npj Digital Medicine

https://doi.org/10.1038/s41746-026-02442-7

Article in Press

Robust and interpretable unit level causal inference in neural networks for pediatric myopia

Received: 18 July 2025

Accepted: 5 February 2026

Cite this article as: Jin, Z., Kang, M., Zhao, W. *et al.* Robust and interpretable unit level causal inference in neural networks for pediatric myopia. *npj Digit. Med.* (2026). https://doi.org/10.1038/s41746-026-02442-7

Zihui Jin, Mengtian Kang, Wuyan Zhao, Wenjin Gui, He Li, Yongfang Tu, Yongjun Huo, Canqing Yu, Weihua Song, Ningli Wang, Xu Yang & Shi-Ming Li

Robust and Interpretable Unit Level Causal Inference in Neural Networks for Pediatric Myopia

Zihui Jin[1†], Mengtian Kang[2,3,4†], Wuyan Zhao[1], Wenjin Gui[1], He Li[5], Yongfang Tu[5], Yongjun Huo[5], Canqing Yu[6,7], Weihua Song[8*], Ningli Wang[2,3,4*], Xu Yang[1*], Shi-Ming Li[2,3,4*]

[1*]School of Computer Science and Technology, Beijing Institute of Technology, Beijing, 100081, China.
[2*]Beijing Tongren Eye Center, Beijing Tongren Hospital, Beijing Institute of Ophthalmology, Beijing Key Laboratory of Intelligent Diagnosis Technology and Equipment for Optic Nerve-Related Eye Diseases, Capital Medical University, Beijing, 100005, China.
[3*] National Engineering Research Center for Ophthalmology, Beijing, 100111, China.
[4*]Engineering Research Center of Ophthalmic Equipment and Materials, Ministry of Education, Beijing, 100069, China.
[5]Anyang Eye Hospital, Henan, 455000, China.
[6]Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, 100191, Beijing.
[7]Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, 100191, Beijing.
[8*]Beijing Municipal Geriatric Medical Research Center, Department of Neurology, Xuanwu hospital Capital Medical University, Beijing, 100053, China.

*Corresponding author(s). E-mail(s): liuliyue1118@163.com; wningli@vip.163.com; pyro_yangxu@bit.edu.cn; lishiming81@163.com;
Contributing authors: jinzihui@bit.edu.cn; kangmengtian@163.com;
†These authors contributed equally to this work.

Abstract

Understanding causal mechanisms in deep learning is essential for clinical adoption, where interpretability and reliability are critical. Most existing AI systems act as black boxes, limiting transparency in medicine. We propose a causal inference framework integrated into neural networks to assess the influence of individual features on predictions. Using a prospective pediatric ophthalmology cohort of over 3,000 children with longitudinal follow-up, our method estimates direct and indirect causal effects through intervention. Applied to myopia progression in children, our approach not only achieved good performance but also identified clinically plausible causal pathways. Refutation experiments with multiple falsification strategies confirm the robustness and reliability of causal effects. The approach is model-agnostic and suitable for digital health interventions requiring explainability. By incorporating unit-level causal reasoning into deep learning, this work advances transparent and reliable AI systems aligned with the goals of precision medicine and equitable healthcare.

Keywords: Causal Inference, Neural Network Attribution , Interpretable Deep Learning, Digital Health, Pediatric Ophthalmology

1 Introduction

Myopia has become the most prevalent refractive disorder and a leading cause of vision impairment globally [1, 2]. In East Asia, the adolescent myopia rate has reached alarming levels of 80–90%, with up to 20% developing high myopia—associated with sight-threatening complications such as retinal detachment and macular degeneration [3–5]. The increasing incidence in younger populations calls for a deeper understanding of the multifactorial etiology of myopia [6–8].

Traditional statistical methods, such as linear and logistic regression, have been widely used to explore myopia-related risk factors [9, 10]. However, these models primarily capture correlations and often fail to represent the complex, nonlinear interactions inherent in biological systems [11–13]. Furthermore, few studies incorporate causal frameworks, limiting mechanistic insight and weakening the translational value for designing targeted interventions [14].

With the rise of digital medicine, artificial neural networks (ANNs) offer an opportunity to leverage large-scale, real-world health data for predictive modeling [15–18]. Nevertheless, deep learning models are frequently criticized for their "black-box" nature, making it difficult to explain predictions or guide clinical action [19–23]. To address this limitation, post-hoc interpretability techniques such as SHAP (SHapley Additive exPlanations) have been widely adopted to attribute model outputs to input features [24, 25]. While such methods improve transparency, they primarily provide

correlation-based explanations and do not fully capture the underlying causal mechanisms. Interpretability and reliability are particularly essential in medical settings, where clinical decisions must be transparent, justifiable, and safe.

To bridge this gap, there is growing interest in combining causal inference with neural architectures, enabling not only accurate predictions but also insight into the underlying mechanisms [26–29]. Such approaches are especially valuable for digital interventions, where understanding the impact pathways of behavioral, physiological, or environmental factors can inform scalable and personalized healthcare strategies.

In this study, we propose a causal modeling framework that integrates intervention-based causal attribution within deep neural networks. Using a prospective pediatric ophthalmology cohort tracking over 3,000 children longitudinally, our framework disentangles both direct and indirect causal effects of input features on myopia progression, enhancing model transparency and clinical interpretability.

To achieve this, we categorize input neurons into three functional types—Isolated, Pure, and Confounded Units—based on their positions in the learned causal structure. We design targeted attribution strategies for each, including a domain-adaptive meta-learning approach for estimating causal effects under confounding bias. In addition to robust predictive performance, our model enables the reconstruction of interpretable causal pathways within the network, offering a mechanistic view of how input signals propagate through the architecture.

Importantly, our approach is model-agnostic and can be generalized to other clinical contexts that demand trustworthy, explainable AI. By integrating deep learning, causal reasoning, and real-world validation, our work contributes to the development of reliable digital interventions that meet the interpretive and ethical standards of modern healthcare.

2 Results

2.1 Neurons' Causal Structure Reveals Functional Categories of Input Neurons

The causal explanation framework proposed in this study is model-agnostic, requiring that the prediction model be trained prior to causal analysis. We first constructed a binary classification model for myopia prediction using a feedforward neural network (as shown in Fig. 1b) trained on a pediatric ophthalmic cohort dataset collected via annual surveys in City Anyang, Henan Province, China, from 2011 to 2017 (see Section 4.1). As illustrated in Fig. 1a, the model input consists of 16 variables spanning behavioral,

physiological, dietary, environmental, and genetic domains (Table 1). The result of this model for myopia classification is summarized in Table 2, the model achieved an accuracy of 0.933, indicating strong discriminative ability between the two classes. This result confirms that the binary myopia classification outcomes provide a robust foundation for subsequent causal attribution analyses, ensuring that the interpretability of the model is built upon clinically reliable predictions.

To uncover the underlying causal relationships among input features, we applied a constraint-based causal discovery algorithm combining the PC algorithm [30] with Degenerate Gaussian Scoring and dimensionality reduction via PCA. PCA was used solely as a preprocessing step to mitigate collinearity and improve numerical stability in conditional independence testing, while ensuring that the final causal graph is defined exclusively over the original 16 clinical and physiological variables (as shown in Fig. 1c, d). The final causal graph was constructed using the TETRAD software suite [31]. As shown in Fig. 1e, the resulting directed acyclic graph (DAG) reveals 15 causal edges connecting 10 of the 16 variables. Notably, six variables — parental myopia (PWG), gender (GENDER), near and distance accommodative ability (NAR, DAR), and corneal curvature (K1, K2) — were found to be direct causes of both axial length (AL) and cycloplegic refraction (CR). Height and red meat consumption (REDM) showed direct effects on AL only, while AL was identified as a direct cause of CR.

Based on Fig. 1f, we categorized the corresponding input-layer neurons in the trained neural network into three structurally distinct types: **Isolated Units**, **Pure Units**, and **Confounded Units**. This categorization is not merely descriptive; rather, it enables a principled decomposition of causal contributions within the neural network. By aligning unit types with distinct structural roles in the causal graph, we introduce a unified framework to estimate causal effects under different confounding conditions, thus bridging structural causal inference and neural attribution in a theoretically grounded manner.

Isolated Units (Fig. 1f left one) are input neurons that do not have any causal relation with other input features. Their contributions to the prediction model are assumed to be independent. Six variables fall under this category: CB, EGG, NW, PULSE, WHIM, and DTO.

Pure Units (Fig. 1f middle one) are input neurons that influence the output only through downstream mediators. For example, variable A causally influences variable B, which in turn affects the prediction output. This corresponds to a chain-like causal path. Variables such as PWG, GENDER, NAR, DAR, K1, K2, HEIGHT, and REDM are identified as Pure Units based on their positions in the causal graph.

Confounded Units (Fig. 1f right one) are input neurons whose effect on the output is entangled with that of another causally related variable, forming a v-structure or a backdoor path. In our study, AL and CR are identified as Confounded Units, both influenced by multiple upstream variables and interlinked causally.

This categorization is central to the subsequent causal attribution analysis. By structurally disentangling how input neurons contribute to predictions, we establish a scalable and interpretable framework for assessing causal roles within neural networks in medical applications.

2.2 Causal Attribution Experiments on Isolated Units Reveal Consistency and Deviation from Prior Knowledge

We performed causal attribution experiments on the six Isolated Units identified in the causal graph to assess the individual causal effect of each input-layer neuron on the output predictions of the trained neural network. In this context, attribution refers specifically to estimating the average treatment effect (ATE) of a single input neuron on each output class in the binary classification task (myopic vs. non-myopic).

Each Isolated Unit was treated as an intervention variable and directly input into a causal attribution algorithm to quantify its effect. Since the model produced a binary prediction for myopia status (Yes/No), we conducted separate interventional analyses on both output classes.

Pulse rate (PULSE) has been previously reported to be negatively associated with myopia risk [32–35], possibly due to increased ocular blood flow influencing eye growth. As shown in Fig. 2a, increasing the intervention value of PULSE leads to a rising trend in the 'No' (non-myopic) output neuron's ATE and a decreasing trend in the 'Yes' (myopic) output neuron's ATE. This suggests a negative causal relationship between PULSE and myopia, in agreement with prior clinical findings and indicating that the model has accurately captured this relationship.

In contrast, carbonated beverage consumption frequency (CB) and egg consumption frequency (EGG) have been reported to be positively associated with myopia[36–39]. However, as shown in Fig. 2b, c, the model exhibits an inverted trend: CB shows a positive effect on the 'No' neuron and a negative effect on the 'Yes' neuron, while EGG shows the opposite pattern. These observations indicate that the model's learned causal behavior for these features is inconsistent with established domain knowledge. A likely explanation is the uneven distribution of CB and EGG consumption across subgroups in our dataset, which may have influenced the model's learned patterns.

In summary, among the six Isolated Units, the model correctly captures causal trends for PULSE and DTO, but misrepresents those for NW, CB, and EGG. The effect of WHIM appears negligible and is excluded from further interpretation [40–45]. Detailed attribution results of variables other than CB, EGG and PULSE are shown in Supplementary material Figure S1-S3.

2.3 Causal Attribution on Pure Units Confirms Model Reliability Across Mediated Paths

For Pure Units, causal attribution can be conducted without adjusting for additional covariates, as their causal effects are either direct or fully mediated by identified intermediate variables. The attribution process involves identifying all causal paths—both direct and indirect—between the input neuron and the output neuron. Assuming linear causal relationships among the variables in the dataset, we performed causal inference using linear regression applied to the structure encoded in the causal graph. The resulting causal effect graph is shown in Fig. 3a,b, where solid lines represent positive causal effects and dashed lines indicate negative effects.

