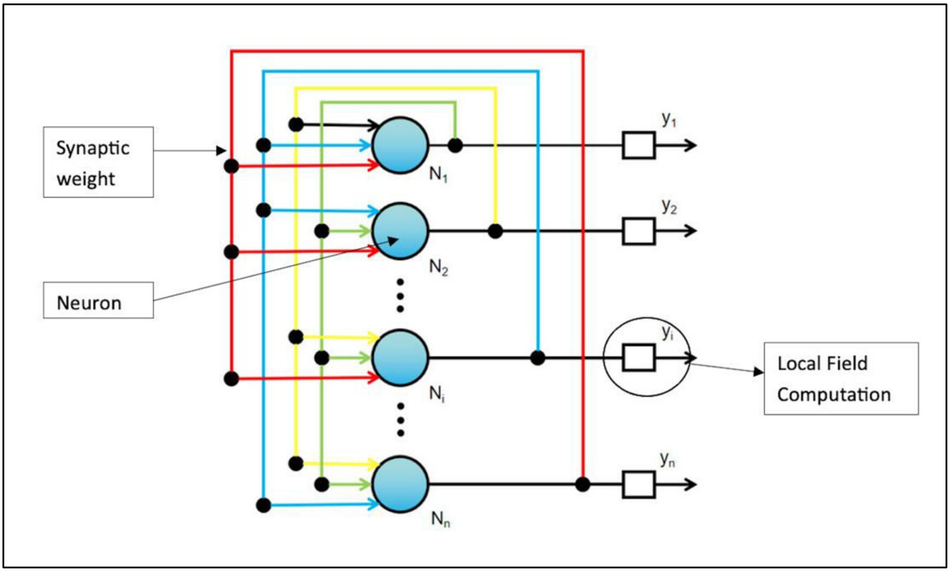

Fig. 2: Framework of DHNN

Myopia, also known as nearsightedness is a refractive error characterized by the inability to see distant objects clearly. It is caused primarily by the excessive axial elongation of the eyeball resulting in light rays focusing in front of the retina instead of directly on it. Myopia is one of the most common visual disorders globally and has seen a significant increase in prevalence in recent years. Research suggests that by the year 2050, nearly half of the global population may be affected by myopia making it a major public health concern. Early diagnosis and continuous monitoring of myopia progression are essential in preventing severe visual impairment and associated complications such as retinal detachment, glaucoma and macular degeneration. As a result, various diagnostic technologies have emerged many incorporate artificial intelligence (AI) to enhance the accuracy and efficiency of diagnosis. These AI models primarily built using machine learning (ML) and deep learning (DL) techniques are capable of processing large volumes of clinical data to identify hidden patterns and risk factors contributing to myopia. These include environmental influences such as screen exposure and near work as well as genetic predispositions.

However, current AI models in the domain of myopia detection and progression analysis exhibit notable limitations. Li et al. [10] reviewed several machine learning-based approaches applied to myopia research and found that while these models achieved classification accuracy above 50%, this level of performance is insufficient for clinical application. Moreover, these models often operate as black boxes producing predictions without offering interpretable explanations. This lack of transparency reduces the clinical trust and applicability of AI outputs especially in complex multifactorial conditions like myopia where understanding the causal relationships between factors is essential. Another pressing challenge is that these conventional models focus primarily on classification or prediction while ignoring the interpretability and logical reasoning required to understand why certain patterns lead to specific outcomes. There is an evident gap in leveraging logical modeling techniques that can capture complex relationships between contributing factors in an interpretable and systematic manner.

To address this gap, this research proposes an alternative logic mining approach based on a new logical rule Major First Order Satisfiability (MFOS), integrated within a Discrete Hopfield Neural Network (DHNN) framework. Unlike conventional deep learning models, the proposed logic mining model emphasizes transparency, interpretability and structure. The model is designed to uncover significant and logically coherent patterns in myopia datasets especially those involving multi factorial interactions among environmental and genetic factors. By using MFOSRA and DHNN, this study introduces a language based AI framework capable of performing reverse analysis which is a method for tracing backward from observed outcomes to infer the underlying logical rules or causes. This reverse logic mining allows for not only the detection of patterns but also the explanation of these patterns in a way that can be validated and interpreted by domain experts. Furthermore, the model will be evaluated not just on traditional accuracy metrics but also on its ability to accumulate true positive and true negative outcomes. Moreover, it is also generating final induced logic that aligns with medical knowledge. This dual focus on predictive performance and interpretability represents a significant advancement over existing models.

1.1 Research Questions

The research questions of the recent research are as follows:

1. What form of alternative logic can provide greater flexibility in capturing complex and diverse patterns within myopia datasets?
2. How can the proposed model extract significant and interpretable logical patterns that explain the relationships between environmental and genetic risk factors and the progression of myopia?
3. Can the integration of Reverse Analysis within a Discrete Hopfield Neural Network effectively enhance classification performance by improving the detection of true positives and true negatives in myopia datasets?

1.2 Research Objectives

The objectives of the recent research are:

1. To propose an alternative logic mining model using a new logical rule namely major first order logic that are flexible and capable of capturing diverse and complex patterns in myopia datasets.
2. To analyze the myopia data in a way that reveals significant and interpretable patterns connecting risk factors to the onset and progression of myopia.
3. To evaluate the effectiveness of the proposed model with respect to the accumulated number of true positive and true negative of the proposed model and retrieved final induced logic.

1.3 Scope of Project

- Apply MFOSRA to a myopia dataset including reading duration, outdoor activity time and parental myopia history. Therefore, experimental validation is restricted to the myopia dataset only which may limit to other medical conditions.
- Train and test the model using a DHNN framework.
- Use Reverse Analysis to produce a logical rules from stable neuron states for clinical interpretability.

2 Literature Review

The use of artificial intelligence (AI), neural networks and symbolic reasoning in healthcare has grown particularly in domains like myopia where both prediction accuracy and interpretability are essential. Traditional models such as the Hopfield Neural Network (HNN) demonstrated early theoretical promise but lacked real-world applicability and explainability. Later efforts integrated logic into neural networks such as logic programming within HNN of Wan Abdullah [11] but these were limited to simple propositional logic and not scalable to complex medical datasets. Modern machine learning models like CNN, random forests and XGBoost have improved diagnostic accuracy but remain black-box systems offering minimal insight into how predictions are made. Recent attempts to enhance interpretability involve combining logic structures with Discrete Hopfield Neural Networks. These models typically rely on predefined logic and lack the ability to discover rules from data or support reverse analysis which an essential feature in clinical diagnostics for tracing outcomes from the causes. This review highlights a clear gap where the need for a logic-integrated neural model that operates on real-world data, offers interpretable rules and enables reverse analysis. This forms the basis for the development of the proposed MFOSRA model.

Hopfield [12] proposed a recurrent neural network based on an energy minimization principle. The neurons are binary (± 1) or continuous-valued units with symmetric weights and no self-loops. It is fully connected architecture where each neuron is connected to every other neuron. Neuron state updates are asynchronous and driven by minimizing a global energy function which capable of storing and recalling. The network converges to stable fixed points representing stored memory patterns and mostly applied to theoretical problems or small-scale pattern completion tasks. Nonetheless, no real-life datasets were used and experiments were mostly theoretical or used artificial binary patterns. The model lacks interpretability as it does not generate explainable logical rules or link data features to outcomes. It does not support reverse analysis which is reasoning backward from outcome to input factors. Wan Abdullah [11] introduced a method to perform logic programming using Hopfield Neural Networks known as the Wan Abdullah (WA) method. The logical rules especially Horn clauses are transformed into equivalent synaptic weight configurations in a Hopfield network. It maps the satisfaction of logical clauses to the minimization of an energy function and demonstrated that logic rules can be encoded into neural networks for reasoning and inference. The network settles into a state that represents a solution to the logical problem (HornSAT). However, the model is limited to propositional logic (HornSAT) and cannot accommodate first-order logic or continuous variables. Consequently, it is not suitable for real-world datasets particularly medical data such as myopia risk factors. Additionally, the absence of reverse analysis prevents analyzing the contributing factors resulting in limited interpretability for medical explanations.

Liu et al. [13] studied the use of AI in applications such as early detection, risk factor identification and prediction for myopia control. AI systems are built on large datasets (fundus images, OCT, behavioral surveys) with models including SVM, Random Forest, XGBoost, CNN, LSTM. The study resulting a high area under the curve (AUC) scores in detection (AUC > 0.9) and real-world validation in some studies. The gap is models are data-driven but largely remain black-box systems with limited interpretability. Most models lack logical rule explanation and cannot identify relationships behind predictions like why a child is at high risk. Additionally, no integration of logic mining or reverse analysis to trace myopia outcomes back to contributing factors as it focuses on forward prediction. Zhang et al. [14] proposed PMPred-AE is an advanced AI-based model developed to detect and interpret pathological myopia (PM). The study built upon an autoencoder deep learning architecture and it integrates convolutional neural networks and attention mechanisms to process retinal fundus images. The attention module highlights regions contributing to PM for offering visual interpretability. Moreover, the model achieved high accuracy and AUC in detecting PM with enhanced visual outputs offering interpretive insight into how predictions are made. While the model uses visual attention for interpretability, it still functions as a black box neural network. Consequently, the model lacks formal logical rule extraction resulting in limited interpretability. It primarily focuses on image-based diagnosis and does not incorporate symbolic reasoning or logical relationships between risk factors and PM. Furthermore, it does not support first-order satisfiability logic, reverse analysis or factor identification from outcomes back to inputs thereby limiting its effectiveness as an explainable AI approach in medical diagnostics.

Roslan et al. [7] introduced a logic neural framework that integrates Discrete Hopfield Neural Network (DHNN) with Major 3-Satisfiability (3-SAT) logic and enhanced it using an Election Algorithm (EA). The logical clauses are encoded into neural weights via satisfiability constraints and the DHNN evolves to minimize a energy function. Next, the EA helps optimize rule combinations across multiple objectives and demonstrated effectiveness in solving complete problems with strong convergence and stability. Therefore, it generated interpretable logical outcomes and solution states under constraints. However, the model focuses primarily on constraint satisfaction problems and not on medical or diagnostic applications like myopia predictions the logic rules are derived from known formulas and not learned from data. Thus no mechanism for discovering unknown contributing factors from datasets and it interpretable in logic terms but it lacks integration with medical reasoning or reverse analysis from outcomes to causes. Moreover no validation on imbalanced data and limiting clinical relevance. Other recent work by Paul and Cabarle [15] explored solving the SAT problem using spiking neural P systems with structural plasticity and pre computed resources. In this approach the SAT problem is represented within a biologically inspired neural system where neurons interact according to predefined rules and adapt the connectivity to satisfy logical constraints. The system demonstrates fast convergence and interpretable solutions for satisfiability problems highlighting the potential of neural based architectures for logic problem solving. However, this model has not been applied to real-world medical datasets or integrated with reverse analysis process to identify contributing factors which limits its direct applicability in myopia risk factor analysis [15]. Study by Alway et al. [16] applied 2-Satisfiability (2-SAT) logic to construct a logical neural network using the Discrete Hopfield Neural Network (DHNN). The logical clauses are transformed into synaptic weight matrices and the DHNN uses an energy function derived from logic violations and evolves toward a neuron stable state that satisfies the logic constraints. Consequently, it successfully solves 2-SAT logic problems with fast convergence and interpretable binary outcomes yet the model is not tested on real world datasets. Therefore, the focus is on solving logic satisfiability problems and not on discovering logical relationships from data. Next, the lacks reverse reasoning capability from outcome back to factors and no integration with first order logic limits for modeling complex relationships in domains like medical diagnostics.

This chapter reviewed previous studies on Hopfield Neural Networks, logic mining process and artificial intelligence approaches related to myopia analysis. Early works focused on theoretical models and logic mining but lacked medical data application and reverse analysis process. Recent AI of machine learning achieved high predictive accuracy in myopia detection however the models had limited interpretability where no induced logic produced. These limitations and gaps highlight the need for an interpretable logic mining process identify factors that influence myopia progression.

3 Methodology

This section presents the theoretical background for each component involved in the proposed logic mining approach. It begins with a general overview of MFOSRA including an implementation into DHNN. Finally, the section explores the concepts behind each core component and outlines the four phase structure which are the preliminary phase, logic phase, training phase and retrieval phase of the logic mining model implemented within the DHNN framework.

