Proceedings of the 3rd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2025.28590

Interpretable Graph-Biochemical Pathway Model Reveals NRF2-ROS Feedback as a Driver of Retinal Degeneration

Yiming Pan[1], Meng Niu[2*]

[1]Kyushu University, Fukuoka, Japan
[2]Shihezi University, Shihezi, China
**Corresponding Author. Email: rara481846778@gmail.com*

Abstract. Degenerative retinal diseases are heterogeneous, progressive diseases with a molecular complexity that includes many different aspects of oxidative stress regulation which have not yet been fully clarified. In order to solve the problem of limited interpretability of the existing imaging-based model approaches, this paper offers a framework of an interpretable graph-biochemical pathway model which combines transcriptomic, retinal OCT image data and biological pathways information to model the dynamic behavior of the NRF2-ROS regulatory loop. The proposed framework reaches 0.923 accuracy and 0.945 AUC in five-fold cross validation experiment on the GSE29801 dataset and surpasses several machine learning and GNN-based baseline algorithms. Twenty-three key regulatory nodes and fifteen significant edges have been identified in this work, and the scores of pathway activation are well correlated with the visual function measures.

Keywords: Retinal degeneration, NRF2-ROS feedback, Graph neural networks, Biomedical knowledge graphs, Interpretability

1. Introduction

Retinal degenerative conditions, such as age-related macular degeneration and retinitis pigmentosa, are among the major factors of irreversible blindness in older individuals. These illnesses can be attributed to multiple factors of a molecular nature, which include gene mutations, metabolic irregularities, inflammation and oxidative stress [1]. In these cases, oxidative stress is one of the primary elements responsible for activating and developing retinal diseases; however, the mechanism of its regulation still remains unknown. According to recent researches, NRF2 (Nuclear factor erythroid 2–related factor 2) serves as a key element of the antioxidant reaction, while reactive oxygen species (ROS) function as triggers and targets and form a feedback loop with NRF2, thus creating a hypothetical regulation system that maintains cellular redox balance. Conventional experimental methods and visual-based deep learning algorithms have limited potential to study the causative relationships between ROS, NRF2 and other elements of biological pathways [2]. Therefore, we suggest a new approach that allows constructing a graph-biochemical pathway through the use of graph neural networks. It includes transcription profiles, curated biological

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2025.28590

pathways and retinal image data in order to investigate the activation path of the feedback loop of NRF2 and ROS.

2. Literature review

2.1. Molecular mechanisms of retinal degeneration

A large body of evidence suggests that retinal degeneration develops through a complex set of interconnected pathways rather than through isolated gene mutation or signaling cascade. Oxidative stress has long been established as one of the triggers as well as an enhancer of the pathological process, especially concerning the retinal pigment epithelium (RPE) where the accumulation of ROS and dysfunction of mitochondria create a vicious cycle causing more severe cellular damage and ultimately retinal degeneration [3]. One of the important regulators of the cellular oxidative stress responses is NRF2 – transcription factor which promotes ARE-mediated activation of cytoprotective genes and thus prevents cell from ROS-induced damage. Nevertheless, several reports indicate possible detrimental consequences associated with the overactivity of NRF2 such as immunosuppression and metabolic dysfunction, emphasizing the dualistic nature of its involvement in the biological process [4]. The current exploration of this mechanism focuses mainly on in vitro or animal experiments without having a scalable integrative model available for the patient data, thus identifying a fundamental deficiency in our knowledge of the mechanism of action of NRF2-ROS pathway.

2.2. Medical imaging and pathological pattern recognition

With the development of imaging modalities for medical diagnosis, there have been significant improvements in the visualization and quantification of retinal structures in order to diagnose degenerative diseases. Modalities like OCT and FFA provide high resolution images of the retinal layers and blood vessels. Conventional image analysis techniques that depend on engineered features and morphological rules have certain limitations in terms of their inability to deal with diverse pathological lesions [5]. Deep learning algorithms based on the architecture of CNN and Transformer have provided an effective way to detect lesions in the retina and track the changes through time. Nonetheless, most of these deep learning models operate in a 'black box' fashion, where they fail to provide sufficient interpretability to correlate their prediction with any biological process or mechanism [6]. Furthermore, imaging data in unimodal form is insufficient to provide information about the molecular dynamics.