Two representative examples are shown in Fig. 3c, d, where we analyze the effect of height (HEIGHT) and gender (GENDER). As the intervention value of HEIGHT increases, the response of the 'No' (non-myopic) output neuron shows a decreasing trend, while the 'Yes' (myopic) output neuron shows an increasing trend. This trend is consistent with previous research findings that people with greater height are more likely to suffer from myopia[46–52]. It should be emphasized, however, that the causal pathway identified by our framework is HEIGHT → AL → CR, rather than a direct HEIGHT → Myopia link. This indicates that height acts indirectly through ocular growth and refractive status, and should not be interpreted as a direct or modifiable risk factor for myopia. The trend after the intervention on gender (GENDER) is the same as previous findings, that is, among adolescents, the incidence of myopia in females is higher than that in males[53–60]. However, such associations may vary with individual-level heterogeneity.

Among the eight Pure Units, the model correctly reflects the causal trends for NAR, DAR, K1, K2, HEIGHT, GENDER, and PWG[61–89]. Notably, while axial length (AL) is widely recognized as the primary anatomical determinant of myopia and a direct driver of cycloplegic refraction (CR), several clinical studies have reported that accommodative function—including near and distance accommodative ability (NAR and DAR)—may also contribute to AL elongation and myopia progression [90–93]. Therefore, the NAR/DAR → AL path revealed by our framework should be interpreted

as a data-driven attribution pattern that aligns with certain clinical observations, while requiring further validation. Detailed attribution results of variables other than HEIGHT and GENDER are shown in Supplementary material Figure S4-S7. The effect of REDM remains inconclusive in the existing literature and is therefore not interpreted in terms of fit correctness. However, our method revealed a weak but directional causal path from REDM to AL (as shown in Fig. 3b $S_5 \rightarrow S_9$), followed by a stronger negative causal influence from AL to CR (as shown in Fig. 3b $S_9 \rightarrow S_{10}$). Although the direct effect of REDM on AL is marginal, as shown in Fig. 3e, the existence of this indirect pathway suggests a potential mediating mechanism by which dietary patterns might influence refractive development. This finding, uncovered through our causal attribution framework, suggests a hypothesis-generating signal that warrants further investigation in nutritional ophthalmology and longitudinal dietary studies.

2.4 Meta-Learning-Based Attribution for Confounded Units Demonstrates Stable Causal Estimation

To robustly assess causal effects under confounding, especially in cases where input neurons share common ancestors with output nodes, conventional regression-based methods often yield biased estimates. To address this challenge, we introduce a domain-adaptive meta-learning algorithm that leverages covariate balancing through propensity scores. The key idea is to learn a causal representation that generalizes across input distributions, enabling unbiased estimation even when confounders are present. This framework not only corrects for selection bias but also enhances robustness under covariate shift—making it well-suited for causal inference in complex medical prediction tasks. Given that these input neurons are causally entangled with others and subject to confounding with the output layer, we first identified and controlled for confounding variables using the backdoor criterion, enabling the estimation of unbiased causal effects.

To evaluate the performance of different causal effect estimation models, we adopt the **Rscore** metric based on the R-Learner framework. The Rscore is defined as:

$$\text{Rscore} = 1 - \frac{\hat{L}_R}{L_{\text{base}}} \tag{1}$$

where $\hat{L}_R$ is the R-loss obtained from cross-validation using the estimated treatment effect model $\hat{\tau}(x)$, and L_{base} is the baseline loss computed using a constant average treatment effect.

This metric assesses the ability of the estimated individualized treatment effect to explain residual variation, in comparison to a constant effect. A higher Rscore indicates better performance in modeling heterogeneous treatment effects. Our proposed meta-learning model achieved superior accuracy, with the best result obtained using a GBDT-based estimator (Rscore = 0.3850). Further improvements were achieved via an ensemble method combining GBDT and Random Forest with a 10:1 weight ratio, yielding an Rscore of 0.6277, indicating strong attribution performance(see Supplementary material for details).

To mitigate bias introduced by discretizing continuous variables, we applied both equal-width and equal-frequency discretization strategies to the two Confounded Units—axial length (AL) and cycloplegic refraction (CR)—and estimated their average causal effects (ACE) on the model's binary outputs ('No' and 'Yes'). As shown in Fig. 4a, c, AL exhibits a positive causal effect on myopia risk, consistent with established clinical findings[94–96]. Similarly, CR shows a pattern where lower refractive power corresponds to increased myopia risk[97–99]. Equal-width discretization provides a clearer causal trend compared to equal-frequency discretization.

Importantly, the causal attribution trends for both AL and CR are aligned with current medical understanding, indicating that the model correctly internalized known relationships between these variables and myopia.

2.5 Refutation Experiments Validate Causal Robustness

To evaluate the internal validity and robustness of our estimated causal effects, we conducted a series of refutation experiments inspired by the DoWhy framework [100]. These experiments were designed to determine whether our results hold under systematic perturbations to data and model assumptions.

Five types of intervention strategies were employed: bootstrap resampling, random addition of artificial confounders, subset validation, placebo treatment, and dummy outcome substitution.

The results of the Pure Units refutation experiment are shown in Fig. 3f. It can be seen from the figure that the error rates (err%) obtained in the five refutation experiments are all less than 10%, and there is no situation where the error rates of the five refutation experiments are all higher than 1%.

The results of the Confounded Units refutation experiment are shown in Fig. 4b, d. For AL, equal-frequency discretization produced lower refutation error, while CR showed greater stability with equal-width discretization. These results collectively demonstrate

that the model is capable of producing accurate and stable causal attributions even under confounding conditions.

3 Disscusion

In this study, we present a causal learning–based interpretability framework for neural networks, tailored to the clinical prediction of pediatric myopia. By abstracting a feedforward neural network into a structured causal system, we quantify the causal contribution of each input neuron to the model output and implement targeted attribution strategies according to the neuron's structural role in the learned causal graph. Using a longitudinal dataset collected in City Anyang, Henan Province, China (2011–2017), our method integrates causal discovery, mediation analysis, and domain-adaptive estimation to produce interpretable and robust neural reasoning.

The inferred causal graph, constructed using a constraint-based algorithm (PC with Degenerate Gaussian Scoring), revealed stable directed dependencies among the 16 input features. Based on this structure, we categorized input neurons into three functionally distinct types—Isolated, Pure, and Confounded Units—each representing a unique causal configuration. Attribution results for Isolated Units aligned with established risk factors in several cases: for instance, increased pulse rate and outdoor activity were found to have protective effects against myopia, consistent with previous studies[101]. In contrast, misalignment was observed for some dietary variables (e.g., carbonated beverage and egg consumption), suggesting either data limitations or prior model biases. Pure Units, evaluated through linear path-based inference, demonstrated biologically plausible effects—e.g., the positive causal influence of height on myopia risk was found to be mediated through axial elongation. For Confounded Units such as axial length (AL) and cycloplegic refraction (CR), our meta-learning–based estimation—guided by the backdoor criterion—produced coherent and stable causal effects. This was further validated by a suite of counterfactual perturbation experiments.

These results demonstrate that black-box neural networks can be decomposed into biologically meaningful, causally interpretable substructures[102, 103]. The model's ability to internalize known physiological mechanisms, particularly around AL and CR—which directly modulate the eye's focal characteristics—confirms its biological plausibility. Overall, the causal pathways identified by our framework demonstrate strong biological plausibility and align closely with established clinical knowledge. Throughout the study, we placed particular emphasis on clinical interpretability, and for each major result we cross-referenced supporting ophthalmic literature to ensure consistency with prior findings. We fully acknowledge the central role of axial length

(AL) as the dominant anatomical determinant of myopia, and our model's detection of the AL → CR relationship is fully consistent with this consensus. At the same time, our framework is inherently data-driven: the causal paths it reveals should be understood as internal attribution patterns within the predictive model, which require cautious interpretation in light of existing clinical evidence. Notably, we found that many of these data-driven paths are corroborated by previous studies, reinforcing the clinical validity of our approach. Compared to conventional attribution methods like SHAP or LIME[104–108], which often fail under confounding or produce unstable saliency maps, our causal framework ensures interpretability by explicitly modeling conditional dependencies and causal paths. This not only improves explanation reliability but also enables insight into systemic bias or misattribution patterns—critical for deploying AI safely in clinical settings.

Importantly, our method has direct implications for the design of digital interventions and early prevention strategies. The identification of modifiable behavioral risk factors, such as outdoor activity, as causal precursors to myopia progression suggests that individualized lifestyle adjustments could be guided by our model outputs. Additionally, we identify discrepancies between learned model logic and medical consensus, which could be used to iteratively improve neural architectures or training protocols, enhancing both performance and clinical acceptability.

At the same time, we note that several features exhibited attribution patterns that diverged from previously established clinical or epidemiological knowledge. These discrepancies may, at least in part, arise from distributional characteristics of the dataset—such as uneven representation of certain behaviors across age or socioeconomic subgroups—which can shape the model's learned associations. It is important to emphasize that our causal framework is primarily intended to disentangle and visualize the internal attribution logic of predictive models, rather than to fully correct for population-level confounding. Consequently, while the framework enhances interpretability and highlights biologically plausible pathways, residual distributional imbalances may still produce effects inconsistent with consensus knowledge. We regard this as a constructive signal, since such mismatches can reveal potential data artifacts or, alternatively, point to novel hypotheses warranting further validation.

Despite these promising findings, several limitations remain. First, the accuracy of the causal discovery process relies on the completeness and quality of observational data. In particular, the presence of missing covariates or unobserved confounders may bias the inferred graph structure. Furthermore, our method was validated on complete longitudinal data, which may not fully reflect real-world scenarios where data are often

incomplete or irregularly sampled. Using complete cases may introduce selection bias if missingness is not completely at random. To address incomplete data in future applications, our framework could be extended through multiple imputation techniques for missing values before causal discovery, or through modified causal discovery algorithms that handle missing data directly. Second, our mediation model currently assumes linear causal effects, while robustness is verified through counterfactual tests, highly nonlinear interactions may not be fully captured. Kernel-based or neural causal inference methods may be needed to improve generalizability. Third, while this study chose a feedforward architecture, our framework is designed to be model-agnostic due to the decoupling of causal discovery from neural attribution. In practice, extending this method to other architectures would primarily involve substituting the attribution module while maintaining the structural causal categorization logic. For instance, when applying the framework to convolutional neural networks (CNNs) for retinal imaging, input units could be defined as high-level features extracted from specific regions of interest, and attribution could be estimated via Grad-CAM or Integrated Gradients. For transformer-based models handling genomic sequences, the nodes could represent token embeddings or attention heads, utilizing Attention Rollout or perturbation-based attribution to quantify causal contributions.

Looking ahead, our work opens several directions for innovation. (1) Embedding causal graphs into neural network training could lead to structure-aware architectures that align model connectivity with known biomedical pathways. (2) Extending our framework to cross-modal causal interpretation—bridging electronic health records with imaging or sensor data—would further improve clinical relevance. Such an extension would leverage the framework's modularity, where the universal "causal skeleton" derived from data governs the interpretation of diverse neural components across different sub-networks. (3) As causal attributions can highlight mislearned associations, they may serve as real-time monitors for algorithmic fairness and bias. (4) With temporally dense datasets, our framework can evolve into dynamic causal modeling, enabling the simulation of patient-specific interventions for myopia prevention or progression control.

In conclusion, we introduce a rigorous and generalizable framework for unit-level causal attribution in neural networks, validated in a clinically relevant setting. By focusing on the interpretation of the model in causal principles, we offer a pathway toward more transparent, robust, and trustworthy AI systems for digital medicine.

4 Methods

4.1 Dataset

We used the Dataset of a prospective longitudinal school-based cohort study—the Anyang Childhood Eye Study(ACES)[109], one of the largest longitudinal studies of myopia in school-aged children in China. The dataset was collected through annual follow-up surveys from 2011 to 2017 and included 3,112 first-grade students at baseline, who were followed through sixth grade (school grades 1–6, spanning six academic years). Each year corresponded to a single visit per participant, with examinations conducted once annually at their schools. At baseline, participants had a mean age of 7.1 years (range: 5.7–9.3 years), and 57.8% were male. Data collection was conducted in collaboration with local primary schools under the supervision of certified medical professionals and epidemiologists. The study received ethics approval from the Institutional Review Board of Beijing Tongren Hospital, Capital Medical University, and informed consent was obtained from all participants and their guardians. Written informed consent was obtained from all participants and their parents or legal guardians.