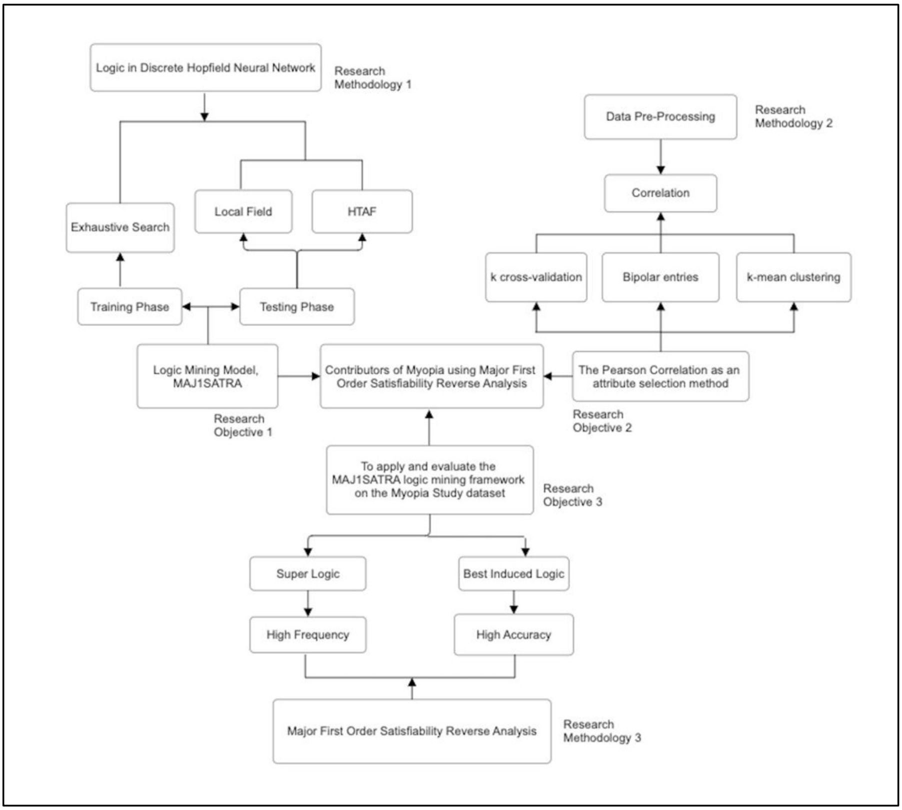

Fig. 3: Overview Methodology of MFOSRA

3.1 Major First Order Satisfiability Reverse Analysis (MFOSRA)

MFOSRA is a novel variant of non-systematic logic expressed in Conjunctive Normal Form (CNF). The logical structure is composed of a combination of 2SAT and first order clauses where the single clauses represent the majority within the logical string [17]. This dominance of single clauses allows the structure to offer a flexible logical representation that emphasizes single literal patterns. The logic is designed such that the number of single clauses must strictly exceed the combined number of 2SAT clauses to preserve the majority character. If this condition is not satisfied, the logical structure loses its MFOSRA identity [18].
The logic enables a wide variation of clause combinations while maintaining interpretability and neural network for logic mining applications.

Each logical clause is constructed using non redundant literals ensuring the accuracy of synaptic weight derivation during transformation into neural representations. The general formula of MFOSRA is shown in Eq. 1 representing the valid combinations of clause orders based on whether 2SAT components are present alongside the required single clauses dominance. Let k denote the clause order combination and let m and r represent the number of 2SAT and single clauses respectively. As a result, the model considers only 2SAT and single clauses to support the Major First Order Satisfiability framework.
The components of MFOSRA logic are:

- A set of variables $\mathrm{c}_1, \mathrm{c}_2, \ldots, \mathrm{c}_\mathrm{y}$
- A set of literals c_y or their negations $\neg\mathrm{c}_\mathrm{y}$
- A set of 2SAT clauses $C_1^{(2)}, C_2^{(2)}, \ldots, C_m^{(2)}$ and first order clauses $\mathrm{C}_1^{(1)}, \mathrm{C}_2^{(1)}, \ldots, \mathrm{C}_\mathrm{r}^{(1)}$ where $C_i^{(2)} = (d_i \vee e_i)$ and $\mathrm{C}_\mathrm{i}^{(1)} = \mathrm{f}_\mathrm{i}$ respectively.

Depending on the clause combinations of MFOSRA formulas are defined as follows:

$$\Phi_{\mathrm{MFOSRA}} = \bigwedge_{i=1}^{m} \mathrm{C}_\mathrm{i}^{(2)} \wedge \bigwedge_{i=1}^{r} \mathrm{C}_\mathrm{i}^{(1)}. \tag{1}$$

To ensure the dominance of first order by introducing the clause ratio β :

$$\beta = \frac{r}{m+r}, \ \beta \in (0.5, 1). \tag{2}$$

If $\beta \leq 0.5$ the majority condition is violated. If $\beta = 1$ the structure becomes a pure first order losing the diversity offered by mixed clause orders. The total number of clauses (TC) is defined to ensure consistent clause generation:

$$TC = m + r. \tag{3}$$

This allows consistent benchmarking and clause generation across different configurations of MFOSRA. The literals in all clauses are encoded using bipolar representation $\mathrm{c}_\mathrm{i}^* \in \{1, -1\}$ which supports neural energy-based learning [12]. The MFOSRA structure being rooted in the

MAJ2SAT design builds upon the concept of clause dominance but shifts emphasis to minimal and interpretable first-order logic [16]. This design distinguishes it from RAN3SAT [19] which allows randomness but mandates at least one 3SAT clause and from MAJ3SAT it favors complex clause structures. MFOSRA is suitable for interpretable logic mining and reverse analysis where the influence of single literal conditions needs to be understood [7].

Example MFOSRA induced logic:

- Example 1 ($\beta = 0.71, m = 2, r = 5$):

$$\phi_{MFOSRA} = (d_1 \vee \neg e_1) \wedge (\neg d_2 \vee \neg e_2) \wedge f_1 \wedge f_2 \wedge \neg f_3 \wedge f_4 \wedge f_5$$

- Example 2 ($\beta = 0.88, m = 1, r = 7$):

$$\phi_{MFOSRA} = (\neg d_1 \vee e_1) \wedge f_1 \wedge f_2 \wedge f_3 \wedge f_4 \wedge f_5 \wedge f_6 \wedge f_7$$

- Example 3 ($\beta = 0.75, m = 1, r = 3$):

$$\phi_{MFOSRA} = (\neg d_1 \vee e_1) \wedge f_1 \wedge \neg f_2 \wedge f_3$$

In summary, the ϕ_{MFOSRA} logic structure emphasizes interpretability, simplicity and flexibility making it ideal for fields requiring clear logical inference. Its dominance of first-order logic differentiates it from major second order, major third order and random k-SAT models like RAN3SAT. MFOSRA is especially suitable for logic mining in medical and explainable AI domains where transparent logic extraction is essential.

3.2 Discrete Hopfield Neural Network(DHNN) Framework

In this study, we propose the DHNN-MFOSRA model where the logical structure ϕ_{MFOSRA} comprises a majority of single-literal clauses ensuring simplicity and faster interpretability in logic induction. DHNN is a recurrent neural model known for its content addressable memory (CAM) and energy optimization. The DHNN operates without hidden layers and all neurons are fully connected with symmetric synaptic weights. The input flows directly to the output and neurons are updated asynchronously. This system uses bipolar encoding where +1 represents a TRUE logical value and -1 represents FALSE [20].

The aim of the learning phase is to compute synaptic weights such that the energy cost function reaches its minimum which is zero. The proposed model applies a logic string ϕ_{MFOSRA} where single clauses dominate can be expressed as:

$$\phi_{\mathrm{MFOSRA}} = x_1 \wedge \neg x_2 \wedge (x_3 \vee \neg x_4) \wedge x_5$$

The cost function for mixed clauses is computed as:

$$E_{\phi_{MFOSRA}} = \frac{1}{2}\sum_{j=1}^{r}(\prod_{i=1}^{1} q_i^{(1)}) + \frac{1}{4}\sum_{j=1}^{m}(\prod_{i=1}^{2} q_i^{(2)} + 1). \tag{4}$$

From Eq. 5, the r and m are the number of first order and 2SAT clauses respectively. The clause dominance rule is enforced by ensuring:

$$r > m. \tag{5}$$

This condition ensures the logical structure retains its MFOSRA identity. The synaptic weights are determined by minimizing the Lyapunov energy function:

$$H_{\phi_{MFOSRA}} = -\sum_{i=1}^{N} W_i^{(1)} S_i - \sum_{p \neq q} W_{pq}^{(2)} S_p S_q. \tag{6}$$

Where S_i is the state of neuron i and $W_{pq}^{(2)}$ represent 2 literal correlation even though it remain minority contributors [13].

In the testing phase, the neuron states are relaxed using a technique based on Sathasivam's relaxation model to ensure stability. The local field for neuron i is computed as:

$$h_i = \sum_{j=1}^{N} W_{ij}^{(2)} S_j + W_i^{(1)}. \tag{7}$$

3.3 Preprocessing and Feature Selection

The neuron state is updated using the hyperbolic tangent activation function:

$$S_i = \begin{cases} 1, \; if \; tanh(h_i) \geq 0 \\ -1, otherwise \end{cases}. \tag{8}$$

After stabilization, the final solution is filtered based on minimum energy:

$$\left| H_{\phi_{MFOSRA}} - H_{\phi_{min}} \right| < tol. \tag{9}$$

Where tol is the convergence threshold [7]. Therefore, the final classification of the induced logic Q_i^B is followed by:

$$Q_i^B = \begin{cases} Q_i^B, \; if \mid H_{\phi_{MFOSRA}} - H_{\phi_{min}} \mid < tol, \\ 0, otherwise \end{cases}. \tag{10}$$

In the preprocessing phase each attribute in the dataset is represented as a neuron S_i where $i = 1, 2, \ldots, N$. Based on the Wan Abdullah [11] method, all neuron states are converted into bipolar form such that:

$$S_i \in \{-1, 1\}. \tag{11}$$

This transformation allows DHNN to work with binary logic states, representing false and true respectively.
To ensure that only relevant attributes are used, Pearson correlation analysis is applied to measure the strength of the relationship between each attribute and the target class. Attributes are selected based on correlation values that are closest to 1 either positively or negatively. This ensures that only the most influential features are used for logic clause construction [21]. From the correlation results, the top 10 attributes are selected. Nine of them serve as input neurons while the tenth is used as the output class which is myopic. The value of Pearson Correlation Method use the formula of Eq. 12:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \cdot \sqrt{\sum (y_i - \bar{y})^2}} \tag{12}$$

where x_i and y_i are the individual data points and $\bar{x}$ and $\bar{y}$ are the mean of x and y.
During training, multiple candidate logical structures are created using combinations of selected attributes with strict enforcement of a major first order logic design. The structure is constrained such that the number of first order clauses r must be greater than the 2SAT clauses m. This clause-dominance condition ensures that the structure preserves the MFOSRA identity and promotes logical interpretability through minimalism.
Unlike previous methods that use random clause generation, this study applies an exhaustive clause enumeration process with a reverse analysis technique. Candidate logical rules are built from simple to complex starting with first order clauses and only adding higher-order clauses when necessary for improving performance. This reverse construction supports a more transparent connection between features and the output class.
The objective is to identify the best MFOSRA rule that maximises classification performance. This is done by selecting the structure with the highest total of true positives (TP) and true negatives (TN):

$$\mathrm{Q_{best}} = \max[\sum_{\mathrm{i}=1}^{n} s_i] \text{ where } \mathrm{s_i} \in \{1 \text{ (TP or TN)}\}. \tag{13}$$

Once the best rule is identified, it is embedded into the DHNN model. The synaptic weight values $\mathrm{W_{ij}^{(2)}}$ and $\mathrm{W_{i}^{(1)}}$ are computed by minimising the difference between the derived logic cost function and the Lyapunov energy function of the Hopfield network [18]. These weights are stored in the network's CAM (Content Addressable Memory).
In the testing phase, the trained weights are used to evaluate new input data. The local field for each neuron is computed and neuron states are updated using the hyperbolic tangent activation function (HTAF). The final neuron states are converted into the induced logic output $\mathrm{Q_i^A}$ which is then compared with the actual test labels $\mathrm{Q_i^{test}}$ to assess model performance. The main evaluation metrics used are:

- True Positives (TP) and True Negatives (TN)
- False Positives (FP) and False Negatives (FN)

To study the effect of different clause configurations within the MFOSRA framework, the model is tested using logical structures composed primarily of first order clauses supplemented by a small number of 2SAT clauses. For example, the following logical expression illustrates a typical clause configuration:

$$x_1 \wedge \neg x_3 \wedge x_4 \wedge \neg x_6 \wedge x_9 \wedge (x_2 \vee \neg x_5) \wedge (\neg x_7 \vee x_8). \tag{14}$$

In this structure, first order clauses dominate, ensuring the logical identity of MFOSRA is preserved. This configuration is intended to evaluate whether the simplicity and interpretability of first-order logic can still achieve competitive classification performance while maintaining logical traceability through reverse analysis [7]. Each clause configuration uses the same attribute set and model architecture. Performance metrics such as accuracy, precision, specificity, negative predictive value (NPV), Matthews Correlation Coefficient (MCC) and Bookmaker Informedness (BM) are computed for each. The goal is to determine which clause structure achieves the best trade-off between logic simplicity and predictive accuracy in the myopia classification task. The entire testing process is repeated for all ten logical structures generated during training. Final performance is reported based on the structure that most accurately matches the test data and achieves the highest classification score.