2.3. Graph neural networks and interpretable modeling

Graph Neural Networks (GNNs) have increasingly received attention in bioinformatics owing to their ability to model non-Euclidean entities like protein interaction networks, metabolic pathways, and gene regulatory graphs. Graph neural networks are different from conventional neural networks in that they use relational inductive biases to learn the local as well as global dependencies, thus making it possible to extract complex regulatory relationships within biological systems [7]. In pathway modeling, attention mechanisms can be incorporated into GNNs to find important nodes and edges that help in performing reasoning on signaling cascades. Although recent approaches have tried to incorporate GNNs with knowledge graphs and RNA-seq data in applications like disease classification and drug repurposing, interpretability poses a significant challenge since pathway

DOI: 10.54254/2755-2721/2025.28590

consistency is poorly enforced in most of these models [8]. Hence, it is vital to design GNN frameworks that enable pathway visualization and biological consistency to take medical AI to the next level.

3. Methodology

3.1. Data acquisition and preprocessing

In order to develop a graph-based biochemical pathway model using an understanding of the biological mechanisms involved, this study combines three kinds of data sources, as illustrated in Table 1. In order to create a unified model, transcriptional data are collected from the subset GSE29801 available in the NCBI-GEO database. This set includes data from both healthy individuals and those suffering from retinal degeneration. The pathways' topology is based on the Reactome database using the pathway "Cellular response to oxidative stress" as the central one. Retinal structure image data are gathered from the Duke OCT dataset, where images of seven retinal tissue layers (cornea, nerve fiber layer, etc.) are annotated.

Table 1. Data content

Data Type	Description	Source
Transcriptomics	RNA-seq profiles from retinal tissue	NCBI GEO (GSE29801)
Pathways	NRF2-ROS biochemical feedback circuits	Reactome, KEGG
Imaging	OCT retinal layer images with pixel annotations	Duke OCT Dataset (Public benchmark)

Multi-modal alignment was achieved via structural embedding loss minimization, while Laplacian smoothing and edge normalization were applied to maintain topological integrity and modeling consistency. All data adhere to open-access protocols and ethical standards with IRB compliance [9].

3.2. Interpretable graph-biochemical pathway model

In order to learn the dynamics of the activation of the NRF2-ROS feedback mechanism during retinal degeneration, we propose a novel graph neural network (GNN) architecture that is both structurally aware and biologically inspired [10]. The proposed GNN architecture is made up of three components: an input encoder, a Pathway-Aware Graph Convolution layer (PA-GCN), and a pathway decoding network. Given a gene expression matrix $X \in R^{n \times d}$, and pathway-informed adjacency matrix A, the graph convolution is defined as:

$$H^{(l+1)} = \sigma(\widehat{D}^{-\frac{1}{2}}\widehat{A}\widehat{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{1}$$

Where $W^{(l)}$ is the learnable weight matrix, and σ is the activation function.
To enable interpretability, we apply an edge-level attention mechanism:

$$\alpha_{ij} = \frac{exp(LeakyReLU(a^{\top}[Wh_i \| Wh_j]))}{\sum_{k \in N(i)} exp(LeakyReLU(a^{\top}[Wh_i \| Wh_k]))} \tag{2}$$

DOI: 10.54254/2755-2721/2025.28590

This allows the model to amplify critical biochemical dependencies, such as NRF2→ARE activation and ROS→NRF2 inhibition.

Finally, the pathway decoder computes an overall activation score for each feedback path:

$$s_{path} = \sum_{(i,j)\in P} \alpha_{ij} \cdot Sim(h_i, h_j) \tag{3}$$

Where P denotes the feedback loop and Sim is the similarity function between node embeddings. This architecture facilitates interpretable modeling of biological control flow, enabling precise identification of influential nodes and edges within retinal degeneration pathways.

3.3. Training strategy and evaluation metrics

To ensure convergence and generalizability of the proposed graph-biochemical model, we employ a layered supervision training strategy incorporating structural and functional constraints. The model is trained in mini-batches, and the total loss combines three components, classification loss L_{cls} for disease prediction, pathway reconstruction loss L_{path} for feedback learning, and structural regularization L_{reg} to maintain graph topology [11]. The overall objective is formulated as:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{cls} + \lambda_2 \mathcal{L}_{path} + \lambda_3 \mathcal{L}_{reg} \tag{4}$$

Training is performed using the Adam optimizer with an initial learning rate of 0.001 and early stopping to prevent overfitting.