All children provided a written informed consent form signed by their parents, and verbal consent was also obtained from each child. This study adhered to the tenets of the Declaration of Helsinki. Ethics committee approval was obtained from the Institutional Review Board of Beijing Tongren Hospital, Capital Medical University (TRECKY2018-030).

The survey captured a broad range of data across behavioral, physiological, environmental, dietary, and hereditary domains. After preprocessing, 2,748 records with complete and valid longitudinal data were retained for analysis (88.3% of baseline participants). These records spanned a 6-year follow-up period, enabling the construction of a rich temporal feature space suitable for causal discovery and model training. In this survey, myopia was classified as a binary outcome (0 = non-myopia, 1 = myopia) according to cycloplegic refraction (CR > -0.5D for non-myopia, CR <= -0.5D for myopia). The myopia classification was determined based on whether the child developed myopia at any point during the 6-year follow-up period, rather than solely at the last visit, thereby capturing incident myopia cases throughout the study period.

To minimize potential methodological bias, different preprocessing strategies were applied depending on the temporal characteristics of each variable. For dynamic ophthalmic measures (e.g., CR), we derived annual progression rates rather than simple averages. For behavioral and dietary variables (e.g., NW, DTO, CB, EGG, REDM, WHIM), cumulative values across six years were used to capture long-term exposure.

For relatively stable biometric and physiological measures (e.g., NAR, DAR, PULSE, AL, K1, K2), six-year averages were computed to reduce random fluctuations and measurement error. Static variables such as gender and parental myopia (PWG) were retained as baseline values.

After data preprocessing (see Supplementary material for details), 16 variables were finally selected and used as input for model training and causal attribution. A full list of selected variables and their definitions is provided in Table 1.

Table 1: Variables of the pediatric myopia dataset from the Anyang Childhood Eye Study

VARIABLE	VARIABLE MEANING	DATA TYPE	DATA RANGE	UNITS
CR	cycloplegic refraction	continuous	-5.4-8.7	diopter(D)
AL	axial length	continuous	20-34	mm
NW	total nearwork load	continuous	0-102	hour(h)
DTO	distance-viewing time outdoors	continuous	0-57	hour(h)
NAR	near accommodative ability of right eye	continuous	-5.1-2.2	diopter(D)
DAR	distance accommodative ability of right eye	continuous	-6.6-4.4	diopter(D)
HEIGHT	height	continuous	97-143	cm
PULSE	pulse rate	continuous	52-140	Bpm
GENDER	gender	binary	0,1	0 represents a girl 1 represents a boy
CB	carbonated beverage consumption frequency	discrete	1-5	times per week
EGG	egg consumption frequency	discrete	1-4	times per week
REDM	red meat consumption frequency	discrete	1-5	times per week
WHIM	white meat consumption frequency	discrete	1-5	times per week
PWG	number of parents wearing glasses	discrete	0-2	person
K1	corneal keratometry in both eyes	continuous	38-71	mm
K2	corneal keratometry in both eyes	continuous	38-183	mm
Myopia	myopia or not	binary	0,1	0 represents non-myopia 1 represents myopia

4.2 Neural Network Architecture

We constructed a feedforward neural network (FNN) model to predict myopia status based on the preprocessed features derived from the dataset. The network architecture consists of an input layer with 16 neurons corresponding to the selected variables, two

fully connected hidden layers with 12 and 8 neurons respectively, and an output layer with 2 neurons representing binary classification outcomes (myopia vs. non-myopia). The network structure is illustrated in Fig.1b.

Each neuron in the hidden layers uses the Rectified Linear Unit (ReLU) as the activation function to introduce non-linearity. The final output layer employs the softmax function to produce probability distributions over the two classes. The model is trained using the cross-entropy loss function, and optimized via stochastic gradient descent with backpropagation.

To identify optimal model performance, we conducted a series of hyperparameter tuning experiments, adjusting the learning rate, batch size, and number of training epochs. The best-performing configuration was selected based on validation accuracy. Final performance metrics of the four-layer FNN model on the test set are reported in Table 2, demonstrating high accuracy and robustness across multiple evaluation criteria.

To benchmark performance, we also implemented five traditional machine learning models—Logistic Regression, Naive Bayes, Random Forest, K-Nearest Neighbors, and Support Vector Machine—using the same input features. As shown in Table 2, the 4-layer FNN achieved consistently high performance across all evaluation metrics. Its sensitivity was comparable to that of SVM and Logistic Regression, all of which reached the highest values among the tested models (within the allowable statistical error), thereby reducing the risk of missed myopia cases. Compared with other baseline models, the FNN also provided a more balanced trade-off between sensitivity and specificity, indicating its robustness in distinguishing myopia versus non-myopia. Furthermore, given its representative "black-box" nature, the FNN serves as an appropriate foundation for our proposed causal explanation framework, which aims to move beyond predictive accuracy to provide interpretable and clinically meaningful insights.

Table 2: Performance metrics of the four-layer feedforward neural network and traditional machine learning models

Model	Accuracy	Sensitivity	Specificity	F1 Score	AUC
4-layer FNN	0.935 ± 0.011	0.939 ± 0.010	0.927 ± 0.019	0.947 ± 0.009	0.976 ± 0.007
SVM	0.863 ± 0.008	0.867 ± 0.013	0.854 ± 0.016	0.888 ± 0.007	0.921 ± 0.014
Logistic Regression	0.935 ± 0.016	0.941 ± 0.009	0.926 ± 0.030	0.948 ± 0.013	0.977 ± 0.008
Naive Bayes	0.845 ± 0.018	0.783 ± 0.027	0.949 ± 0.011	0.864 ± 0.018	0.948 ± 0.010
K-Nearest Neighbors	0.922 ± 0.012	0.936 ± 0.015	0.899 ± 0.011	0.938 ± 0.010	0.978 ± 0.008
Random Forest	0.936 ± 0.014	0.940 ± 0.007	0.928 ± 0.027	0.948 ± 0.011	0.977 ± 0.009

4.3 Causal Discovery

Causal discovery, a key approach for inferring causal relationships from observational data via graphical models like DAGs, enables researchers to uncover intervention-relevant structures in the absence of randomized trials[110]. In this study, we adopted a constraint-based causal discovery approach, specifically the **Peter-Clark (PC) algorithm**[30], to identify causal relationships among the 16 input variables of the myopia prediction model. The PC algorithm begins with a fully connected undirected graph and iteratively removes edges by testing conditional independence. It then determines edge directions using v-structure identification and Meek rules[111], producing a partially directed acyclic graph (CPDAG) that represents the underlying causal structure.

A critical challenge in biomedical datasets is the presence of mixed-type variables, including both continuous and categorical features. Traditional independence tests (e.g., Pearson correlation, Chi-squared test, or Fisher's Z-test) are limited to continuous or discrete data and often fail in mixed domains.

To address this, we integrated the **Degenerate Gaussian Likelihood Ratio Test (DG-LRT)** into the PC algorithm. This method originally proposed by Andrews[112, 113] and refined for mixed data, uses a transformed representation of categorical variables via one-hot encoding, followed by degenerate Gaussian modeling. The resulting structure is especially well-suited for medical applications where variable types and scales vary widely.

This design highlights the cross-disciplinary adaptability. It is capable of capturing statistically reasonable causal structures across hybrid domains, bridging data mining, statistical learning, and clinical reasoning.

For stability and completeness, the causal discovery process was conducted on the entire dataset rather than a training split, ensuring that the learned structure reflected global dependencies among variables.

The final causal graph extracted from our dataset included 10 out of the 16 input variables, forming 15 directed edges (as shown in Fig. 1e). This structure serves as the backbone for following causal attribution and interpretation. The detailed steps of the algorithm are presented in Algorithm 1.

Algorithm 1: PC Algorithm with Degenerate Gaussian Likelihood Ratio Test

Input: Dataset D with mixed-type variables

Output: Causal graph G

Initialize G as a complete undirected graph over all variables;

Set separation sets $S(i, j) \leftarrow \emptyset$ for all pairs (i, j);

for $l = 0$ *to maximum neighborhood size* **do**

  foreach *variable pair* (X_i, X_j) *with* $|Adj(X_i) \setminus \{X_j\}| \geq l$ **do**

    foreach *subset* $Z \subseteq Adj(X_i) \setminus \{X_j\}$ *with* $|Z| = l$ **do**

      if $X_i \perp\!\!\!\perp X_j \mid Z$ *using DG-LRT* **then**

        Remove edge (X_i, X_j) from G;

        Set $S(i, j) \leftarrow Z$;

Identify v-structures based on separation sets;

Apply Meek rules to orient remaining edges;

return G;

4.4 Attribution Algorithm

In this study, causal inference is conducted at two levels: estimating the effect of input-layer neurons on output neurons within a neural network (i.e., attribution), and evaluating inter-feature causal effects in the dataset using regression-based techniques.

Causal effects can be defined at different levels. Assuming binary treatment $T \in \{0, 1\}$, the most commonly used is:

Average Treatment Effect (ATE):

$$\text{ATE} = \mathbb{E}[Y(1)] - \mathbb{E}[Y(0)]$$

Isolated Units. Isolated Units refer to specific input neurons in a neural network whose causal effects on output neurons are to be independently assessed. Based on the operator framework introduced in Section *Causal Inference*, we formalize the average causal effect (ACE) from input neuron x_i to output neuron y, as shown in Fig. 1f left one.

For a binary variable x, its ACE on another variable y is defined as:

$$\text{ACE}(x) = \mathbb{E}[y \mid \text{do}(x = 1)] - \mathbb{E}[y \mid \text{do}(x = 0)] \tag{2}$$

For continuous-valued variables, we extend this to:

$$\mathrm{ACE}_{x_i} = \frac{1}{high^i - low^i} \int_{low^i}^{high^i} \mathbb{E}[y \mid \mathrm{do}(x_i = \alpha)]\, d\alpha \tag{3}$$

To mitigate the inefficiency and variance introduced by high-dimensional sampling, we approximate the intervention expectation using a second-order Taylor expansion of the network function f.

Let μ be the mean input vector, and $f(\mu)$ be the smooth neural network function. For a small perturbation δ in the input, the Taylor expansion around μ is:

$$f(\mu + \delta) \approx f(\mu) + \nabla f(\mu)^\top \delta + \frac{1}{2} \delta^\top \nabla^2 f(\mu) \delta \tag{4}$$

Taking expectation over δ where $\mathbb{E}[\delta] = 0$ leads to:

$$\mathbb{E}[y \mid \mathrm{do}(x_i = \alpha)] \approx f(\mu) + \frac{1}{2} \mathrm{tr}\left(\nabla^2 f(\mu) \cdot \mathrm{Cov}\right) \tag{5}$$

This allows efficient estimation of intervention effects without full re-evaluation over the dataset.

The algorithm estimates the causal effect of input neuron x_i on the output neuron y over a specified intervention interval $[low^i, high^i]$ by dividing it into num uniform sub-intervals.

The full procedure is shown in Algorithm 2.