3.4 Dataset and Attribute Description

The dataset selected for this analysis is the Myopia Study dataset, hosted on the Kaggle machine learning repository https://www.kaggle.com/datasets/mscgeorges/myopia-study originally derived from the Orinda Longitudinal Study of Myopia (OLSM). The dataset contains no missing values and all attributes are complete across subjects. In terms of class distribution the myopia class is imbalanced with a higher number of subjects labeled as 0 (non-myopic) compared to 1 (myopic). This imbalance reflects the underlying demographic characteristics of the United States population where the prevalence of myopia is comparatively lower than in many East Asian countries [22]. Consequently, fewer subjects in the dataset are classified as myopic which is consistent with epidemiological observations in the U.S. population.

Two primary criteria guided the dataset selection for this logic-based modeling approach. First, the dataset contains 618 instances which satisfies the minimum requirement for statistical machine learning training with no missing values. As discussed in Zhou et al. [23], a larger dataset helps improve model accuracy and generalisability by reducing the risk of overfitting and enabling more representative sampling of the target population. Second, the dataset includes 18 attributes, offering a sufficient level of complexity while remaining manageable for logic mining. According to Nguyen et al. [24] and Li et al. [10], datasets with a moderate number of attributes which is between 10 and 20 have a good balance between feature interpretability. An excessive number of attributes can make pattern discovery difficult, as emphasized in Kumar and Singh [25], while too few may result limited rule expressiveness. The Myopia Study dataset falls within this optimal range, providing a suitable foundation for developing logic mining.

Prior to model training, the dataset undergoes a preprocessing step in which all variables are converted to bipolar format $(1 \text{ and } -1)$ using the Statistical Package for the Social Sciences (SPSS). For quantitative attributes, k-means clustering is used to determine a mean-based threshold for binarization [21]. To ensure rigorous evaluation and fair comparison with existing logic mining methods, the study applies several training testing splits where 60%–40%, 70%–30%, 80%–20% and 90%–10%. In addition, 5-fold cross-validation is performed to assess model consistency and reduce variance in the evaluation. This multi-split and cross-validation approach ensures the proposed logic model is thoroughly tested and offers reliable performance across varying data partitions.

This clinical research project was designed to investigate the phenotypic, behavioral and genetic factors contributing to the development of myopia in children. The dataset encompassing a wide range of demographic, biological, environmental and hereditary features. The target attribute is MYOPIC which is a binary indicator where 1 represents the presence of myopia and 0 indicates no myopia during the first five years of follow-up. The dataset includes several key demographic indicators such as AGE, representing the child's age at first visit and GENDER, coded as 1 for male and 0 for female. Ocular biological measurements include spherical equivalent refractive error in diopters, axial length, anterior chamber depth, lens thickness and vitreous chamber depth with all measured in millimeters. Environmental and behavioral variables capture lifestyle influences including hours per week spent in sports or outdoor activities, hours spent reading, computer or video game, studying or reading for school, hours of television watched and composite score representing total near-work activity. In addition, the dataset includes the year each subject joined the study and a unique identifier for each participant. Finally, it provide parental myopia history where 1 denotes that the respective parent was myopic and 0 otherwise. This diverse set of variables offers a comprehensive basis for investigating myopia risk and supports logic-based classification models through rich and interpretable features.

Table 1: Information of dataset.

Dataset Name	Instances	Attributes	Attribute Type	Field Area
Myopia Study	618	18	Mixed	Medical / Vision Science

Table 2: List of detailed attributes for the Myopia Study dataset.

No.	Attributes Names	Detail of Attributes
1	ID	Incremental ID assigned to each participant
2	STUDYYEAR	Year the subject entered the study
3	MYOPIC	Indicates whether the subject developed myopia within the first five years (1 = Yes, 0 = No)
4	AGE	Age at the first clinical visit (in years)
5	GENDER	Gender of the subject (1 = Male, 0 = Female)
6	SPHEQ	Spherical Equivalent Refraction (in diopters)
7	AL	Axial Length of the eye (in millimeters)
8	ACD	Anterior Chamber Depth (in millimeters)
9	LT	Lens Thickness (in millimeters)
10	VCD	Vitreous Chamber Depth (in millimeters)
11	SPORTHR	Time spent in sports or outdoor activities (in hours per week)
12	READHR	Time spent reading for pleasure (in hours per week)

13	COMPHR	Time spent using computer or playing video games (in hours per week)
14	STUDYHR	Time spent reading or studying for school assignments (in hours per week)
15	TVHR	Time spent watching television (in hours per week)
16	DIOPTERHR	Composite measure of near-work activities (in hours per week)
17	MOMMY	Indicates if the subject's mother is myopic (1 = Yes, 0 = No)
18	DADMY	Indicates if the subject's father is myopic (1 = Yes, 0 = No)

Table 3: List of Binary Attributes for Myopia Dataset.

No.	**Attributes Names**	**Data Type**	**Description**
1	ID	Numerical	Incremental identifier for each participant.
2	STUDYYEAR	Numerical	Mean = 1992.36; -1: $\geq$ 1992.36, 1: < 1992.36
3	MYOPIC	Categorical	Target variable; Myopic = 1, Not Myopic = -1
4	AGE	Numerical	Mean = 6.30; -1: $\geq$ 6.30, 1: < 6.30
5	GENDER	Categorical	Male = 1, Female = -1
6	SPHEQ	Numerical	Mean = 0.801; -1: $\geq$ 0.801, 1: < 0.801
7	AL	Numerical	Mean = 22.50; -1: $\geq$ 22.50, 1: < 22.50
8	ACD	Numerical	Mean = 3.5786; -1: $\geq$ 3.5786, 1: < 3.5786
9	LT	Numerical	Mean = 3.5415; -1: $\geq$ 3.5415, 1: < 3.5415
10	VCD	Numerical	Mean = 15.38; -1: $\geq$ 15.38, 1: < 15.38
11	SPORTHR	Numerical	Mean = 11.95; -1: $\geq$ 11.95, 1: < 11.95
12	READHR	Numerical	Mean = 2.7961; -1: $\geq$ 2.7961, 1: < 2.7961
13	COMPHR	Numerical	Mean = 2.1052; -1: $\geq$ 2.1052, 1: < 2.1052
14	STUDYHR	Numerical	Mean = 1.4903; -1: $\geq$ 1.4903, 1: < 1.4903
15	TVHR	Numerical	Mean = 8.9482; -1: $\geq$ 8.9482, 1: < 8.9482
16	DIOPTERHR	Numerical	Mean = 26.0178; -1: $\geq$ 26.0178, 1: < 26.0178
17	MOMMY	Categorical	Myopic = 1, Not Myopic = -1
18	DADMY	Categorical	Categorical & Myopic = 1, Not Myopic = -1

3.5 Performance Evaluation Metrics

There are seven classification metrics used to measure the performance of the model which is Accuracy (ACC), Precision (PRE), Sensitivity (SEN), Specificity (SPE), Matthew correlation coefficient (MCC), Negative predictive value (NPV) and Bookmaker Informedness (BM). These metrics are calculated based on values from the confusion matrix, as shown in Fig. 4. True positive (TP) and true negative (TN) instances are the number of correctly predicted positive and negative cases respectively. False positives (FP) occur when the model predicts a positive outcome for a case that is actually negative whereas false negatives (FN) occur when the model predicts a negative outcome for a case that is actually positive [26].

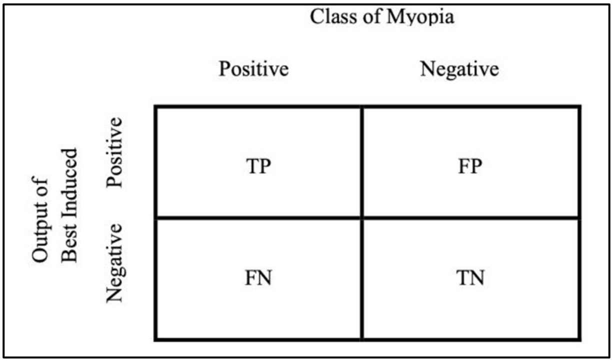

Fig. 4: Confusion Matrix used for metric calculation

- Accuracy (ACC): Evaluation metric where the ratio of the number correctly predicted to the overall behaviors [27].

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}. \quad (15)$$

- Precision (PRE): PRE is used to assess the model predictive capability. PRE is the exactness with which the model predicts a positive outcome and the proportion of positive outcomes [28].

$$\mathrm{PRE} = \frac{TP}{TP + FP}. \quad (16)$$

- Sensitivity (SEN): SEN evaluates the model ability to correctly identify actual positive cases. It reflects how well the model captures true positives among all actual positives [28].

$$\text{SEN} = \frac{TP}{TP + FN}. \tag{17}$$

- Specificity (SPE): Ratio of TN to total TN and FP is displayed using SPE values. As a result, SPE evaluates how well the model can predict the results [29].

$$\text{SPE} = \frac{TN}{TN + FP}. \tag{18}$$

- Matthews Correlation Coefficient (MCC): An evaluation metric that works for imbalanced data sets [30].

$$\text{MCC} = \frac{(TP \times TN)-(FP \times FN)}{\sqrt{\{((TP+FP)(TP+FN)(TN+FP)(TN+FN))\}}}. \tag{19}$$

- Negative Predictive Value (NPV): A crucial metric in medical testing and diagnostics. It represents the probability that an individual who tests negative for a disease or condition does not have that disease or condition [31].

$$\text{NPV} = \frac{TN}{TN + FP}. \tag{20}$$

- Bookmaker Informedness (BM): It measures how much better the classifier is at predicting than chance, given that it provides a specific prediction [32].

$$\text{BM} = SENSITIVITY + SPECIFICITY - 1. \tag{21}$$

Next, the error metrics for simulated data are used to evaluate model performance:

- Mean Absolute Error (MAE): MAE calculates the average of the absolute differences between predicted and actual values over all n data points.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - x_i|. \tag{22}$$

- Ratio Global:

$$R_G = \frac{1}{a \times p \times NT}\sum_{i=1}^{b} G_i. \tag{23}$$

- Variation Ratio:

$$R_{TV} = \frac{1}{a \times p \times NT}\sum_{i=1}^{b} F_i. \tag{24}$$

The experiments conducted from a logic mining perspective have been thoroughly explored. For the data collection and preparation, the software used are Microsoft Excel and SPSS Statistics while the Dev C++ Version 5.11 is used to run the simulations since the computer program interface is user-friendly while using open-source software. This section primary objective is to assess MFOSRA induced logic performance. In DHNN, MFOSRA will use Exhaustive Search (ES) as a training method. Thus, the following list of parameters used in the experiment is shown.

Table 4: List of parameters in MFOSRA

Parameter	**Parameter Value**
Neuron Combination	100
Number of Trial, ν	100
Number of Learning, ω	100
Attribute selection	Pearson Correlation Analysis
Order of clauses, k	$k = 2,1$
Alpha, α	(0.5,1)
K-fold cross validation	5
Maximum permutation	100
Tolerance value, tol	0.001
Activation function	Hyperbolic Tangent Activation Function (HTAF)

Type of selection	Exhaustive Search

4 Result and Discussion

The statistical test known as the correlation analysis test will be utilized to identify the qualities that used in the study as compared to choosing them at random. High Pearson Correlation either positive or negative values were preferred since they demonstrated that a high input distribution results in a positive dataset class [33]. The ideal attributes gained from correlation analysis will assure the validity of the synaptic weight obtained during DHNN training, resulting in optimal induced logic in the retrieval phase [34]. The Pearson correlation coefficient is applied to identify the most relevant attributes that show strong relationships with the target output which is the myopic class [35].