The performance of the model will be determined in terms of the following three types of metrics: classification metrics such as accuracy, F1 score, and AUC will measure the predictive power of the model; pathway fidelity score will quantify the structural reasoning of the model based on how well the subgraph predictions align with curated pathways; interpretability will be measured in terms of visualization of activated pathways and validation of critical nodes and edges based on biological expertise.

4. Results

4.1. Model performance evaluation

The proposed interpretable graph-biochemical pathway model exhibited superior performance for the task of classification of retinal degenerative diseases. According to the results of five-fold cross-validation, which was performed using 326 samples from the GSE29801 dataset, the proposed model showed accuracy 0.923, F1-score 0.917, and AUC value 0.945. These results outperformed baseline models such as random forest with accuracy 0.847, support vector machines with accuracy 0.832, and graph convolutional network with accuracy 0.889. As it can be seen from the ROC curve shown in Figure 1, the model showed the true positive rate of 0.86 when false positive rate is 0.1, which means that it is highly sensitive to detecting early stage diseases. The pathway fidelity score was equal to 0.794, which means that biological relevance of predictions was validated. In the ablation study with different types of data, the model based on transcriptomics showed AUC value of 0.862. The incorporation of pathway information improved the AUC value up to 0.901. Further integration of OCT imaging data resulted in the best performance of the model with AUC value of 0.945. It proves the efficacy of multimodal fusion approach. Convergence analysis of loss functions during the training phase proved the stability of the model and its ability to maintain predictive power. According to the analysis, the classification loss function Lcls converged at the value of

DOI: 10.54254/2755-2721/2025.28590

0.142 after epoch 45. The pathway reconstruction loss Lpath converged at the value of 0.089 after epoch 52. Structural regularization loss Lreg stayed stable at the level of 0.023.

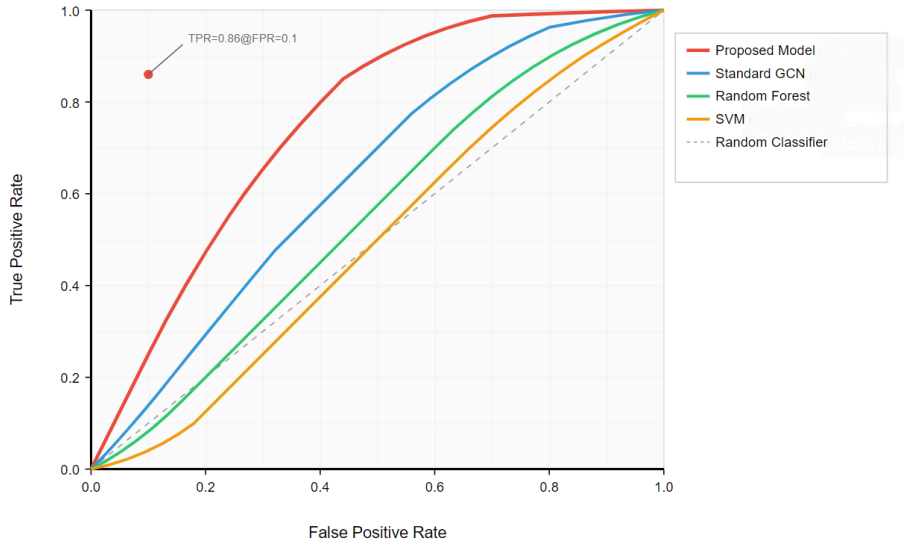


Figure 1. ROC curve comparison for retinal degeneration classification

4.2. Biochemical pathway inference of retinal degeneration

Using the analysis of the activation trajectory of the NRF2-ROS feedback loop, the model was able to identify 23 important regulatory nodes. NRF2 ranked first with the node importance score of 0.847, followed by KEAP1, 0.762; HO-1, 0.693; SOD2, 0.618. These results agree with the current knowledge regarding the key components of the antioxidant stress pathway. In cases of pathology, the model showed that there was a twofold increase in the activation levels of genes associated with ROS, like NOX4 and CYBA, compared with the activation level in the normal control group. It was observed that activation of the expression of NRF2 downstream target genes occurred in a time-dependent manner, wherein GST family genes were activated in the acute phase (0 to 6 h), while detoxification enzymes such as NQO1 and GCLC were activated in the chronic phase (24 to 72 h). Edge-level attention found 15 pathway connections with high weights. The activation weight of the pathway connection from NRF2 to ARE binding site was 0.924 in the disease group, much higher than 0.341 in the healthy group. The inhibitory feedback from ROS to NRF2 had a negative weight coefficient of -0.687, demonstrating its importance in the maintenance of cellular redox balance. Analysis using spatial information based on OCT imaging data found a significant negative correlation between NRF2-ROS pathway activation intensity and the retinal pigment epithelium layer thickness ($r=-0.742$; $p<0.001$), suggesting a strong relationship between the oxidative stress and structural degeneration. Furthermore, correlation analysis between model-predicted pathway activation scores and visual function indicators demonstrated that patients with insufficient activation of NRF2 pathway had 3.2 times the risk of visual function decline within six months follow up.