Algorithm 2: Approximate Causal Effect for Isolated Input Unit

Input: Output neuron y, target input neuron x_i, intervention interval $[low^i, high^i]$, number of intervals num, mean input vector μ, input covariance matrix Cov, neural network function $f(\cdot)$

Output: Intervention expectation $E[y \mid \text{do}(x_i)]$ for each $\alpha \in [low^i, high^i]$

Initialize $Cov[x_i][:] \leftarrow 0$, $Cov[:][x_i] \leftarrow 0$;

Initialize $intervention\_expectation \leftarrow [\ ]$;

Set $\alpha \leftarrow low^i$;

while $\alpha \leq high^i$ **do**

  $\mu[i] \leftarrow \alpha$;

  $e \leftarrow f(\mu) + \frac{1}{2} \cdot \text{tr}(\nabla^2 f(\mu) \cdot Cov)$;

  Append e to $intervention\_expectation$;

  $\alpha \leftarrow \alpha + \frac{high^i - low^i}{num}$;

return $intervention\_expectation$;

Pure Units. This section discusses how to attribute causal effects for input neurons that are causally dependent on other inputs but are not affected by confounders—referred to as Pure Units. We assume these neurons are connected to other input neurons via causal pathways, but no common ancestor influences both the input and output neurons (i.e., no confounding exists).

As illustrated in Fig. 1f middle one, let l_{11} be the intervened input neuron and l_{n1} be the output neuron. In this structure, l_{11} can influence l_{n1} through multiple paths, such as the direct path P_1 and an indirect path P_2 via intermediate neuron l_{12}, but no confounders are present.

In this scenario, the causal effect of l_{11} on l_{n1} includes both direct and indirect components:

$$\text{ACE}(l_{11} \to l_{n1}) = \text{ACE}_{P_1} + \text{ACE}_{P_2} + \cdots \tag{6}$$

Each term corresponds to a causal path, and the effect along each path can be estimated as the product of edge weights along that path. For instance, the average causal effect along the indirect path P_2 from $l_{11} \to l_{12} \to l_{n1}$ is given by:

$$\text{ACE}_{P_2} = (\mathbb{E}[l_{12} \mid \text{do}(l_{11} = \alpha_2)] - \mathbb{E}[l_{12} \mid \text{do}(l_{11} = \alpha_1)]) \cdot \beta_{l_{n1} \sim l_{12}} \tag{7}$$

where $\beta_{l_{n1} \sim l_{12}}$ denotes the regression coefficient from l_{12} to l_{n1}. Equation 7 reflects the mediated effect of l_{11} on l_{n1} through l_{12}. The direct effect $\beta_{l_{n1} \sim l_{11}}$ accounts for path P_1.

Once all relevant paths are identified and estimated, we update the mean vector μ and covariance matrix Σ of the training data to reflect the intervened distribution. The expected effect of l_{11} on l_{n1} under intervention $\mathrm{do}(l_{11} = \alpha)$ is then computed using the intervention expectation approximation from the previous subsection:

$$\mathbb{E}[l_{n1} \mid \mathrm{do}(l_{11} = \alpha)] \approx f(\mu) + \frac{1}{2} \cdot \mathrm{tr}\left(\nabla^2 f(\mu) \cdot \Sigma\right) \tag{8}$$

This method effectively aggregates the impact of multiple causal paths while maintaining computational efficiency. By correcting for potential bias induced by intermediaries, it enables robust causal interpretability in neural networks.

The full procedure is shown in Algorithm 3.

Algorithm 3: Estimate Causal Effect for Pure Input Unit

Input: Output neuron y;
Target input neuron x_i;
Set of causal paths $\{P_1, P_2, \ldots, P_k\}$ from x_i to y;
Intervention values α_1, α_2 for x_i;
Neural network function $f(\cdot)$;
Dataset D
Output: Estimated average causal effect $\mathrm{ACE}(x_i \to y)$
Initialize $total\_effect \leftarrow 0$;
foreach *path* $P_j \in \{P_1, \ldots, P_k\}$ **do**
  Identify intermediate neuron z along path P_j such that $x_i \to \cdots \to z \to y$;
  foreach $\alpha \in \{\alpha_1, \alpha_2\}$ **do**
    Set input vector μ with $x_i \leftarrow \alpha$ and other inputs fixed at their mean;
    Feed forward through $f(\cdot)$ to compute $\mathbb{E}[z \mid \mathrm{do}(x_i = \alpha)]$;
  Compute $\Delta z \leftarrow \mathbb{E}[z \mid \mathrm{do}(x_i = \alpha_2)] - \mathbb{E}[z \mid \mathrm{do}(x_i = \alpha_1)]$;
  Estimate regression coefficient $\beta_{y \sim z}$ using training data D;
  Compute $path\_effect \leftarrow \Delta z \cdot \beta_{y \sim z}$;
  Update $total\_effect \leftarrow total\_effect + path\_effect$;
return $total\_effect$

Confounded Units. In real-world neural networks, causal dependencies among input neurons are often interleaved with hidden confounding variables. A confounder is a variable that simultaneously influences both the intervention (input neuron) and the outcome (output neuron), thereby creating backdoor paths that induce selection bias and invalidate naive causal estimation.

As shown in Fig. 1f right one, the input neuron l_{11} affects output neuron l_{n1} via the direct causal path $P_1 : l_{11} \to l_{n1}$. However, the presence of neuron l_{12}, which also

influences l_{n1} and is influenced by l_{11}, forms a backdoor path $l_{11} \leftarrow l_{12} \rightarrow l_{n1}$. This makes l_{12} a confounder and it needs to be adjusted to avoid biased estimation of the average treatment effect (ATE) of l_{11} on l_{n1}.

To estimate unbiased causal effects in the presence of confounders, we adopt a four-stage framework that integrates graphical causal identification, propensity-based domain adaptation, and ensemble-based meta-learning[114]. Specifically:

Confounder Identification: Based on a predefined causal graph, we identify a sufficient set of covariates Z that satisfy the backdoor criterion, thereby blocking all spurious associations between the treatment variable X and the outcome Y.

Propensity Score Estimation: We estimate the propensity score $e(z) = P(X = 1 \mid Z = z)$ using a supervised learning model. These scores are subsequently used to reweight samples and mitigate distributional imbalance across treatment groups, enabling domain adaptation.

Ensemble Meta-Learning for Outcome Modeling: To model potential outcomes, we train ensemble-based predictors using gradient boosted decision trees (GBDT) and random forests (RF), combined with a fixed weight ratio of 10:1. These models are trained separately on the treatment and control groups with inverse-propensity weighting to account for confounding.

Causal Effect Estimation: Individual treatment effects (ITEs) are approximated by computing residuals between observed and predicted counterfactual outcomes. A final regression model is fitted to these residuals to estimate the conditional average treatment effect $\hat{\tau}(x)$. Prior to modeling, continuous covariates such as axial length (AL) and cycloplegic refraction (CR) are discretized using both equal-width and equal-frequency discretization to satisfy the discrete input requirements of the meta-learning model.

Algorithm 4: Domain-Adaptive Causal Attribution via Ensemble-based Meta-Learning

Input: Input-output pairs (X, Y), confounders Z, binary intervention $X \in \{0, 1\}$

Output: Estimated conditional average treatment effect $\hat{\tau}(x)$

Step 1: Propensity Score Estimation

Train a model to estimate $e(z) = P(X = 1 \mid Z = z)$ using observed confounders Z;

Step 2: Ensemble Outcome Modeling (Meta-Learner)

Train ensemble predictors (GBDT + RF, weight ratio 10:1) with inverse propensity weighting:

- For control group $(X = 0)$:
  $\hat{\mu}_0(x) \leftarrow \texttt{EnsembleModel}(Y^0 \sim X^0, \text{weight} = \frac{e(X^0)}{1-e(X^0)})$
- For treatment group $(X = 1)$:
  $\hat{\mu}_1(x) \leftarrow \texttt{EnsembleModel}(Y^1 \sim X^1, \text{weight} = \frac{1-e(X^1)}{e(X^1)})$

Step 3: Residual-based Individual Treatment Effect Estimation

foreach *sample i in treatment group* **do**
  $\hat{D}_1^i \leftarrow Y_1^i - \hat{\mu}_0(X_1^i)$

foreach *sample j in control group* **do**
  $\hat{D}_0^j \leftarrow \hat{\mu}_1(X_0^j) - Y_0^j$

Step 4: Final Treatment Effect Modeling

Discretize continuous covariates (e.g., AL, RA) via equal-width and equal-frequency binning;

Fit regression model: $\hat{\tau}(x) \leftarrow \texttt{MetaRegressor}(\hat{D}_0 \cup \hat{D}_1 \sim X_0 \cup X_1)$;

return $\hat{\tau}(x)$

4.5 Refutation Experiments

To evaluate the robustness and reliability of the estimated causal effects, we conduct a series of refutation experiments following the methodology proposed in DoWhy [100]. These experiments are designed to examine whether the estimated effect is stable under a variety of controlled perturbations. Specifically, five types of falsification strategies are applied:

Bootstrap Validation (BV): This method uses resampling with replacement from the original dataset to generate synthetic bootstrap datasets. A reliable causal effect estimator should produce consistent estimates across these samples. The error rate is defined as:

$$ERR_{BV} = \left| \frac{NewEffect - EstimatedEffect}{EstimatedEffect} \right| \tag{9}$$

Add Random Common Cause (ARCC): This method randomly introduces artificial confounding variables into the dataset. A robust estimator should show minimal change in causal effect when noise variables are added. The deviation indicates sensitivity to confounding bias. The error rate is calculated as:

$$ERR_{ARCC} = \left| \frac{NewEffect - EstimatedEffect}{EstimatedEffect} \right| \tag{10}$$

Data Subsets Validation (DSV): This method randomly selects a subset of the data as the new evaluation set, and recomputes the causal effect. A stable estimation algorithm should yield similar results across data splits. The error rate is defined as:

$$ERR_{DSV} = \left| \frac{NewEffect - EstimatedEffect}{EstimatedEffect} \right| \tag{11}$$

Placebo Treatment (PT): In this test, the original treatment variable is replaced by a randomly permuted variable (i.e., placebo). Since there is no actual treatment effect, a valid causal estimator should report an effect close to zero. The error rate is defined as:

$$ERR_{PT} = |NewEffect| \tag{12}$$

Dummy Outcome (DO): In this setting, the outcome variable Y is replaced with a randomly generated variable. If the causal inference method is valid, it should yield an estimated causal effect near zero. The error rate is:

$$ERR_{DO} = |NewEffect| \tag{13}$$

These falsification strategies provide a comprehensive suite of counterfactual diagnostics. A reliable causal estimation method should demonstrate stability across bootstrap samples, insensitivity to irrelevant variables, robustness to subsampling, and null effects under placebo or dummy variable conditions.

5 Data Availability

The datasets analyzed in the current study are not publicly available due to patient privacy purposes but are available upon reasonable request to the corresponding author Shi-Ming Li(lishiming81@163.com).

6 Code Availability

The code can be made available upon reasonable request to the corresponding author Xu Yang(pyro_yangxu@bit.edu.cn).

7 Acknowledgments

This research was funded by the National Key Research and Development Program of China 2025YFE0101231. This project was funded by the Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education (Grant NO.2025102). This project was funded by the National Natural Science Foundation of China (82471113). This project was funded by the Beijing Natural Science Foundation (L248023). This project was funded by the Excellent Youth Talents Program of Capital Medical University (A2307). This project was funded by the Capital health research and development of special grant (2024-2G-1081). This project was funded by the Beijing New-star Plan of Science and Technology Cross-cooperation Project (20250484983).

8 Author Contributions

Conceptualization: X.Y., Z.J., and W.S.; methodology: Z.J., and M.K.; investigation: Z.J. and W.Z., W.G., and W.S.; visualization: W.Z. and Z.J.; data collection: H.L., Y.T., and Y.H.; data analysis: N.W., and S.L.; funding acquisition: X.Y. , C.Y., and S.L.; project administration: X.Y. and W.S.; supervision: X.Y.; writing – original draft: Z.J., M.K., and W.Z.; writing – review & editing: all authors

9 Declaration of Interests

The authors declare no competing interests.

10 Supplementary material

Supplementary material can be found in the submission materials.