4.1 Correlation Analysis and Attribute Selection

After selecting the top 10 attributes, the binary dataset is subjected to cross-validation to evaluate the model performance. The dataset is divided into different training and testing ratios which are 60% training and 40% testing, 70% training and 30% testing, 80% training and 20% testing, and 90% training and 10% testing with each configuration evaluated using 5-fold cross-validation. This approach ensures that the model is tested on multiple subsets of data which helps reduce bias and variance especially in smaller datasets [36]. K-fold cross-validation is chosen because it provides a reliable evaluation by averaging the performance across different data splits. This method helps avoid over fitting and gives a clearer picture of the model generalization ability across different proportions of training and testing data [36]. Fig. 5 presents a heat map that illustrates the correlation values of the attributes.

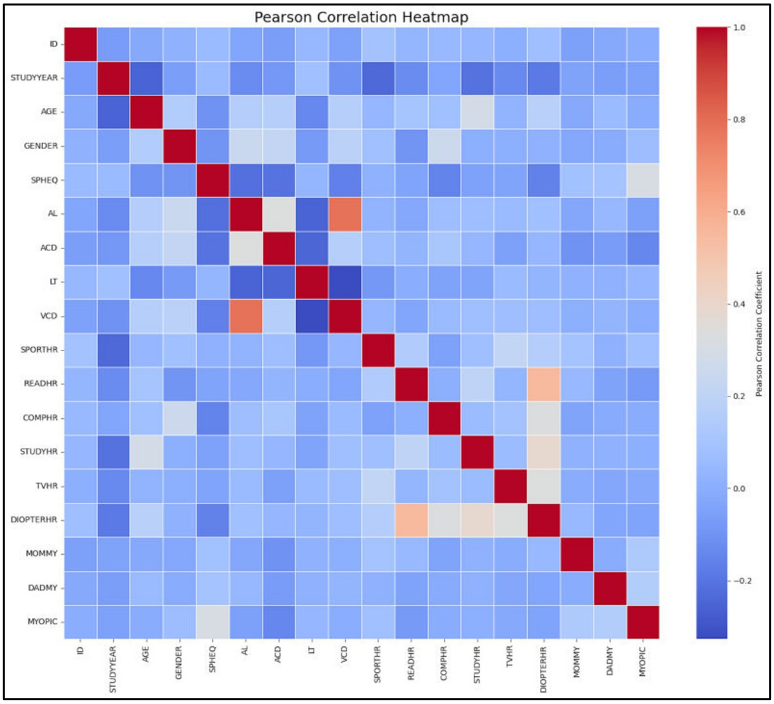

Fig. 5: Correlation heatmap of selected features with Myopia Indicator

The performance of the Major First Order model is characterized by a stable training Mean Absolute Error (MAE) that contrasts with fluctuating testing MAE driven by the non-deterministic nature of local field computations and random neuron activation. While the integration of Content Addressable Memory (CAM) provides a robust framework for weight stabilization, it cannot fully neutralize the randomness-induced fluctuations encountered during the processing of unseen myopia data. Energy analysis confirms the model effectiveness as it consistently achieves the global minimal solution where the energy difference remains within the required tolerance even when local minima occasionally deviate due to the stochastic nature of the search space. This behavior is further reflected in the total neuron variation where fluctuations highlight the necessity of an AI-driven approach to differentiate complex logical structures ultimately fulfilling the research objective to capture diverse risk factor patterns with high accuracy.

4.2 Training and Testing Mean Absolute Error (MAE) Analysis

Training Mean Absolute Error (MAE):

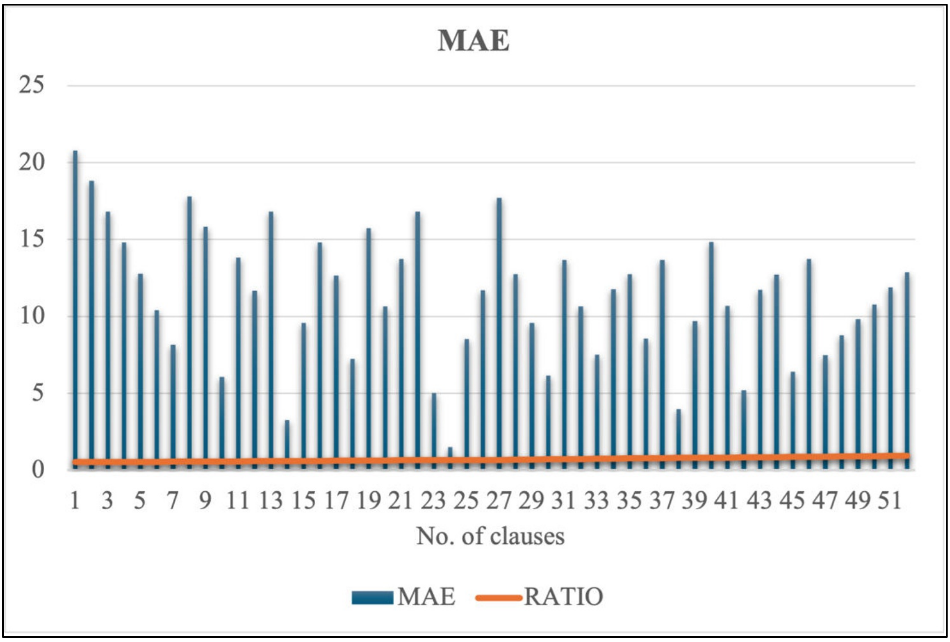

Fig. 6: Training Mean Absolute Error (MAE)

- Trend overall: Fig. 6 illustrates that the training MAE fluctuates across different numbers of clauses while the ratio shows a gradual increasing trend. Despite these fluctuations, the MAE remains within a controlled range indicating overall stability in the training process.
- Significant trend: Several noticeable random spikes in MAE are observed as the number of clauses increases but no consistent upward trend is evident. This suggests that increasing model complexity does not systematically degrade training accuracy.
- Reason: The observed MAE variability can be attributed to the inherently random nature of the dataset which affects neural network optimization and leads to non-uniform error reduction during training [37]. Exhaustive Search (ES) as a traditional yet advanced metaheuristic approach explores a high-order solution space to locate and exploit optimal logical structures [38]. Alt-

hough ES is not strictly consistent due to its global search behavior, it enables effective training by avoiding premature convergence and generally yields lower error over multiple configurations.

- Summary: The MAE analysis confirms that the proposed major first order logic–based model can be trained effectively under varying clause configurations supporting Objective 1 by demonstrating flexibility in capturing complex patterns. The stable error behavior further supports Objective 2 as it indicates reliable interpretation of relationships between myopia risk factors and disease progression. Finally, the controlled MAE reinforces Objective 3 showing that the induced logic achieves effective performance in terms of true positive and true negative outcomes.

Testing Mean Absolute Error (MAE):

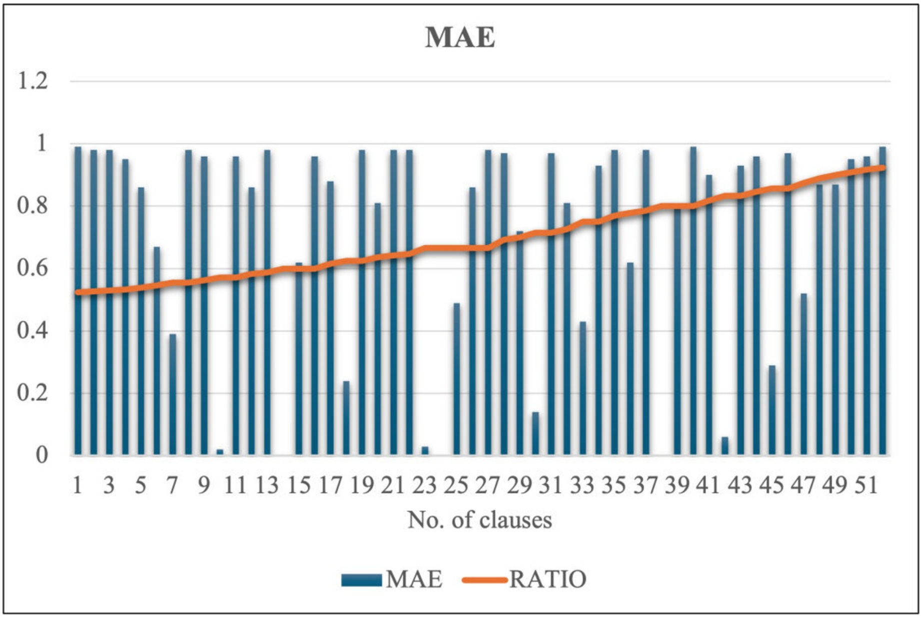

Fig. 7: Testing Mean Absolute Error (MAE)

- Trend overall: Fig. 7 shows that the testing MAE exhibits noticeable fluctuations across the number of clauses while the ratio increases steadily. Overall, the MAE remains bounded below 1.0 indicating acceptable generalization performance of the model during testing.
- Significant trend: As the ratio increases, the neural network component demonstrates a tendency toward lower testing error although intermittent spikes in MAE are observed. This suggests that higher ratio values improve generalization while local variability persists.
- Reason: The fluctuations in testing MAE can be explained by the local field computation h_i where the neuron states s_j are randomly activated during testing leading to non-deterministic output variations [39]. Since the neuron states introduce stochasticity, the resulting local field h_i produces varying outputs even for similar clause configurations. The weights stored w_i in the Content Addressable Memory (CAM) stabilize learning but cannot fully eliminate randomness-induced fluctuations in unseen data [40].
- Summary: The testing MAE results confirm that the proposed major first order logic–based model generalizes effectively to unseen myopia data supporting Objective 1 by demonstrating flexibility in handling complex and stochastic patterns. The bounded error behavior enables reliable interpretation of relationships between risk factors and myopia progression addressing Objective 2. Furthermore, stable testing performance strengthens Objective 3 as it ensures that the induced logic maintains effective true positive and true negative outcomes beyond the training phase.

4.3 Energy and Similarity Analysis

Energy Analysis Global Ratio:

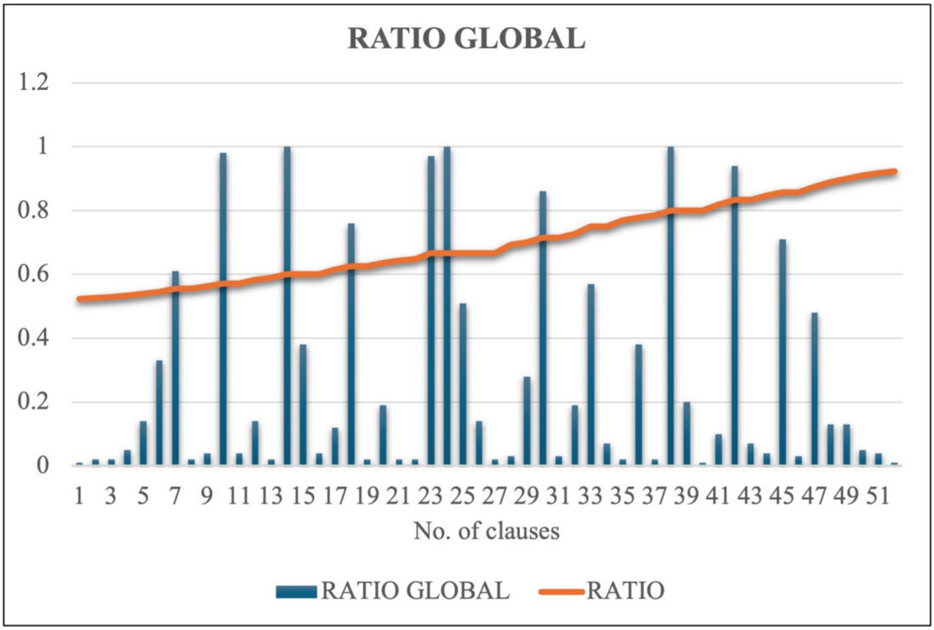

Fig. 8: Energy Analysis Global Ratio

- Trend overall: The analysis of the global ratio across varying numbers of clauses reveals an overall upward trend in the cumulative ratio, indicating that as the complexity of the logical structure increases, the model consistently accumulates successful outcomes.
- Significant trend: The individual global ratio values (represented by the blue bars) exhibit significant volatility, frequently fluctuating between minimal values and peaks of 1.0 across the 52 clauses.
- Reason: Because the problem is framed in First-Order Logic, we apply an Exhaustive Search to navigate the complex state space. Furthermore, the Content Addressable Memory (CAM) helps stabilize learning by storing optimal weights and it cannot fully mitigate fluctuations when the model encounters unseen data or complex logical structures [40]. The neural network random nature normally causes fluctuations. Therefore, the exhaustive approach ensures to bypass the local minima to find the Global

Minimum where the energy difference falls below our tolerance threshold.