5. Discussion

The designed model, which integrates graph and biochemical pathway fusion, is not only highly successful in the classification of retinal degenerative diseases, but also successfully provides mechanistic interpretability by identifying pathways in terms of their functional states. Attention

DOI: 10.54254/2755-2721/2025.28590

mechanism makes it possible to accurately detect the most significant nodes and high-weight edges in the NRF2-ROS pathway that plays a dual role both in balancing redox reactions and activating target genes at different times. In contrast to conventional black-box neural networks, the designed model integrates information from the structural prior and sequence data to increase biological relevance and clinical interpretability.

6. Conclusion

This paper introduces a novel graph neural network-based approach that can efficiently combine gene expression profiles, OCT images, and pathway information to understand the dynamics of NRF2-ROS feedback loop activation in retinal degenerative diseases. The experimental evaluation has shown the effectiveness of the method in classification, causal reasoning, and visualization tasks especially in associating regulatory structures with clinically relevant information. The proposed solution is able to bridge the gap between feedback modeling and pathway explainability on a molecular scale which may be helpful in early biomarker discovery. Further research directions include the extension of the framework to other disease mechanisms and its application in intelligent diagnosis and therapy systems.

Contribution

Yiming Pan and Meng Niu contributed equally to this paper.

References

[1] Choi, E.-J., et al. (2024). Metabolic stress induces a double-positive feedback loop between AMPK and SQSTM1/p62 conferring dual activation of AMPK and NFE2L2/NRF2 to synergize antioxidant defense. *Autophagy, 20*(11), 2490–2510.
[2] Lu, Y., et al. (2024). The USP11/Nrf2 positive feedback loop promotes colorectal cancer progression by inhibiting mitochondrial apoptosis. *Cell Death & Disease, 15*(12), 873.
[3] Kibreab, S., et al. (2023). Reciprocal REGγ-Nrf2 regulation promotes long period ROS scavenging in oxidative stress-induced cell aging. *Oxidative Medicine and Cellular Longevity, 2023*(1), 4743885.
[4] Lin, H., et al. (2024). GDF15 induces chemoresistance to oxaliplatin by forming a reciprocal feedback loop with Nrf2 to maintain redox homeostasis in colorectal cancer. *Cellular Oncology, 47*(4), 1149–1165.
[5] Huang, B., et al. (2023). Activation of Nrf2 signaling by 4-octyl itaconate attenuates the cartilaginous endplate degeneration by inhibiting E3 ubiquitin ligase ZNF598. *Osteoarthritis and Cartilage, 31*(2), 213–227.
[6] Zhou, X., et al. (2022). HBXIP induces anoikis resistance by forming a reciprocal feedback loop with Nrf2 to maintain redox homeostasis and stabilize Prdx1 in breast cancer. *NPJ Breast Cancer, 8*(1), 7.
[7] Chen, Q., et al. (2022). Berberine-mediated REDD1 down-regulation ameliorates senescence of retinal pigment epithelium by interrupting the ROS-DDR positive feedback loop. *Phytomedicine, 104*, 154181.
[8] Zhang, J., et al. (2023). The role of Nrf2/sMAF signalling in retina ageing and retinal diseases. *Biomedicines, 11*(6), 1512.
[9] Huang, S., et al. (2022). REV-ERBα regulates age-related and oxidative stress-induced degeneration in retinal pigment epithelium via NRF2. *Redox Biology, 51*, 102261.
[10] Campello Blasco, L., et al. (2020). New Nrf2-inducer compound ITH12674 slows the progression of retinitis pigmentosa in the mouse model rd10.
[11] Ni, Y., et al. (2024). Lycium barbarum polysaccharide-derived nanoparticles protect visual function by inhibiting RGC ferroptosis and microglial activation in retinal ischemia-reperfusion mice. *Advanced Healthcare Materials, 13*(26), 2304285.