References

[1] Liang, J.: Trend and projection of myopia in children and adolescents from 1990 to 2050: A comprehensive systematic review and meta-analysis". Br J Ophthalmol **362**, 109 (2025)

[2] Lee, Y., Keel, S., Yoon, S.: Evaluating the effectiveness and scalability of the world health organization myopiaed digital intervention: Mixed methods study. JMIR Public Health and Surveillance **10**(1), 66052 (2024)

[3] Yang, B.-Y.: Significance of school greenspaces in preventing childhood myopia. The Innovation **5**(1) (2024)

[4] Morgan, I.G., Ohno-Matsui, K., Saw, S.-M.: Myopia. The Lancet **379**(9827), 1739–1748 (2012)

[5] Rose, K.A., French, A.N., Morgan, I.G.: Environmental factors and myopia: paradoxes and prospects for prevention. Asia-Pacific journal of ophthalmology **5**(6), 403–410 (2016)

[6] Dolgin, E.: The myopia boom. Nature **519**(7543), 276 (2015)

[7] Morgan, I.G., He, M., Rose, K.A.: Epidemic of pathologic myopia: what can laboratory studies and epidemiology tell us? Retina **37**(5), 989–997 (2017)

[8] Ng, D.S., Lai, T.Y.: Insights into the global epidemic of high myopia and its implications. JAMA ophthalmology **140**(2), 123–124 (2022)

[9] Ghorbani Mojarrad, N., Williams, C., Guggenheim, J.A.: A genetic risk score and number of myopic parents independently predict myopia. Ophthalmic and Physiological Optics **38**(5), 492–502 (2018)

[10] Chua, S.Y., Sabanayagam, C., Cheung, Y.-B., Chia, A., Valenzuela, R.K., Tan, D., Wong, T.-Y., Cheng, C.-Y., Saw, S.-M.: Age of onset of myopia predicts risk of high myopia in later childhood in myopic singapore children. Ophthalmic and Physiological Optics **36**(4), 388–394 (2016)

[11] Huang, T., Xu, H., Wang, H., Huang, H., Xu, Y., Li, B., Hong, S., Feng, G., Liu, G., Jiang, D., et al.: Artificial intelligence for medicine: Progress, challenges, and perspectives. The Innovation Medicine **1**(2) (2023)

[12] Xie, Y.: Values and limitations of statistical models. Research in social stratification and mobility **29**(3), 343–349 (2011)

[13] Yang, Y., Xu, S., Hong, Y., Cai, Y., Tang, W., Wang, J., Shen, B., Zong, H., Yu, G.: Computational modeling for medical data: From data collection to knowledge

discovery. The Innovation Life **2**(3), 100079–1 (2024)

[14] Grebovic, M., Filipovic, L., Katnic, I., Vukotic, M., Popovic, T.: Overcoming limitations of statistical methods with artificial neural networks. In: 2022 International Arab Conference on Information Technology (ACIT), pp. 1–6 (2022). IEEE

[15] Almeida, J.S.: Predictive non-linear modeling of complex data by artificial neural networks. Current opinion in biotechnology **13**(1), 72–76 (2002)

[16] Patra, J.C., Pal, R.N., Chatterji, B., Panda, G.: Identification of nonlinear dynamic systems using functional link artificial neural networks. IEEE transactions on systems, man, and cybernetics, part b (cybernetics) **29**(2), 254–262 (1999)

[17] Aziz, R., Verma, C., Srivastava, N.: Artificial neural network classification of high dimensional data with novel optimization approach of dimension reduction. Annals of Data Science **5**(4), 615–635 (2018)

[18] Aggarwal, C.C., *et al.*: Neural Networks and Deep Learning vol. 10. Springer, ??? (2018)

[19] Luo, M., Yang, W., Bai, L., Zhang, L., Huang, J.-W., Cao, Y., Xie, Y., Tong, L., Zhang, H., Yu, L., *et al.*: Artificial intelligence for life sciences: A comprehensive guide and future trends. The Innovation Life **2**(4), 100105–1 (2024)

[20] He, C., Ma, M., Wang, P.: Extract interpretability-accuracy balanced rules from artificial neural networks: A review. Neurocomputing **387**, 346–358 (2020)

[21] Zhang, Y., Tiňo, P., Leonardis, A., Tang, K.: A survey on neural network interpretability. IEEE transactions on emerging topics in computational intelligence **5**(5), 726–742 (2021)

[22] Fan, F.-L., Xiong, J., Li, M., Wang, G.: On interpretability of artificial neural networks: A survey. IEEE Transactions on Radiation and Plasma Medical Sciences **5**(6), 741–760 (2021)

[23] Zhang, Z., Beck, M.W., Winkler, D.A., Huang, B., Sibanda, W., Goyal, H., *et al.*: Opening the black box of neural networks: methods for interpreting neural network models in clinical applications. Annals of translational medicine **6**(11),

216 (2018)

[24] Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. Advances in neural information processing systems **30** (2017)

[25] Cremades, A., Hoyas, S., Vinuesa, R.: Additive-feature-attribution methods: A review on explainable artificial intelligence for fluid dynamics and heat transfer. International Journal of Heat and Fluid Flow **112**, 109662 (2025)

[26] Moraffah, R., Karami, M., Guo, R., Raglin, A., Liu, H.: Causal interpretability for machine learning-problems, methods and evaluation. ACM SIGKDD Explorations Newsletter **22**(1), 18–33 (2020)

[27] Ahmad, O., Béreux, N., Baret, L., Hashemi, V., Lecue, F.: Causal analysis for robust interpretability of neural networks. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 4685–4694 (2024)

[28] Xu, G., Duong, T.D., Li, Q., Liu, S., Wang, X.: Causality learning: A new perspective for interpretable machine learning. arXiv preprint arXiv:2006.16789 (2020)

[29] Pearl, J.: Causality. Cambridge university press, ??? (2009)

[30] Spirtes, P., Glymour, C.N., Scheines, R.: Causation, Prediction, and Search. MIT press, ??? (2000)

[31] Ramsey, J.D., Zhang, K., Glymour, M., Romero, R.S., Huang, B., Ebert-Uphoff, I., Samarasinghe, S., Barnes, E.A., Glymour, C.: Tetrad—a toolbox for causal discovery. In: 8th International Workshop on Climate Informatics, pp. 1–4 (2018). National Center for Atmospheric Research

[32] Baksh, J., Lee, D., Mori, K., Zhang, Y., Torii, H., Jeong, H., Hou, J., Negishi, K., Tsubota, K., Kurihara, T.: Myopia is an ischemic eye condition: A review from the perspective of choroidal blood flow. Journal of clinical medicine **13**(10), 2777 (2024)

[33] Li, S., Pan, Y., Xu, J., Li, X., Spiegel, D.P., Bao, J., Chen, H.: Effects of physical exercise on macular vessel density and choroidal thickness in children. Scientific Reports **11**(1), 2015 (2021)

[34] Yang, Y.S., Koh, J.W.: Choroidal blood flow change in eyes with high myopia. Korean journal of ophthalmology: KJO **29**(5), 309 (2015)

[35] SHIH, Y.-F., HORNG, I.-H., YANG, C.-H., LIN, L.L.-K., PENG, Y., HUNG, P.-T.: Ocular pulse amplitude in myopia. Journal of Ocular Pharmacology and Therapeutics **7**(1), 83–87 (1991)

[36] Zhang, D., Wu, M., Yi, X., Shi, J., Ouyang, Y., Dong, N., Gong, G., Guo, L., Zhou, L.: Correlation analysis of myopia and dietary factors among primary and secondary school students in shenyang, china. Scientific Reports **14**(1), 20619 (2024)

[37] Hao, Z., Jiang, M., Ma, J., Wu, T., Liu, Y., Zhou, J.: Association between sugar-sweetened beverages intake and myopia: Results from the national health and nutrition examination survey 2001–2008. Ophthalmic Epidemiology, 1–8 (2025)

[38] Kim, J.-M., Choi, Y.J.: Association between dietary nutrient intake and prevalence of myopia in korean adolescents: evidence from the 7th korea national health and nutrition examination survey. Frontiers in Pediatrics **11**, 1285465 (2024)

[39] Mrugacz, M., Zorena, K., Pony-Uram, M., Lendzioszek, M., Pieńczykowska, K., Bryl, A.: Interdependence of nutrition, physical activity, and myopia. Nutrients **16**(19), 3331 (2024)

[40] Dutheil, F., Oueslati, T., Delamarre, L., Castanon, J., Maurin, C., Chiambaretta, F., Baker, J.S., Ugbolue, U.C., Zak, M., Lakbar, I., *et al.*: Myopia and near work: a systematic review and meta-analysis. International journal of environmental research and public health **20**(1), 875 (2023)

[41] Huang, H.-M., Chang, D.S.-T., Wu, P.-C.: The association between near work activities and myopia in children—a systematic review and meta-analysis. PloS one **10**(10), 0140419 (2015)

[42] Lingham, G., Mackey, D.A., Lucas, R., Yazar, S.: How does spending time outdoors protect against myopia? a review. British Journal of Ophthalmology **104**(5), 593–599 (2020)

[43] Mei, Z., Zhang, Y., Jiang, W., Lam, C., Luo, S., Cai, C., Luo, S.: Efficacy of outdoor interventions for myopia in children and adolescents: a systematic review

and meta-analysis of randomized controlled trials. Frontiers in Public Health **12**, 1452567 (2024)

[44] Philipp, D., Vogel, M., Brandt, M., Rauscher, F.G., Hiemisch, A., Wahl, S., Kiess, W., Poulain, T.: The relationship between myopia and near work, time outdoors and socioeconomic status in children and adolescents. BMC Public Health **22**(1), 2058 (2022)

[45] He, M., Xiang, F., Zeng, Y., Mai, J., Chen, Q., Zhang, J., Smith, W., Rose, K., Morgan, I.G.: Effect of time spent outdoors at school on the development of myopia among children in china: a randomized clinical trial. Jama **314**(11), 1142–1148 (2015)

[46] Kearney, S., Strang, N.C., Cagnolati, B., Gray, L.S.: Change in body height, axial length and refractive status over a four-year period in caucasian children and young adults. Journal of optometry **13**(2), 128–136 (2020)

[47] Tao, L., Wang, C., Peng, Y., Xu, M., Wan, M., Lou, J., Yu, X.: Correlation between increase of axial length and height growth in chinese school-age children. Frontiers in Public Health **9**, 817882 (2022)

[48] Kuoliene, K., Danieliene, E., Tutkuviene, J.: Eye morphometry, body size, and flexibility parameters in myopic adolescents. Scientific Reports **14**(1), 6787 (2024)

[49] Zhang, J., Li, R., Zhang, Y., Tang, W., Ao, D., He, L., Yang, K., Qi, X., Zhou, X.: The association between physical stature and myopia in elementary and junior high school graduates in chongqing, china. Frontiers in Medicine **12**, 1530960 (2025)

[50] Wang, D., Zhao, C., Huang, S., Huang, W., He, M.: Longitudinal relationship between axial length and height in chinese children: Guangzhou twin eye study. Eye Sci **30**(1–6), 12 (2015)

[51] Wang, D., Ding, X., Liu, B., Zhang, J., He, M.: Longitudinal changes of axial length and height are associated and concomitant in children. Investigative ophthalmology & visual science **52**(11), 7949–7953 (2011)

[52] Lu, T.L., Wu, J.F., Ye, X., Hu, Y.Y., Wu, H., Sun, W., Guo, D.D., Wang, X.R., Bi, H.S., Jonas, J.B.: Axial length and associated factors in children: the

shandong children eye study. Ophthalmologica **235**(2), 78–86 (2016)

[53] Alrasheed, S.H., Alghamdi, W.: Systematic review and meta-analysis of the prevalence of myopia among school-age children in the eastern mediterranean region. Eastern Mediterranean Health Journal **30**(4) (2024)

[54] Zhao, L., Jiang, X., Zhang, W., Hao, L., Zhang, Y., Wu, S., Zhu, B., Xu, H.: Prevalence and risk factors of myopia among children and adolescents in hangzhou. Scientific Reports **14**(1), 24615 (2024)