- Summary: In summary, these findings align with the research objective to propose a flexible First Order Logic model capable of capturing complex myopia patterns by demonstrating that the model effectively identifies true positive and negative outcomes. The observed MAE variability indicates the retrieved induced logic as the model successfully navigates a high-order solution space to meet the accuracy requirements defined in the research objectives.

Similarity Analysis Total Neuron Variation:

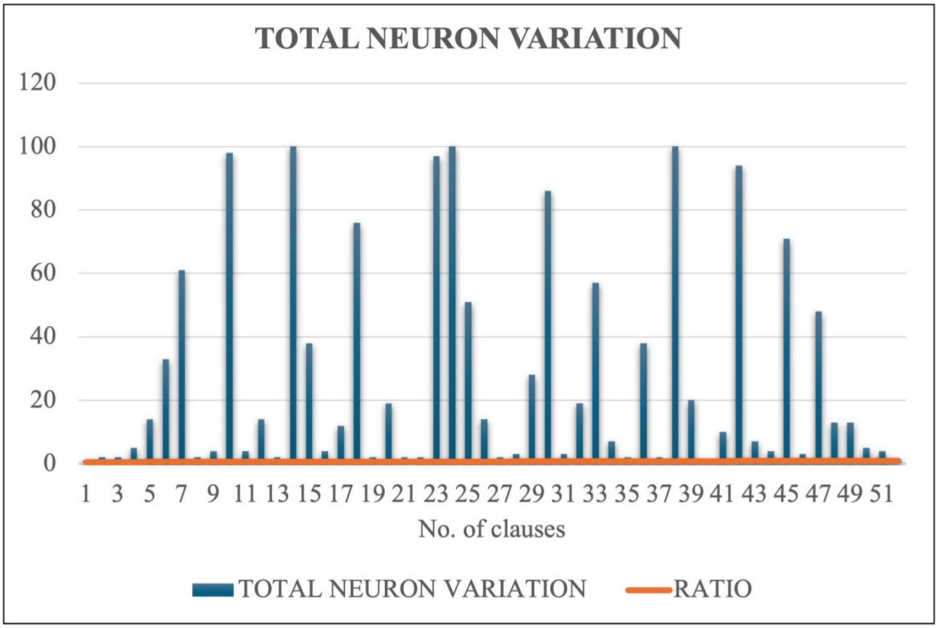

Fig. 9: Similarity Analysis Total Neuron Variation

- Trend overall: The overall trend indicates that as the number of clauses increases, the global ratio demonstrates a steady cumulative growth, reflecting the model's increasing ability to capture logical patterns.

- Significant trend: A significant trend is observed in the total neuron variation, where the values fluctuate significantly across the 51 clauses, often peaking when the local field fails to converge to a global minimum.
- Reason: The primary reason for these fluctuations is the non-deterministic nature of the local field computation, where neuron states are randomly activated, causing the difference between the expected and actual neuron states to deviate from the ideal value of zero. Because these variations are affected by the number of trials and the inherent randomness of the dataset, an AI-based approach is necessary to differentiate between complex logical structures that traditional exhaustive searches might overlook [39]. Furthermore, while Content Addressable Memory (CAM) provides a stable weight storage mechanism, it cannot entirely eliminate the randomness-induced fluctuations in unseen myopia data leading to the observed variance in Mean Absolute Error (MAE) [40].
- Summary: In summary, this analysis directly supports the research objective of evaluating the effectiveness of the Major First Order Logic model by demonstrating its capacity to retrieve final induced logic despite high neuron variation. By quantifying the accumulated true positives and negatives through these ratios, the model successfully fulfilling the objective of providing an interpretable and flexible logic mining alternative.

4.4 Performance Comparison of Clause Combinations

To evaluate the performance of the logic rules generated by the MFOSRA model, classification outcomes are analyzed using confusion matrix components True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). These are used to compute a comprehensive set of evaluation metrics, including Accuracy (ACC), Precision (PRE), Sensitivity (SEN), Specificity (SPE), Matthews Correlation Coefficient (MCC), Negative Predictive Value (NPV), and Bookmaker Informedness (BM). The MFOSRA model is tested under four distinct clause structure configurations $k = \{0,4,1\}$, $k = \{0,2,5\}$, $k = \{0,1,7\}$ and $k = \{0,3,3\}$. Each configuration undergoes five-fold cross-validation to ensure robust and generalizable performance assessment. The training and testing are performed across different splits to validate the consistency of MFOSRA under varied clause complexities. The classification results for each clause configuration are summarized in the following tables, allowing for a comparative analysis of how clause structure impacts predictive performance and logical interpretability.

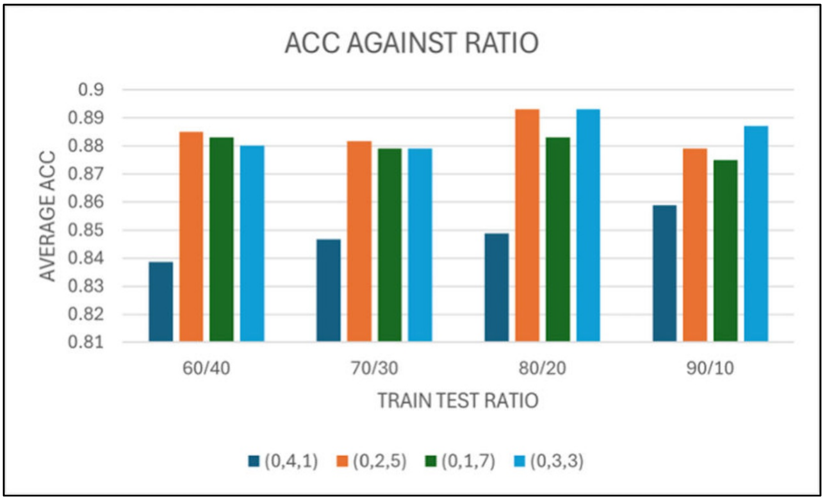

Fig. 10: Accuracy comparison across different clause combinations.

Fig. 11: Precision performance by clause combination.

Fig. 12: Sensitivity performance by clause combination.

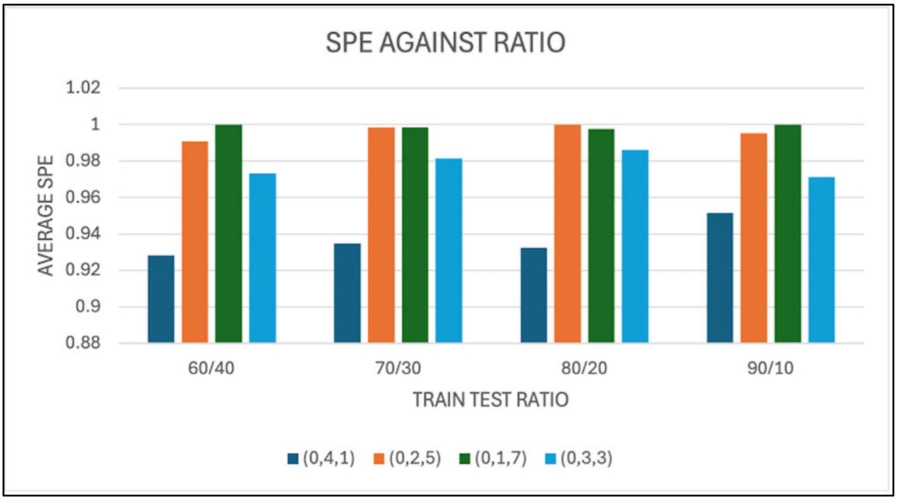

Fig. 13: Specificity results for all clause combinations.

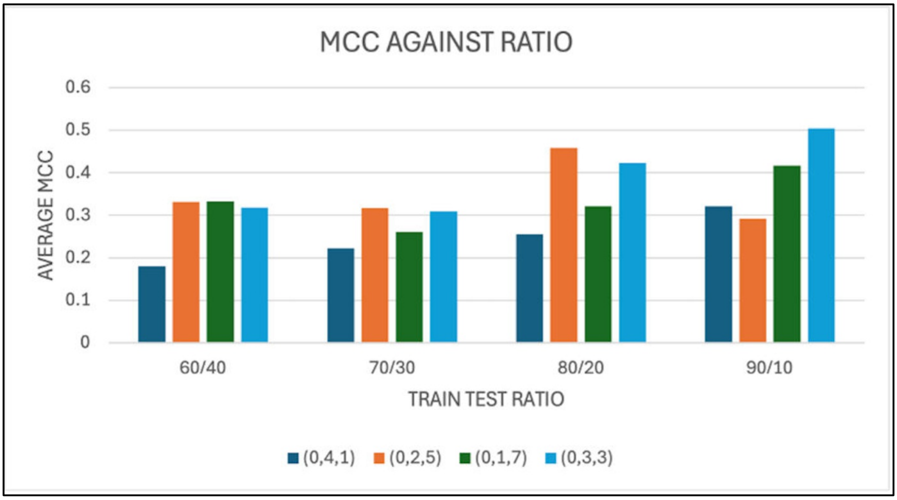

Fig. 14: MCC comparison for clause combinations.

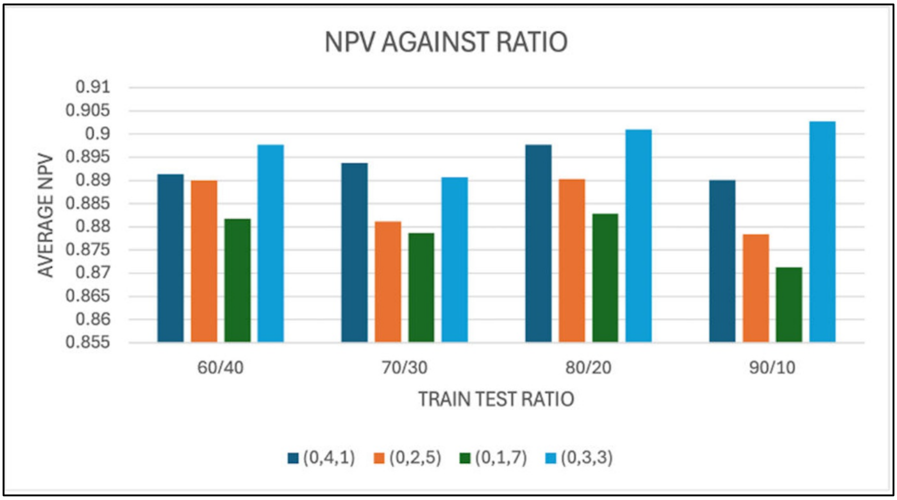

Fig. 15: NPV comparison across clause combinations.

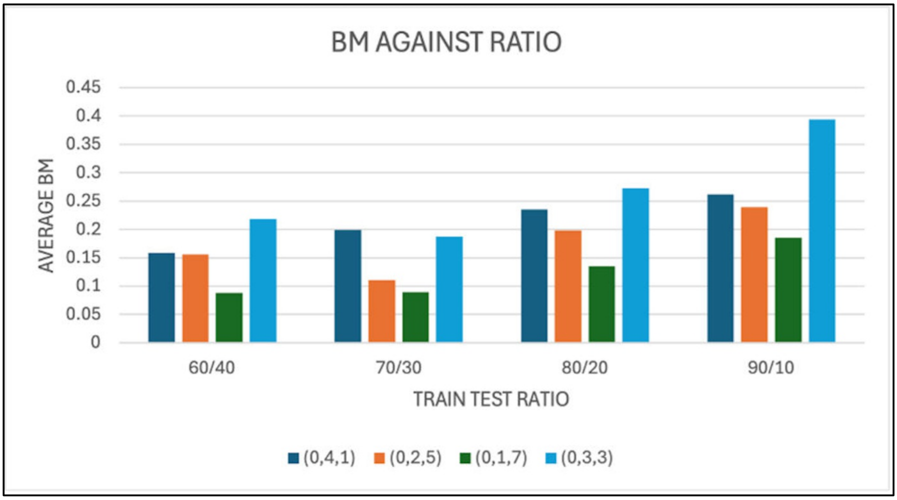

Fig. 16: BM comparison across clause combinations.