[55] Rudnicka, A.R., Kapetanakis, V.V., Wathern, A.K., Logan, N.S., Gilmartin, B., Whincup, P.H., Cook, D.G., Owen, C.G.: Global variations and time trends in the prevalence of childhood myopia, a systematic review and quantitative meta-analysis: implications for aetiology and early prevention. British Journal of Ophthalmology **100**(7), 882–890 (2016)

[56] Groot, A.L., Lissenberg-Witte, B.I., Rijn, L.J., Hartong, D.T.: Meta-analysis of ocular axial length in newborns and infants up to 3 years of age. survey of ophthalmology **67**(2), 342–352 (2022)

[57] Terasaki, H., Yamashita, T., Asaoka, R., Yoshihara, N., Kakiuchi, N., Sakamoto, T.: Sex differences in rate of axial elongation and ocular biometrics in elementary school students. Clinical Ophthalmology, 4297–4302 (2021)

[58] Zadnik, K., Manny, R.E., Yu, J.A., Mitchell, G.L., Cotter, S.A., Quiralte, J.C., Shipp, M.D., Friedman, N.E., Kleinstein, R.N., Walker, T.W., *et al.*: Ocular component data in schoolchildren as a function of age and gender. Optometry and Vision Science **80**(3), 226–236 (2003)

[59] He, X., Sankaridurg, P., Naduvilath, T., Wang, J., Xiong, S., Weng, R., Du, L., Chen, J., Zou, H., Xu, X.: Normative data and percentile curves for axial length and axial length/corneal curvature in chinese children and adolescents aged 4–18 years. British Journal of Ophthalmology **107**(2), 167–175 (2023)

[60] Hashemi, H., Fotouhi, A., Mohammad, K.: The age-and gender-specific prevalences of refractive errors in tehran: the tehran eye study. Ophthalmic epidemiology **11**(3), 213–225 (2004)

[61] Lee, D.C., Lee, S.Y., Kim, Y.C.: An epidemiological study of the risk factors

associated with myopia in young adult men in korea. Scientific reports **8**(1), 511 (2018)

[62] Pandian, A., Sankaridurg, P.R., Naduvilath, T., O'Leary, D., Sweeney, D.F., Rose, K., Mitchell, P.: Accommodative facility in eyes with and without myopia. Investigative ophthalmology & visual science **47**(11), 4725–4731 (2006)

[63] Scheiman, M., Gwiazda, J., Zhang, Q., Deng, L., Fern, K., Manny, R.E., Weissberg, E., Hyman, L., Group, C., *et al.*: Longitudinal changes in corneal curvature and its relationship to axial length in the correction of myopia evaluation trial (comet) cohort. Journal of optometry **9**(1), 13–21 (2016)

[64] Jiang, L., Du, Z., Sun, W., Zhu, S., Xiong, L., Fang, X., Zhou, J., Zhang, Q., Lei, X., Zeng, Q., *et al.*: Associations between corneal curvature and other anterior segment biometrics in young myopic adults. Scientific Reports **14**(1), 8305 (2024)

[65] Carney, L.G., Mainstone, J.C., Henderson, B.A.: Corneal topography and myopia. a cross-sectional study. Investigative ophthalmology & visual science **38**(2), 311–320 (1997)

[66] Fan, Q., Pozarickij, A., Tan, N.Y., Guo, X., Verhoeven, V.J., Vitart, V., Guggenheim, J.A., Miyake, M., Tideman, J.W.L., Khawaja, A.P., *et al.*: Genome-wide association meta-analysis of corneal curvature identifies novel loci and shared genetic influences across axial length and refractive error. Communications biology **3**(1), 133 (2020)

[67] Wang, Y.-M., Lu, S.-Y., Zhang, X.-J., Chen, L.-J., Pang, C.-P., Yam, J.C.: Myopia genetics and heredity. Children **9**(3), 382 (2022)

[68] Jiang, X., Tarczy-Hornoch, K., Cotter, S.A., Matsumura, S., Mitchell, P., Rose, K.A., Katz, J., Saw, S.-M., Varma, R., Consortium, P., *et al.*: Association of parental myopia with higher risk of myopia among multiethnic children before school age. JAMA ophthalmology **138**(5), 501–509 (2020)

[69] Martínez-Albert, N., Bueno-Gimeno, I., Gené-Sampedro, A.: Risk factors for myopia: a review. Journal of clinical medicine **12**(18), 6062 (2023)

[70] Koomson, N.Y., Kobia-Acquah, E., Abdul-Kabir, M., Aderonke, U.M., Kwaw, R.J., Arkhurst, E.E.: Relationship between peripheral refraction, axial lengths

and parental myopia of young adult myopes. Journal of Optometry **15**(2), 122–128 (2022)

[71] Kurtz, D., Hyman, L., Gwiazda, J.E., Manny, R., Dong, L.M., Wang, Y., Scheiman, M., group, C., *et al.*: Role of parental myopia in the progression of myopia and its interaction with treatment in comet children. Investigative ophthalmology & visual science **48**(2), 562–570 (2007)

[72] Ishikuro, M., Fuse, N., Obara, T., Noda, A., Shinoda, G., Orui, M., Uruno, A., Yoshida, S., Takada, N., Fujioka, S., *et al.*: Parental axial lengths and prenatal conditions related to child axial length: The tmm birthree cohort study. AJO International **2**(1), 100088 (2025)

[73] Mutti, D.O., Mitchell, G.L., Moeschberger, M.L., Jones, L.A., Zadnik, K.: Parental myopia, near work, school achievement, and children's refractive error. Investigative ophthalmology & visual science **43**(12), 3633–3640 (2002)

[74] Hui, J., Peck, L., Howland, H.C.: Correlations between familial refractive error and children's non-cycloplegic refractions. Vision research **35**(9), 1353–1358 (1995)

[75] Xiang, F., He, M., Morgan, I.G.: The impact of parental myopia on myopia in chinese children: population-based evidence. Optometry and Vision Science **89**(10), 1487–1496 (2012)

[76] Ma, Y., Lin, S., Zhu, J., Zhao, R., Zhang, B., Yin, Y., Shao, Y., He, X., Xu, X., Zou, H.: Effect of parental myopia on change in refraction in shanghai preschoolers: a 1-year prospective study. Frontiers in Pediatrics **10**, 864233 (2022)

[77] Tang, S.M., Kam, K.W., French, A.N., Yu, M., Chen, L.J., Young, A.L., Rose, K.A., Tham, C.C., Pang, C.P., Yam, J.C.: Independent influence of parental myopia on childhood myopia in a dose-related manner in 2,055 trios: the hong kong children eye study. American Journal of Ophthalmology **218**, 199–207 (2020)

[78] Onyeahiri, C.: Approach to cycloplegic refraction. Canadian Journal of Optometry **86**(4), 11–30 (2024)

[79] Hinkley, S., Iverson-Hill, S., Haack, L.: The correlation between accommodative

lag and refractive error in minors under 18. Austin J Clin Ophthalmol **1**(2), 1–5 (2014)

[80] Lim, L.S., Gazzard, G., Low, Y.-L., Choo, R., Tan, D.T., Tong, L., Wong, T.Y., Saw, S.-M.: Dietary factors, myopia, and axial dimensions in children. Ophthalmology **117**(5), 993–997 (2010)

[81] Xue, C.C., Li, H., Dong, X.-X., Yu, M., Da Soh, Z., Chong, C.C.Y., Jiang, C., Choquet, H., Zebardast, N., Zekavat, S.M., *et al.*: Omega-3 polyunsaturated fatty acids as a protective factor for myopia. American Journal of Ophthalmology **268**, 368–377 (2024)

[82] Yin, C., Gan, Q., Xu, P., Yang, T., Xu, J., Cao, W., Wang, H., Pan, H., Ren, Z., Xiao, H., *et al.*: Dietary patterns and associations with myopia in chinese children. Nutrients **15**(8), 1946 (2023)

[83] Jin, G., Liu, Z., Wang, L., Zhu, Y., Luo, L., Liu, Y.: Corneal biometric features and their association with axial length in high myopia. American journal of ophthalmology **238**, 45–51 (2022)

[84] Merriam, J., Zheng, L.: The relationship of corneal curvature and axial length in adults. Investigative Ophthalmology & Visual Science **46**(13), 864–864 (2005)

[85] Aljuhani, G., Alharbi, M., Alsaidi, R., Alharbi, A.: Axial length and keratometry characteristics of patients undergoing cataract surgery in saudi arabia. Journal of Medicine and Life **17**(6), 620 (2024)

[86] Li, S.-M., Iribarren, R., Kang, M.-T., Li, H., Li, S.-Y., Liu, L.-R., Sun, Y.-Y., Meng, B., Zhan, S.-Y., Rozema, J.J., *et al.*: Corneal power, anterior segment length and lens power in 14-year-old chinese children: the anyang childhood eye study. Scientific reports **6**(1), 20243 (2016)

[87] Tideman, J.W.L., Polling, J.R., Jaddoe, V.W., Vingerling, J.R., Klaver, C.C.: Environmental risk factors can reduce axial length elongation and myopia incidence in 6-to 9-year-old children. Ophthalmology **126**(1), 127–136 (2019)

[88] He, X., Zou, H., Lu, L., Zhao, R., Zhao, H., Li, Q., Zhu, J.: Axial length/corneal radius ratio: association with refractive state and role on myopia detection combined with visual acuity in chinese schoolchildren. PLoS One **10**(2), 0111766

(2015)

[89] Grosvenor, T., Scott, R.: Role of the axial length/corneal radius ratio in determining the refractive state of the eye. Optometry and vision science **71**(9), 573–579 (1994)

[90] Angie, E., Amra, A.A., Sari, M.D.: Correlation between axial length and amplitude of accommodation in myopia patients. Journal of Public Health and Pharmacy **4**(3), 300–317 (2024)

[91] Hughes, R.P., Read, S.A., Collins, M.J., Vincent, S.J.: Changes in ocular biometry during short-term accommodation in children. Ophthalmic and Physiological Optics **40**(5), 584–594 (2020)

[92] Woodman, E.C., Read, S.A., Collins, M.J.: Axial length and choroidal thickness changes accompanying prolonged accommodation in myopes and emmetropes. Vision research **72**, 34–41 (2012)

[93] Read, S.A., Collins, M.J., Woodman, E.C., Cheong, S.-H.: Axial length changes during accommodation in myopes and emmetropes. Optometry and Vision Science **87**(9), 656–662 (2010)

[94] Flitcroft, D.I., He, M., Jonas, J.B., Jong, M., Naidoo, K., Ohno-Matsui, K., Rahi, J., Resnikoff, S., Vitale, S., Yannuzzi, L.: Imi–defining and classifying myopia: a proposed set of standards for clinical and epidemiologic studies. Investigative ophthalmology & visual science **60**(3), 20–30 (2019)

[95] Mutti, D.O., Hayes, J.R., Mitchell, G.L., Jones, L.A., Moeschberger, M.L., Cotter, S.A., Kleinstein, R.N., Manny, R.E., Twelker, J.D., Zadnik, K.: Refractive error, axial length, and relative peripheral refractive error before and after the onset of myopia. Investigative ophthalmology & visual science **48**(6), 2510–2519 (2007)

[96] Meng, W., Butterworth, J., Malecaze, F., Calvas, P.: Axial length of myopia: a review of current research. Ophthalmologica **225**(3), 127–134 (2011)

[97] Tharwat, E., Hassanein, M., Ezzeldin, E.R., Soliman, H.B., Eltantawy, B., Elgazzar, A.F., Abdella, W.S., Abdelkader, A.M.: Effect of cycloplegia on the refractive status of children. African Vision and Eye Health **83**(1), 916 (2024)

[98] Fotedar, R., Rochtchina, E., Morgan, I., Wang, J.J., Mitchell, P., Rose, K.A.:

Necessity of cycloplegia for assessing refractive error in 12-year-old children: a population-based study. American journal of ophthalmology **144**(2), 307–309 (2007)

[99] Khan, H.A., Tran, H., Naduvilath, T.J., Tahhan, N., Ha, T., Sankaridurg, P.: Comparison between cycloplegic and noncycloplegic refraction in young adult myopes. Optometry and Vision Science **101**(7), 470–476 (2024)

[100] Sharma, A., Kiciman, E.: Dowhy: An end-to-end library for causal inference. arXiv preprint arXiv:2011.04216 (2020)

[101] Li, S.-M., Ran, A.-R., Kang, M.-T., Yang, X., Ren, M.-Y., Wei, S.-F., Gan, J.-H., Li, L., He, X., Li, H., *et al.*: Effect of text messaging parents of school-aged children on outdoor time to control myopia: a randomized clinical trial. JAMA pediatrics **176**(11), 1077–1083 (2022)

[102] Doshi-Velez, F., Kim, B.: Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608 (2017)

[103] Sundararajan, M., Taly, A., Yan, Q.: Axiomatic attribution for deep networks. In: International Conference on Machine Learning, pp. 3319–3328 (2017). PMLR

[104] Adebayo, J., Gilmer, J., Muelly, M., Goodfellow, I., Hardt, M., Kim, B.: Sanity checks for saliency maps. advances in neural information processing systems. In: :, p. 31 (2018)

[105] Samek, W., Wiegand, T., Müller, K.-R.: Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296 (2017)

[106] Alvarez-Melis, D., Jaakkola, T.S.: On the robustness of interpretability methods. arXiv preprint arXiv:1806.08049 (2018)

[107] Salih, A.M., Raisi-Estabragh, Z., Galazzo, I.B., Radeva, P., Petersen, S.E., Lekadir, K., Menegaz, G.: A perspective on explainable artificial intelligence methods: Shap and lime. Advanced Intelligent Systems **7**(1), 2400304 (2025)

[108] Kelodjou, G., Rozé, L., Masson, V., Galárraga, L., Gaudel, R., Tchuente, M., Termier, A.: Shaping up shap: Enhancing stability through layer-wise neighbor selection. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol.

38, pp. 13094–13103 (2024)

[109] Li, S.-M., Liu, L.-R., Li, S.-Y., Ji, Y.-Z., Fu, J., Wang, Y., Li, H., Zhu, B.-D., Yang, Z., Li, L., *et al.*: Design, methodology and baseline data of a school-based cohort study in central china: the anyang childhood eye study. Ophthalmic epidemiology **20**(6), 348–359 (2013)

[110] Pearl, J., Mackenzie, D.: The Book of Why: the New Science of Cause and Effect. Basic books, ??? (2018)

[111] Meek, C.: Causal inference and causal explanation with background knowledge. arXiv preprint arXiv:1302.4972 (2013)

[112] Andrews, B., Ramsey, J., Cooper, G.F.: Learning high-dimensional directed acyclic graphs with mixed data-types. In: The 2019 ACM SIGKDD Workshop on Causal Discovery, pp. 4–21 (2019). PMLR

[113] Venter, J.H.: On estimation of the mode. The Annals of Mathematical Statistics, 1446–1455 (1967)

[114] Keith Battocchi, e.a.: EconML: A Python Package for ML-Based Heterogeneous Treatment Effects Estimation. https://github.com/py-why/EconML. Version 0.x (2019)

11 Figures

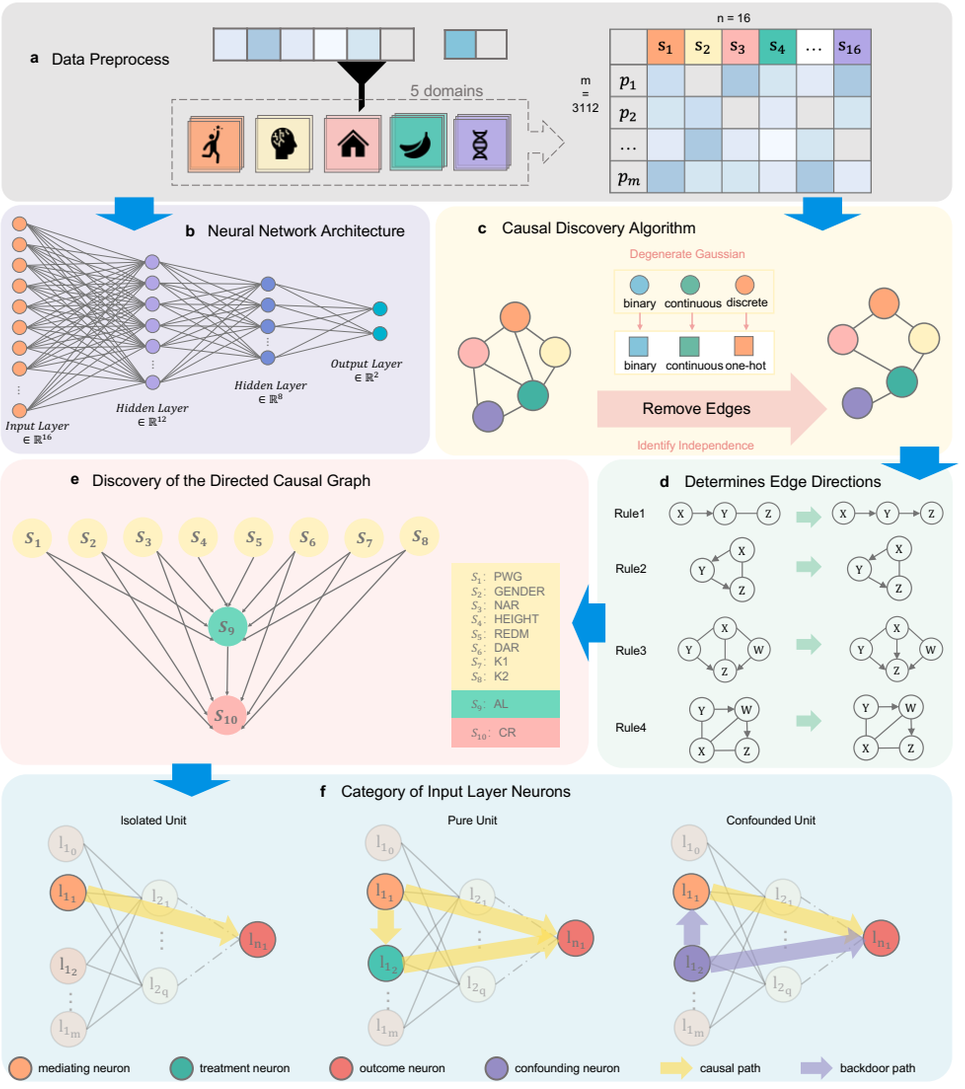


Fig. 1: **The workflow of causal structure discovery and functional categorization of input neurons a** Data Preprocess. Sixteen input variables were selected across five domains: behavioral, physiological, environmental, dietary, and hereditary. **b** Architecture of the neural network (NN). **c** Flow of removing edges via causal discovery algorithm. **d** Four rules for determining edge directions. **e** The directed causal graph of key behavioral and genetic factors related to myopia, involving 10 of the 16 input variables, connected by a total of 15 edges. **f** Three categorized units of the corresponding input-layer neurons in NN.

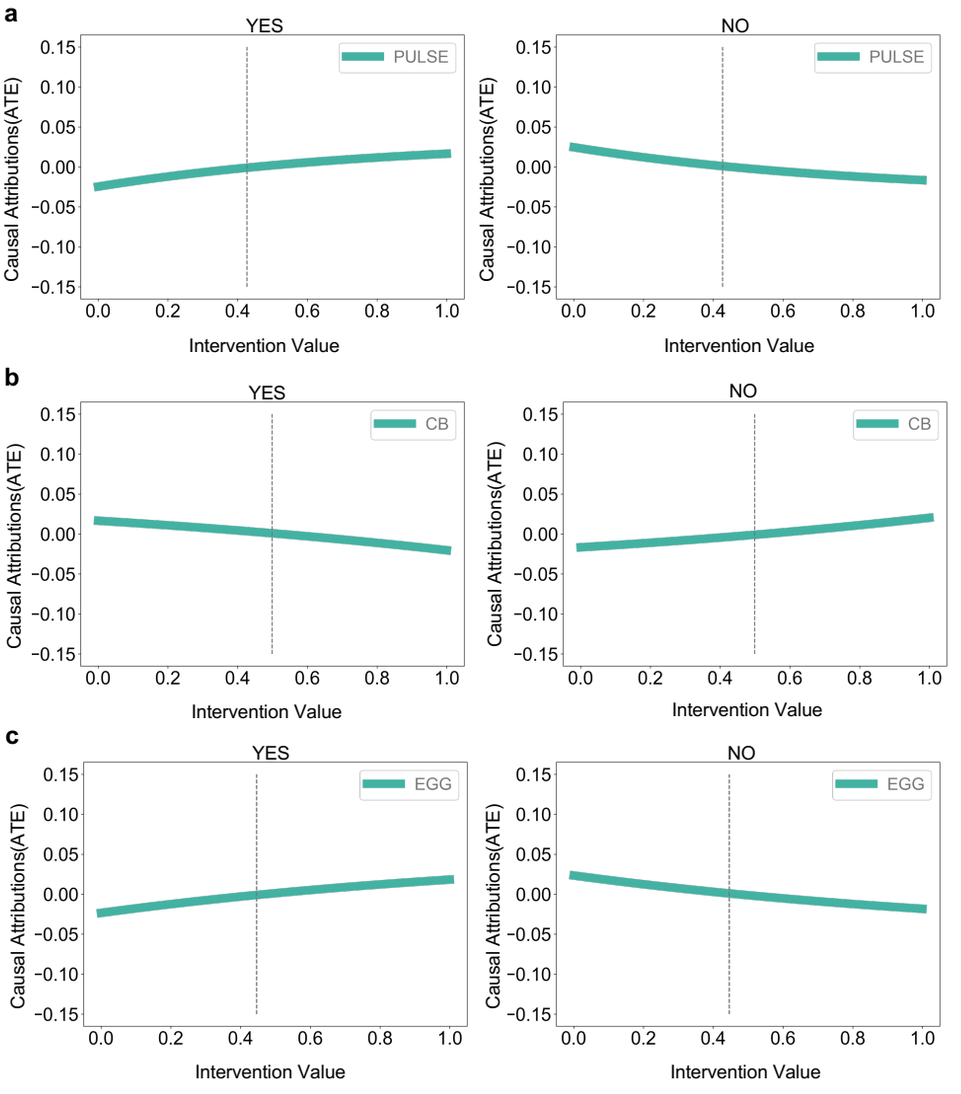
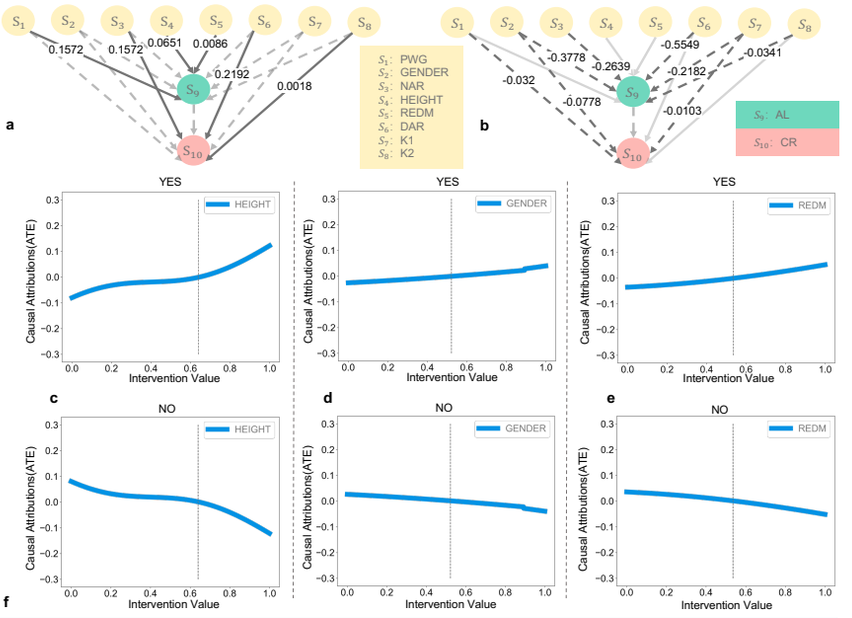