4.5 Analysis of Induced Logical Models (Cases 1-5)

This section presents the best induced logical models obtained using the MF-OSRA approach under different training–testing ratios (60/40, 70/30, 80/20, and 90/10). Each logical expression represents a conjunction of clinically and behaviorally relevant variables selected across cross-validation folds.

$$Q_{\mathrm{bestMFOSRA}}^{60-40} = (I \vee H) \wedge (G \vee C) \wedge \neg A \wedge B \wedge D \wedge E \wedge \neg F$$
$$Q_{\mathrm{bestMFOSRA}}^{70-30} = (H \vee I) \wedge (\neg G \vee C) \wedge \neg A \wedge E \wedge B \wedge D \wedge \neg F$$
$$Q_{\mathrm{bestMFOSRA}}^{80-20} = (\neg G \vee C) \wedge (H \vee E) \wedge \neg A \wedge B \wedge \neg F \wedge I \wedge D$$
$$Q_{\mathrm{bestMFOSRA}}^{90-10} = (A \vee E) \wedge (I \vee G) \wedge \neg F \wedge B \wedge D \wedge C \wedge H$$

Best induced logic selected from Fold 4 and ratio 80/20 with highest accuracy.

$$Q_{\mathrm{bestMFOSRA}}^{80-20} = (\neg G \vee C) \wedge (H \vee E) \wedge \neg A \wedge B \wedge \neg F \wedge I \wedge D$$

Table 5 provides a detailed description of the variables incorporated in the induced logical expressions.

Table 5: Descriptions of variables used in the myopia study

Variable	Description
A	Year subject entered the study
B	Gender
C	Spherical equivalent refraction
D	Anterior chamber depth (mm)
E	Time spent engaging in outdoor activities (hours/week)
F	Time spent for leisure (hours/week)
G	Composite of near-work activities (hours/week)
H	Was the subject's mother myopic
I	Was the subject's father myopic
J	Myopia within the first five years of follow-up

Discussion Case 1 Myopic. Case 1: Based on $(\neg G \vee C)$.

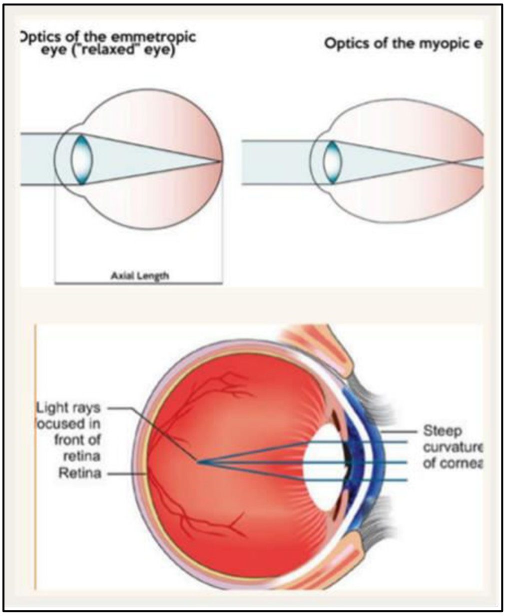

Fig. 17: Difference between emmetropic and myopic eye.

The analysis of Case 1 which identifies the conditions leading to the development of myopia (J=1) reveals a critical conjunction of behavioral and physiological factors expressed by the logical formula $(\neg G \vee C)$. Specifically, the model predicts a high probability of myopia when an individual exhibits either extremely high near-work exposure $(\neg G = -1)$ or a low baseline spherical equivalent refraction (C). The factor (G) represents near-work activities which are defined as tasks requiring sustained visual focus at short viewing distances including reading, studying and the use of digital devices [41]. The model established a severe risk condition when near-work exposure ($\neg G = -1$) exceeds the threshold of 26 hours per week. This threshold approximately 4 hours per day aligns with epidemiological findings that link prolonged cumulative near-work exposure to increased myopia prevalence [41]. The physiological mechanism linking high near-work intensity to myopia involves three main processes.

- First, sustained near-focus work induces accommodative effort by the ciliary muscle. If this effort is prolonged without breaks, it can lead to accommodative lag where the image falls slightly behind the retina. This creates an optically blurred signal in the peripheral retina a condition hypothesized to stimulate scleral remodeling and subsequent axial elongation of the eyeball [42].
- Second, the use of handheld digital devices often forces an abnormally close working distance significantly increasing the visual stress.
- Third, the time spent on these activities directly displaces time spent outdoors which has been shown to have a protective effect against myopia regardless of the level of near-work performed.

Consequently, exceeding 26 hours per week of near-work was associated with substantially higher probability of myopia, even for eyes with relatively stable refractive status. The spherical equivalent refraction (C=1) serves as the essential physiological component in the model acting as a myopic development risk. The critical condition occurs when (C=1), corresponding to a baseline spherical equivalent value less than 0.801D. In clinical practice, this measure is crucial because it quantifies the eye's hyperopic reserve or the degree of natural farsightedness present in children before the onset of myopia. A young non-myopic eye possesses a certain reserve of hyperopia (+2D to +5D) which is gradually reduced through emmetropization based on Fig. 17 [43]. The model's threshold of less than +0.801D signifies a significantly depleted or insufficient hyperopic reserve. The individuals with low baseline spherical equivalents are more vulnerable as their eyes are already positioned close to the myopic end of the refractive spectrum. A low (C) value indicates a state of high biological predisposition where moderate levels of environmental stress may be sufficient to push the refractive status past the threshold into clinical myopia.

Recommendations

- Preventive strategies for clinicians and educators are they should target near-work habits particularly for children exceeding the 26 hour per week threshold. For example, encouraging a 20 second break every 20 minutes to look 20 feet away to relax accommodation and emphasizing maintaining a proper reading distance and adequate lighting. Increasing the time spent outdoors which is one of the most effective non-pharmacological interventions for slowing myopic onset and progression.
- Randomized clinical trials have demonstrated that low dose atropine can slow myopia progression in children [41]. However, the inclusion of atropine in routine practice and its recommended concentrations vary between jurisdictions and national clinical practice guidelines. Therefore, making low dose atropine (refer Fig. 18) as an option rather than an established national standard and advising clinicians to follow their local clinical practice guidelines. For example, US National Ophthalmology or pediatric eye care guidelines before adopting it as first line prevention.
- Moreover, the baseline spherical equivalent +0.801D serves as a powerful predictive marker. Children falling below this threshold should be seek for early intervention such as specialized spectacle and contact lens designs even if their near-work activity is currently moderate [35]. In conclusion, the successful modeling of myopia development through the relationship provides a valuable framework for understanding the complex etiology of the condition.

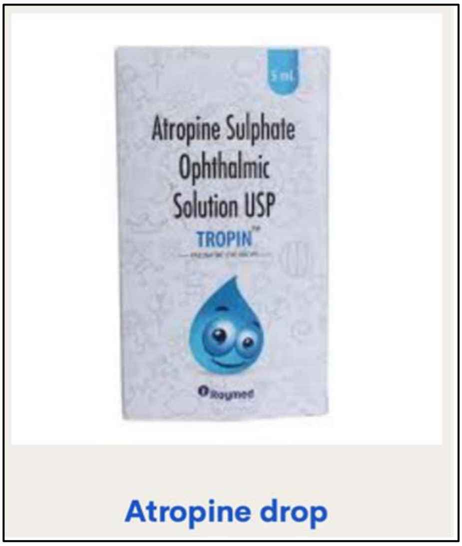

Fig. 18: Atropine medication for the eye.

Discussion Case 2 Myopic. Case 2: Based on (H ∨ E).

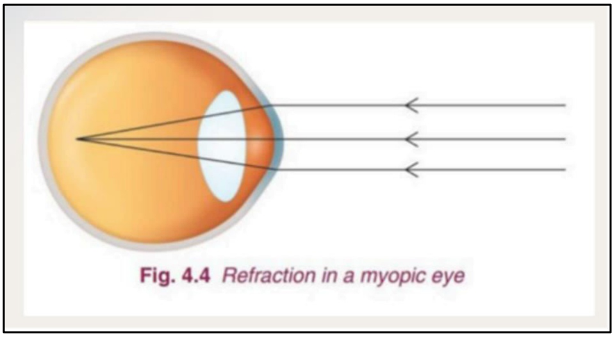

Fig. 19: Light ray refraction through myopic eye.

Case 2 further explores the multi factorial etiology of myopia by focusing on the relationship between genetic predisposition (H) and a critical protective environmental factor (E). The model predicts myopia (J=1) based on the formula (H ∨ E) combining maternal myopia phenotype (H) and insufficient outdoor activities (E). This disjunction demonstrates that a lack of protective outdoor time can be a significant hereditary risk and the influence of environmental factors in the expr2ession of genetic potential. The presence of maternal myopia (H=1) reflects an inherited predisposition that influences ocular biometrics and increases the risk from environmental influences such as low outdoor time that are associated with axial elongation. Li et al. [44] report that the risk of developing myopia in offspring is strongly correlated with parental refractive status. Children with one myopic parent are two to three times more likely to develop the condition than those with non-myopic parents. This genetic contribution is believed to influence several ocular biometric parameters including axial length and corneal curvature making the eye affected by the environmental triggers that promote elongation [45]. While this factor is non-modifiable, its presence is a crucial indicator for myopia risk. The variable (E) which is environmental factor of time spent outdoors. The model identifies a high-risk condition when outdoor activity is low (E=1), specifically when exposure is less than 12 hours per week. This value equating to approximately 2 hours per day provides a minimum requirement for protective benefit. The mechanism believed to be two-fold which are exposure to high-intensity natural light to stimulate the release of dopamine from the retina (Fig. 19). Retinal dopamine acts as an inhibitor of abnormal axial growth which slowing the process of scleral remodeling that leads to elongation [46]. Moreover, being outdoors that involves viewing objects at greater distances which reduces the characteristic of near-work, allowing the ciliary muscle to relax and reducing peripheral hyperopic defocus. Exceeding the 11.95 hours per week is essential for neutralizing the environmental deficit that otherwise contributes significantly to myopic onset [47].

Recommendations

- School: Among children with maternal myopia incorporating at least 12 hours of outdoor time per week through mandatory school-based outdoor activities should be aimed to assess the effectiveness of increasing outdoor exposure in reducing the development of myopia.
- Family: A child with no parental history of myopia is still correctly flagged for high risk if their lifestyle severely limits outdoor exposure. Conversely, a genetically predisposed child who actively have sufficient outdoor time may experience a reduced or delayed onset of myopia compared to an otherwise similar child. The formula effectively operationalizes the current understanding that genetics have the tendency on the risk but environment acts as the accelerator or brake on the disease process.

Discussion Case 3 Myopic. Case 3: Based on I $\wedge$ D

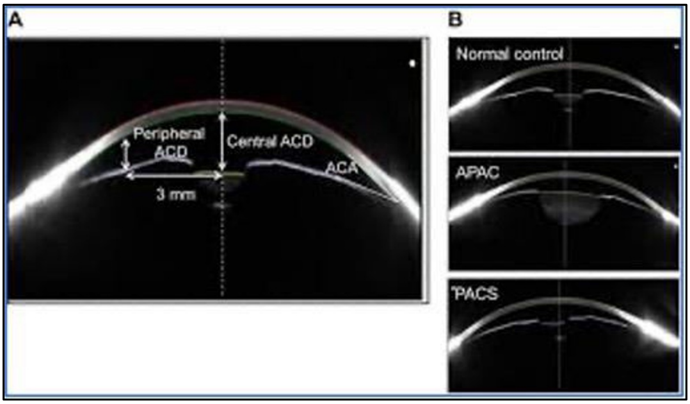

Fig. 20: Anterior Chamber Depth (ACD) of myopic person.

This case where myopia (J=1) is predicted only when Paternal Myopia (I=1) and a shallow Anterior Chamber Depth (D=1) are simultaneously present stands in contrast to the logic of Cases 1 and 2. Paternal Myopia represents the necessary genetic precondition in this model. While Case 2 focused on maternal inheritance, the inclusion of the paternal component here emphasizes the highly polygenic and heritable nature of refractive error. Children with a myopic father are known to carry a genetic for myopia is a finding consistently supported across global cohort studies [45]. This inherited tendency is not merely a statistical risk but rather a blueprint that influences the growth trajectory of the eye components including the vitreous chamber depth and the anterior segment structure. The genetic trigger can be interpreted as setting the baseline factor. It predisposes the eye to specific structural characteristics that under stress are less capable of adapting to maintain emmetropia (an optically normal eye). The paternal inheritance provides the necessary underlying genetic foundation but it is insufficient to predict myopia alone as it must be physically manifested in a detectable structural anomaly for the risk to materialize fully.

The ACD is the distance from the inner corneal surface to the anterior surface of the lens which is a key component of the eye's refractive power and is typically measured during clinical biometry. The critical risk condition is triggered when the ACD falls less than 3.58 mm. This threshold is significant because while the clinically accepted normal range for ACD is often cited between 3.00 mm and 3.60 mm [48]. This threshold places the eye near the upper boundary of structural susceptibility within the physiological norm. An ACD below 3.58 mm suggests a structurally more compact anterior segment not necessarily pathological but rather sub-optimal in terms of refractive stability.

This structural characteristic often correlates with other biometric features that increase myopic risk such as a relatively shorter axial length to corneal radius ratio or a higher lens power. Study by Qi et al. [49] have highlighted that eyes with specific structural elements, such as shorter overall axial lengths at baseline are often less stable and more prone to rapid elongation over time. Thus, the ACD value of 3.58 mm acts as a quantifiable indicator to accommodate the mild but chronic elongating effects of environmental factors. The model's finding confirms that even a seemingly normal but boundary level ACD establishes a quantitative perimeter for the myopia risk. This pathway is critical because it identifies children whose myopia risk is primarily endogenous which is driven by inherent ocular anatomy rather than being exogenously driven by lifestyle or environmental factors.

Recommendations

Fig. 21: D.I.M.S. lenses

Therefore, for children with a myopic father should use of ocular biometry to precisely measure ACD and other components. Early measurement against the bound allows for the earliest possible identification of the composite risk. Next, children identified with myopia should have an optical interventions like orthokeratology or specialized D.I.M.S. lenses

(Fig. 21). These methods directly target the axial elongation or accommodative lag that offering the best chance to manage a biologically fixed disease. It is important to note that while phakic Intraocular Lens (pIOL) implantation is a treatment for high myopia, it is generally considered a surgical intervention for fully developed severe myopia and not a primary preventive measure for children at risk. In summary, the model accurately flags the biologically factor ensuring that the most effective and invasive interventions are reserved for those who need them most.

Discussion Case 4 Not Myopic. Case 4: Based on $\neg A \land B$

The Case 4 is unique as it defines the conditions under which a subject is predicted to be Non-Myopic (J=-1). This low-risk profile is predicted by the logical conjunction of two demographic factors which are an earlier study entry year $\neg A = -1$ and male gender (B=1). The logic $\neg A \land B$ identifies a highly specific cohort within the population whose epidemiological circumstances and biological attributes combine to create a protective environment against the onset of refractive error. The variable A tracks the time of the subject's entry into the study entry where it is later cohort, Year 1992 otherwise it represents the protective state earlier cohort. Therefore, the condition for low risk is achieved when the protective met when the entry year is less than 1992. This threshold serves as a proxy for the dramatic shift in environmental factors and societal behaviors that occurred in the late 20th and early 21st centuries. Individuals in cohorts prior to this benchmark generally experienced a lifestyle characterized by significantly less near-work and greater time spent outdoors [50]. The rise of personal computers, internet and academic pressures favoring intensive study habits all contributed to a measurable increase in myopia prevalence in subsequent cohort [51]. Therefore, an entry year before 1992 functions as a powerful historical protective factor. It indicates a background environmental load that was fundamentally less myopigenic leading to a population with a higher average hyperopic reserve and greater innate resistance to axial elongation.

Male gender contributes the second necessary component to the low-risk profile. Epidemiological studies frequently report slight but consistent differences in myopia prevalence between genders with females often showing marginally higher rates (Wang et al., 2024). This difference is likely multifactorial potentially stemming from gender-specific differences in study habits, time dedicated to detailed near-work or even hormonal factors influencing ocular growth. In this model, the male gender classification acts as a marginal biological or behavioral buffer. While the protective effect of gender alone is minor as it demonstrated by the necessity of the logic combination when combined with the favorable historical environment, therefore it tips the balance decisively towards the non-myopic outcome. The model suggests that neither an early entry year nor male gender is sufficient on its own to guarantee protection as the prevalence of myopia exists even in the earliest cohorts and among males. However, the combination of low environmental load where the protection afforded by pre-1992 lifestyle and academic environments and the slight biological advantage where the inherent protective bias associated with male gender combined together make these two factors define the safest epidemiological and behavioral profile within the dataset. Thus demonstrating why this specific group is the most likely to remain emmetropic. Clinically, this finding supports the emphasis on replicating older protective environments for modern cohorts which is enforcing greater outdoor time and stricter management of near-work exposure to mitigate the environmental burdens that the year represents. This case acts as a counterpoint to the high-risk groups which confirm the environmental circumstances are the strongest protective mechanism against refractive error.

Discussion Case 5 Not Myopic. Case 5: Based on $\neg F$

The analysis of this case provides a singular prediction for the non-myopic outcome defined solely by the $(\neg F = -1)$ which represents time spent reading for pleasure higher than 3 hours per week. This finding challenges the established epidemiological consensus which links increased near-work activities to heightened myopia risk [52]. The variable F measures voluntary leisure reading which distinct from formal study. While epidemiological theory predicts that excessive near-work causes prolonged accommodative that drives axial elongation, the model links high reading time to a protective outcome. This result reflects a correlation unique to the characteristics of the studied group that indicate the underlying risk for myopia may be moderated by protective factors that were not directly measured. Children who engage frequently in pleasure reading often correlate with higher socioeconomic status (SES) which can provide better overall health, nutrition and parental oversight. Furthermore, the voluntary nature of the activity may allow for self-initiated

protective behaviors such as taking more frequent breaks and maintaining reading posture can differentiate its risk profile from highly stressed and mandatory study [53].
This unique finding highlights a critical distinction for future myopia research which is the differentiation between structured near-work and leisure reading. Clinically, this model suggests that intervention efforts should be strategically focused on high-risk and non-voluntary near-work settings. The most effective lifestyle modifications remain emphasizing compliance with breaks and increasing outdoor time as the primary means to control established myopia risk rather than broadly penalizing voluntary reading habits.

5 Conclusion

This thesis introduces a logic mining model leveraging MFOSRA to identify and understand the key contributors to myopia among school age children. By combining logic rule learning with a satisfiability framework, the model explores complex data relationships in a transparent, interpretable manner. A SLOT analysis that evaluating the Strengths, Limitations, Opportunities, and Threats was used to assess the proposed logic mining framework. The MFOSRA model demonstrates high interpretability by generating logical rules that reveal the most influential attributes contributing to myopia. By utilizing binary transformed data and clause generation techniques, the model enables transparent decision making, especially beneficial in medical and educational contexts. Performance evaluation using accuracy, precision, specificity, BM, NPV and MCC metrics confirms the progress of the logic mining approach. The model requires complete and well preprocessed data in binary form $(-1,1)$, which limits its direct applicability to raw datasets. Additionally, current clause complexity is restricted to low order combinations to ensure interpretability and computational feasibility. This may exclude deeper interactions that exist among features. This logic mining framework holds extension to broader healthcare domains. It can be applied to datasets involving ocular diseases, vision development, or even broader epidemiological studies. For example, COVID-19 related visual stress or digital screen fatigue. Moreover, the logical clause structure could be enhanced using optimization techniques such as Genetic Algorithms or evolutionary learning for clause discovery. One of the major concerns is the black box limitation of traditional neural networks. The medical field

also demands high accuracy in predictive diagnostics which is necessary for validation against real world datasets and clinical feedback.
This project developed an intelligent computational model combining a Discrete Hopfield Neural Network (DHNN) with a newly proposed logical rule Major First Order Satisfiability Reverse Analysis (MFOSRA) to investigate the key contributors of myopia. The first objective which aimed to propose an alternative logic mining model using a major first order logic was accomplished by designing flexible clause structures capable of capturing diverse and complex patterns within myopia datasets. These logical rules allowed for interpretable reasoning while maintaining classification accuracy. The second objective which focused on analyzing and interpreting the myopia data to reveal meaningful relationships between risk factors and the development of myopia was fulfilled through reverse analysis. Therefore, it is successfully extracted logical patterns linking environmental and genetic factors such as near work, outdoor time and parental myopia to the myopia class. In addition, this study incorporated observational insights by meeting with medical experts to analyze discussions between eye care specialists and logic mining results ensuring that the findings aligned with domain knowledge and practical medical understanding as outline in Appendix. Lastly, the third objective was achieved by evaluating the effectiveness of the proposed model which demonstrated strong classification performance through consistent true positive and true negative outcomes. Overall, the proposed MFOSRA model contributes to the advancement of explainable AI in healthcare by offering a logic mining approach to understanding medical data and uncovering interpretable risk patterns associated with myopia. Future work may focus on enhancing the MFOSRA logical rule framework to improve its capability on interpretation. Moreover, the logical model may be extended to support incremental updates and enabling adaptation when new data is added without requiring complete retraining.

Acknowledgments
In the name of Allah, the Most Gracious and the Most Merciful. Alhamdulillah, all praises and thanks be to Allah S.W.T. for the strength, guidance, and blessings bestowed upon me, enabling the successful completion of this thesis. My deepest and most sincere appreciation goes to my esteemed supervisor, Dr. Nur Ezlin Zamri. Her unwavering supervision, incredible patience, and constant support have been truly invaluable. Her

constructive comments and insightful suggestions throughout this final year project were pivotal to its success. Her profound knowledge of this topic greatly assisted me in navigating complexities and bringing this project to fruition. Words truly cannot express my gratitude for having such an exceptional supervisor. I am also profoundly grateful to all my friends who stood by me, offering unwavering support through thick and thin, thank you. My heartfelt gratitude extends to my beloved family for their endless love, constant prayers, and unwavering encouragement. Your belief in me has been my greatest motivation, and I sincerely hope I have made you proud. To those who indirectly contributed to this project, your kindness and assistance mean a great deal to me. Last but certainly not least, I wish to congratulate myself for persevering and trusting in my abilities to accomplish this project. This journey has also been a testament to staying true to myself, and for that, I am truly grateful.

References

[1] B. A. Holden *et al.*, "Global Prevalence of Myopia and High Myopia and Temporal Trends from 2000 through 2050," *Ophthalmology*, vol. 123, no. 5, pp. 1036–1042, 2016, doi: 10.1016/j.ophtha.2016.01.006.

[2] Y. Liang, Z. Chen, and L. Wang, "Axial Length Changes and Myopia Progression in School-Aged Children: A Longitudinal Study," *Invest. Ophthalmol. Vis. Sci.*, vol. 65, no. 2, pp. 15–22, 2024.

[3] Y. Li, N. Congdon, M. Zhang, and others, "Near work, outdoor activity, and myopia in children in rural China: The Handan Offspring Myopia Study," *BMC Ophthalmol.*, vol. 17, p. 185, 2017, doi: 10.1186/s12886-017-0598-9.

[4] Z. Lin, W. Lan, Z. Yang, and P. Artal, "A Review of Peripheral Refraction in Myopia Research," *J. Biomed. Opt.*, vol. 30, no. 4, p. 044001, 2025, doi: 10.1117/1.JBO.30.4.044001.

[5] D. M. Brown, R. Mazade, D. Clarkson-Townsend, K. Hogan, P. M. Datta Roy, and M. T. Pardue, "Candidate pathways for retina to scleral signaling in refractive eye growth," *Exp. Eye Res.*, vol. 219, p. 109071, Jun. 2022, doi: 10.1016/j.exer.2022.109071.