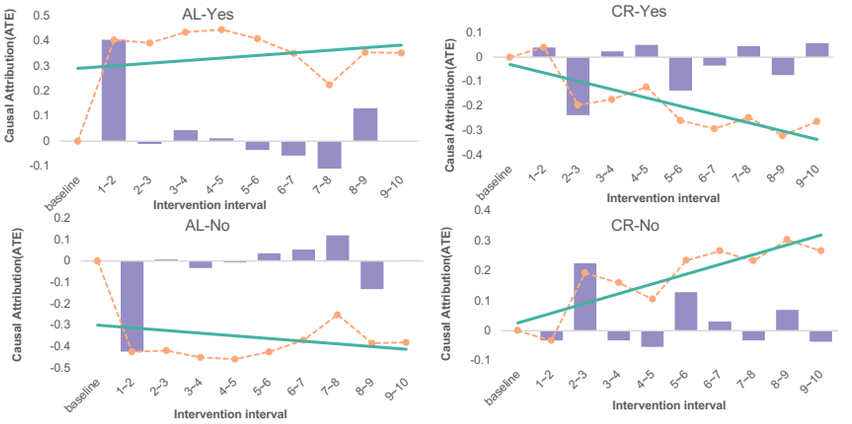
Fig. 2: **Causal attribution results of Isolated Units a** Intervention value of PULSE. **b** Intervention value of CB. **c** Intervention value of EGG. The difference between the ATE value and its average value is presented as the green line, which shows the increasing/decreasing trend of variable for myopic. 'YES' means myopic and 'NO' means non-myopic. The vertical dotted line indicates that the ATE at this location is 0.



treatment	outcome	ATE	BV/err%	ARCC/err%	DSV/err%	PB/err%	DO/err%
PWG	AL	0.1041	0.1035/0.6291%	0.1041/0.0297%	0.1028/1.2914%	0/0%	0.0005/0.0543%
PWG	CR	-0.032	-0.03/6.2595%	-0.0319/0.0275%	-0.0319/0.2702%	0/0%	0.0004/0.0355%
GENDER	AL	-0.3778	-0.3741/0.9775%	-0.3777/0.013%	-0.3716/1.6287%	-0.004/0.4023%	0.01/0.9982%
GENDER	CR	-0.0778	-0.0786/1.1297%	-0.0778/0.0192%	-0.0786/1.1078%	0.0001/0.0106%	0.0042/0.4195%
NAR	AL	-0.2639	-0.2586/2.0231%	-0.2639/0.0037%	-0.2461/6.7741%	0/0%	0.0013/0.1287%
NAR	CR	0.1572	0.1464/6.8549%	0.1572/0.003%	0.1617/2.8822%	0/0%	0.003/0.2974%
HEIGHT	AL	0.0651	0.0645/0.934%	0.0651/0.0096%	0.0645/0.9139%	0/0%	0/0.0019%
REDM	AL	0.0086	0.0085/1.1528%	0.0085/0.0542%	0.0084/1.2745%	0/0%	-0.0003/0.029%
DAR	AL	-0.5549	-0.5698/2.6925%	-0.5549/0.0013%	-0.5607/1.0469%	0/0%	0.0007/0.069%
DAR	CR	0.2192	0.2249/2.5916%	0.2192/0.0003%	0.2218/1.1852%	0/0%	0.0003/0.0331%
K1	AL	-0.2182	-0.2194/0.5381%	-0.2182/0.0026%	-0.2181/0.0575%	0/0%	-0.0004/0.044%
K1	CR	-0.0103	-0.0104/1.3375%	-0.0103/0.0214%	-0.01/2.2884%	0/0%	-0.0029/0.2902%
K2	AL	-0.0341	-0.0368/7.6269%	-0.0341/0.0109%	-0.0371/8.6001%	0/0%	0/0.003%
K2	CR	0.0018	0.0019/3.2185%	0.0018/0.1668%	0.0018/0.9818%	0/0%	0.0016/0.1606%
AL	CR	-0.0813	-0.0784/3.4753%	-0.0813/0.0234%	-0.0785/3.4161%	0/0%	-0.0015/0.1479%

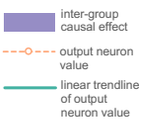
Fig. 3: **Causal attribution results of Pure Units a** Result of positive causal effects. **b** Result of negative causal effects. Solid lines represent positive causal effects and dashed lines indicate negative effects. Numbers on each line represent causal effect value of corresponding causal links. **c-e** Intervention value of HEIGHT, GENDER and REDM. The difference between the ATE and its average value is presented as the blue line, showing the increasing/decreasing trend of variable for myopic. 'YES' means myopic and 'NO' means non-myopic. The vertical dotted line indicates that the ATE at this location is 0. **f** Refutation results of the 15 causal effects.

a

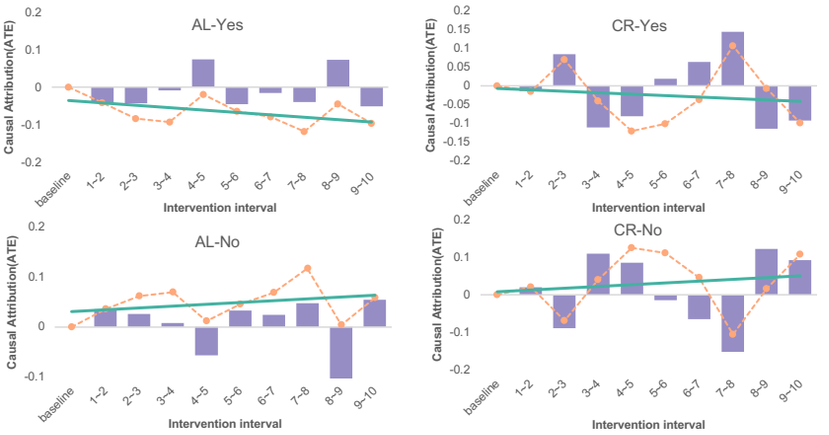


b

ATE	BV/err%	ARCC/err%	DSV/err%	PB/err%	DO/err%
-0.3575	-0.2808/21.46%	-0.3415/4.48%	-0.243/32.03%	-0.0022/0.22%	-0.0941/9.41%
0.365	0.2815/22.86%	0.3414/6.45%	0.2409/34%	0.0066/0.66%	-0.0153/1.53%
0.2811	0.2842/1.11%	0.2685/4.48%	0.2599/7.54%	0.001/0.1%	0.0971/9.71%
-0.2797	-0.2769/1.01%	-0.2672/4.47%	-0.2703/3.35%	0.0065/0.65%	0.0878/8.78%



c



d

ATE	BV/err%	ARCC/err%	DSV/err%	PB/err%	DO/err%
0.0802	0.0740/7.75%	0.0776/3.2%	0.0732/8.67%	0.0005/0.05%	-0.0352/3.52%
-0.0758	-0.0718/5.36%	-0.0760/0.15%	-0.0729/3.87%	0.0005/0.05%	0.0308/3.08%
0.0949	0.0945/0.43%	0.1075/13.26%	0.1138/19.86%	-0.0001/0.01%	-0.0063/0.63%
-0.0955	-0.1192/24.87%	-0.1038/8.7%	-0.1024/7.18%	0.0031/0.31%	0.0363/3.63%

Treatment →Outcome
AL→No
AL→Yes
CR→No
CR→Yes

Fig. 4: **Causal attribution results of Confounded Units** (**a** and **c**) Causal effects of AL and CR on output neurons (YES/NO) under equal-width and equal-frequency discretization, respectively. Equal-width discretization divides AL and CR into 10 groups within their value intervals and calculates the ACE between adjacent discrete groups. The equal-frequency discretization divides variables into 10 groups and the number of data points in each group is the same. The output neuron value (YES/NO) was benchmarked at 0. The corresponding causal effect value was added to each intervention interval, from which the dashed line of the output neuron value change showed. (**b** and **d**) Refutation results of equal-width and equal-frequency discretization, respectively.

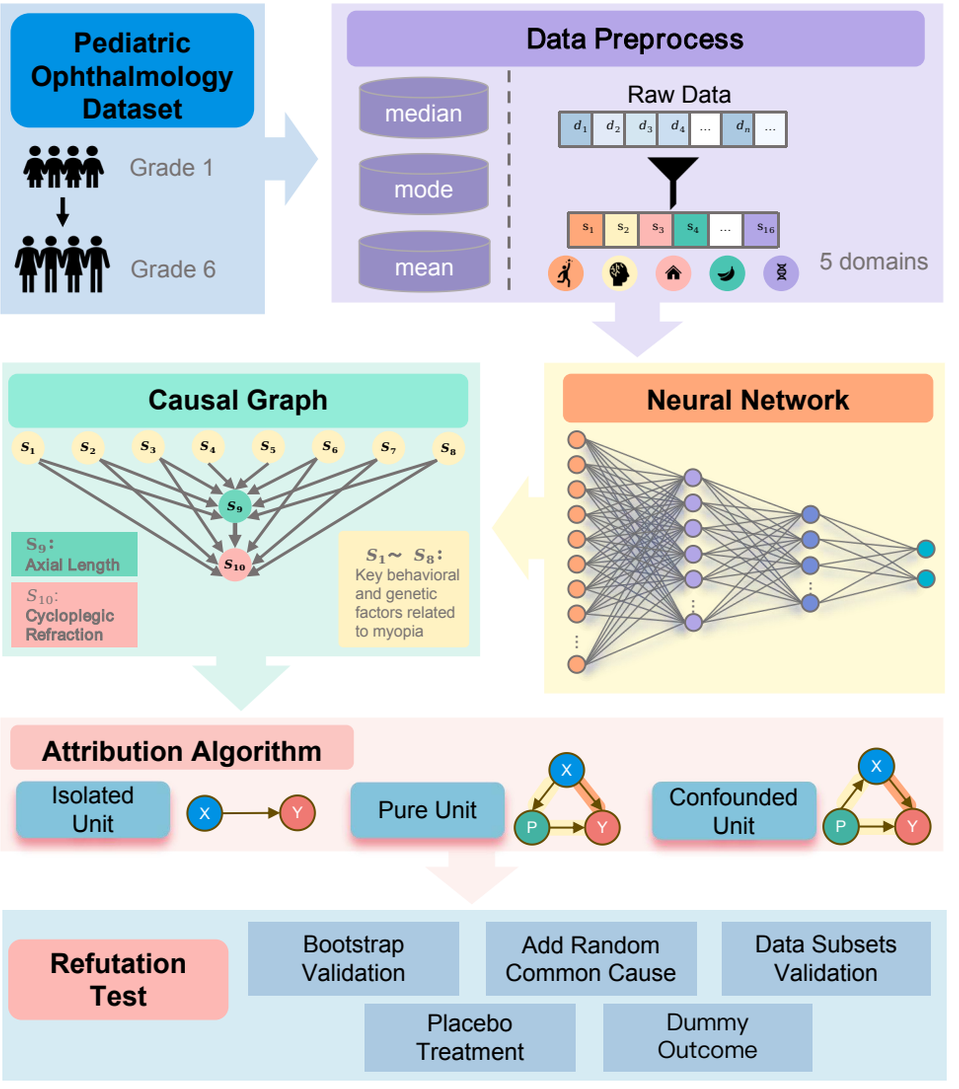


Fig. 5: Graphical abstract