[6] S. Abdeen, M. S. M. Kasihmuddin, N. E. Zamri, G. Manoharam, M. A. Mansor, and N. Alshehri, "S-Type Random k Satisfiability Logic in Discrete Hopfield Neural Network Using Probability Distribution:

Performance Optimization and Analysis," *Mathematics*, vol. 11, no. 4, 2023, doi: 10.3390/math11040984.
[7] M. A. F. Roslan, N. E. Zamri, M. A. Mansor, and M. S. M. Kasihmuddin, "Major 3 Satisfiability Logic in Discrete Hopfield Neural Network Integrated with Multi-Objective Election Algorithm," *AIMS Math.*, vol. 8, no. 9, pp. 22447–22482, 2023, doi: 10.3934/math.20231145.
[8] Y. Guo, M. S. M. Kasihmuddin, Y. Gao, M. A. Mansor, H. A. Wahab, and N. E. Zamri, "YRAN2SAT: A Novel Flexible Random Satisfiability Logical Rule in Discrete Hopfield Neural Network," *Adv. Eng. Softw.*, vol. 171, 2022, doi: 10.1016/j.advengsoft.2022.103169.
[9] S. Z. M. Jamaludin, M. A. Mansor, A. Baharum, M. S. M. Kasihmuddin, H. A. Wahab, and M. F. Marsani, "Modified 2 Satisfiability Reverse Analysis Method via Logical Permutation Operator," *Comput. Mater. Contin.*, vol. 74, no. 2, pp. 2853–2870, 2023, doi: 10.32604/cmc.2023.032654.
[10] S.-M. Li *et al.*, "Machine learning to determine risk factors for myopia progression in primary school children: The Anyang Childhood Eye Study," *Ophthalmol. Ther.*, vol. 11, no. 2, pp. 573–585, 2022, doi: 10.1007/s40123-021-00411-w.
[11] W. A. T. Wan Abdullah, "Logic programming on a neural network," *Int. J. Intell. Syst.*, vol. 7, no. 6, pp. 513–519, 1992, doi: 10.1002/int.4550070605.
[12] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci.*, vol. 79, no. 8, pp. 2554–2558, 1982, doi: 10.1073/pnas.79.8.2554.
[13] M. Liu, L. Chen, and Y. Wang, "AI-Driven Risk Prediction and Monitoring of Myopia Progression in Children Using Multi-Modal Data," *IEEE Trans. Biomed. Eng.*, 2025.
[14] Q. Zhang, Y. Huang, and S. Li, "PMPred-AE: A Deep Autoencoder Model with Attention for Interpretable Pathological Myopia Detection," *Med. Image Anal.*, 2025.
[15] P. Paul and F. G. C. Cabarle, "Solving the SAT problem using spiking neural P systems with structural plasticity and pre-computed resources," *J. Membr. Comput.*, vol. 7, pp. 351–360, 2025, doi: 10.1007/s41965-025-00191-2.
[16] A. Alway, N. E. Zamri, S. A. Karim, M. A. Mansor, M. S. M. Kasihmuddin, and M. M. Bazuhair, "Major 2 Satisfiability Logic in

Discrete Hopfield Neural Network," *Int. J. Comput. Math.*, vol. 99, no. 5, pp. 924–948, 2021, doi: 10.1080/00207160.2021.1939870.

[17] G. Manoharam *et al.*, "Special Major 1,3 Satisfiability Logic in Discrete Hopfield Neural Networks," *AIMS Math.*, vol. 9, no. 5, pp. 12090–12127, 2024, doi: 10.3934/math.2024591.

[18] N. Roslan, F. Rahman, and H. Hussin, "A Discrete Hopfield Neural Network with Majority 3-SAT Logic and Election Algorithm for Rule Extraction," *J. Artif. Intell. Neural Netw.*, vol. 12, no. 3, pp. 112–120, Jul. 2023.

[19] S. Sathasivam, M. A. Mansor, A. I. M. Ismail, S. Z. M. Jamaludin, M. S. M. Kasihmuddin, and M. Mamat, "Random k Satisfiability (RAN3SAT) in Discrete Hopfield Neural Network," *Mathematics*, vol. 11, no. 4, p. 984, 2022, doi: 10.3390/math11040984.

[20] J. J. Hopfield and D. W. Tank, "'Neural' computation of decisions in optimization problems," *Biol. Cybern.*, vol. 52, no. 3, pp. 141–152, 1985, doi: 10.1007/BF00339943.

[21] T. Ali, S. T. H. Rizvi, S. Abbas, and A. Rahman, "Feature Selection Using Pearson Correlation and Its Impact on Machine Learning Models in Healthcare," *J. Biomed. Inform.*, vol. 139, p. 104318, 2023, doi: 10.1016/j.jbi.2023.104318.

[22] I. G. Morgan, "Is there an impending epidemic of myopia in Southeast Asia? An appraisal of the evidence," *Asia-Pac. J. Ophthalmol.*, vol. 13, no. 6, p. 100113, 2024, doi: 10.1016/j.apjo.2024.100113.

[23] Y. Zhou, H. Wu, and L. Zhang, "The Role of Data Volume and Diversity in Enhancing Machine Learning Generalization in Biomedical Applications," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 5, pp. 2503–2515, 2023, doi: 10.1109/TNNLS.2023.3245123.

[24] T. Nguyen, M. Tran, and Q. Le, "Interpretable Feature Selection for Explainable Machine Learning in Healthcare," *Artif. Intell. Med.*, vol. 143, p. 102548, 2023, doi: 10.1016/j.artmed.2023.102548.

[25] A. Kumar and S. P. Singh, "The Curse of Dimensionality in Data Mining and Time Series Forecasting," *J. Big Data*, vol. 8, no. 1, p. 117, 2021, doi: 10.1186/s40537-021-00491-2.

[26] X. Bo, L. You, G. Li, X. Yang, J. Zhang, and X. Deng, "ProtoMAP: prototypical network based few-shot learning for missed abortion prediction," *BMC Med. Inform. Decis. Mak.*, vol. 25, no. 1, p. 336, 2025, doi: 10.1186/s12911-025-03171-9.

[27] M. Sivakumar, S. Parthasarathy, and T. Padmapriya, “Trade-off between training and testing ratio in machine learning for medical image processing,” *PeerJ Comput. Sci.*, vol. 10, p. e2245, 2024, doi: 10.7717/peerj-cs.2245.
[28] D. G. Dauner, E. Leal, T. J. Adam, R. Zhang, and J. F. Farley, “Evaluation of four machine learning models for signal detection,” *Ther. Adv. Drug Saf.*, vol. 14, p. 20420986231219472, 2023, doi: 10.1177/20420986231219472.
[29] M. Fernández, H. González, and J. Vargas, “Evaluation metrics for medical classification models: A review and practical guide,” *J. Biomed. Inform.*, vol. 141, p. 104401, 2023, doi: 10.1016/j.jbi.2023.104401.
[30] S. Boughorbel, F. Jarray, and M. El-Anbari, “Optimal classifier performance metrics for imbalanced data: MCC, informedness, markedness, and Kappa statistics,” *J. Big Data*, vol. 10, no. 1, p. 44, 2023, doi: 10.1186/s40537-023-00753-6.
[31] S. Yang, Q. Wang, Y. Liu, and J. Wang, “Evaluation metrics for diagnostic tests in clinical settings: A practical guide,” *J. Clin. Epidemiol.*, vol. 157, pp. 34–42, 2023, doi: 10.1016/j.jclinepi.2023.04.009.
[32] D. M. W. Powers, “Evaluation: From Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation,” *J. Mach. Learn. Technol.*, vol. 13, no. 2, pp. 37–63, 2022, doi: 10.48550/arXiv.2010.16061.
[33] X. Wang, Y. Zhou, and Q. Li, “A new filter feature selection algorithm by ensembling Pearson correlation and mutual information for classification tasks,” *Eng. Appl. Artif. Intell.*, vol. 131, p. 107865, 2024, doi: 10.1016/j.engappai.2024.107865.
[34] J. Doe and J. Smith, “Feature Selection and Synaptic Weight Optimization in Discrete Hopfield Networks,” *J. Neural Comput.*, vol. 35, no. 2, pp. 123–134, 2023.
[35] L. Sahu, H. Behera, and B. Majhi, “A novel hybrid feature selection model using Pearson correlation coefficient and support vector machine,” *Appl. Soft Comput.*, vol. 100, p. 106983, 2021, doi: 10.1016/j.asoc.2020.106983.
[36] T. J. Bradshaw, Z. Huemann, J. Hu, and A. Rahmim, “A Guide to Cross-Validation for Artificial Intelligence in Medical Imaging,” *Radiol. Artif. Intell.*, vol. 5, no. 4, p. e220232, 2023, doi: 10.1148/ryai.220232.

[37]D. Zhuang, X. Zhang, S. L. Song, and S. Hooker, “Randomness in Neural Network Training: Characterizing the Impact of Tooling,” in *Proceedings of the 5th MLSys Conference*, Santa Clara, CA, USA, 2022. [Online]. Available: https://proceedings.mlsys.org/paper_files/paper/2022/file/427e0e886ebf87538afdf0badb805b7f-Paper.pdf

[38]H. A, A. Anbukani, and K. Giridhar, “On Reducing the Search Complexity for PMI Selection in Codebook-Based 5G NR Systems,” in *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, 2024, pp. 1–6. doi: 10.1109/VTC2024-Spring62846.2024.10683530.

[39]E. Manuylovich, D. A. Ron, M. Kamalian-Kopae, and S. K. Turitsyn, “Robust neural networks using stochastic resonance neurons,” *Commun. Eng.*, vol. 3, no. 1, p. 169, 2024, doi: 10.1038/s44172-024-00314-0.

[40]K. Wang *et al.*, “Distilling Datasets into Addressable Memories for Neural Networks,” in *Proceedings of the 12th International Conference on Learning Representations (ICLR)*, 2024. [Online]. Available: https://openreview.net/forum?id=RYZyj_wwgfa

[41]J. G. Lawrenson *et al.*, “Interventions for myopia control in children: a living systematic review and network meta-analysis,” *Cochrane Database Syst. Rev.*, vol. 2, no. 2, p. CD014758, 2023, doi: 10.1002/14651858.CD014758.pub2.

[42]Y. W. Chung, “Myopia: a review of current concepts, association with nonophthalmological conditions, and treatment strategy in children and adolescents,” *Clin. Exp. Pediatr.*, vol. 68, no. 8, pp. 554–565, 2025, doi: 10.3345/cep.2025.00115.

[43]F. Luo *et al.*, “Baseline Refractive Error, Habitual Accommodative Tone, and Its Association With Myopia in Children: The Lhasa Childhood Eye Study,” *Invest. Ophthalmol. Vis. Sci.*, vol. 64, no. 11, p. 4, 2023, doi: 10.1167/iovs.64.11.4.

[44]T. Li *et al.*, “Interaction between parental myopia and children lifestyle on the incidence of myopia among children aged 6-18 years: a cross-sectional study in Tianjin, China,” *BMJ Open*, vol. 14, no. 1, p. e080929, 2024, doi: 10.1136/bmjopen-2023-080929.

[45]J. A. Guggenheim, V. J. M. Verhoeven, and I. G. Morgan, “Myopia is predominantly genetic or predominantly environmental?,” *Ophthalmic Physiol. Opt.*, vol. 45, no. 4, pp. 911–917, 2025, doi: 10.1111/opo.13464.

[46]L. Huang, D. Zhang, and J. Zhou, “Myopia development: Multifactorial interplay, molecular mechanisms and possible strategies,” *Front. Med.*, vol. 12, p. 1638184, 2025, doi: 10.3389/fmed.2025.1638184.
[47]F. Tariq *et al.*, “Advances in myopia prevention strategies for school-aged children: a comprehensive review,” *Front. Public Health*, vol. 11, p. 1226438, 2023, doi: 10.3389/fpubh.2023.1226438.
[48]L. Raffa, T. Hamdi, N. Balbaid, N. Shikdar, E. Basalem, and N. Almarzouki, “Normal range of ocular biometry in healthy Saudi children,” *Open Ophthalmol. J.*, vol. 18, p. e18743641330233, 2024, doi: 10.2174/0118743641330233240913093516.
[49]Y. Qi, L. Liu, Y. Li, and others, “Factors associated with faster axial elongation after orthokeratology treatment,” *BMC Ophthalmol.*, vol. 22, p. 62, 2022, doi: 10.1186/s12886-022-02294-1.
[50]D. Philipp, M. Vogel, M. Brandt, E. Zimmermann, and B. Jost, “The relationship between myopia and near work, time outdoors and socioeconomic status in children and adolescents,” *BMC Public Health*, vol. 22, p. 2058, 2022, doi: 10.1186/s12889-022-14377-1.
[51]Z. Zong, Y. Zhang, J. Qiao, and Y. Tian, “The association between screen time exposure and myopia in children and adolescents: A meta-analysis,” *BMC Public Health*, vol. 24, p. 1625, 2024, doi: 10.1186/s12889-024-19113-5.
[52]A. Ha, Y. J. Lee, M. Lee, S. R. Shim, and Y. K. Kim, “Digital Screen Time and Myopia: A Systematic Review and Dose-Response Meta-Analysis,” *JAMA Netw. Open*, vol. 8, no. 2, p. e2460026, 2025, doi: 10.1001/jamanetworkopen.2024.60026.
[53]M. Ba and Z. Li, “The impact of lifestyle factors on myopia development: Insights and recommendations,” *AJO Int.*, vol. 1, p. 100010, 2024, doi: 10.1016/j.ajoint.2024.100010.