Graph Neural Networks for Prediction of Gene-Autoimmune Disorder Associations

Meghan Kwon[1], Bermet Abylova[2]
[1]Department of Mathematics and Statistics, UMBC, 1000 Hilltop Cir, Baltimore, MD 21250
[2]Department of Biomedical Engineering, CityU, 83 Tat Chee Ave, Kowloon Tong, Hong Kong

Mentored by Dr. Kwai Wong, University of Tennessee, Knoxville, TN 37996

Abstract

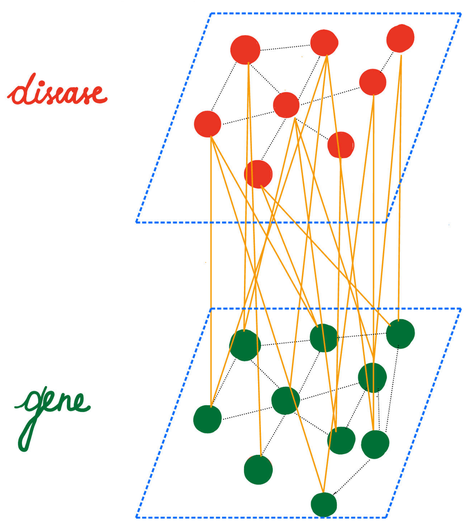
Autoimmune disorders (ADs) are a prevalent and growing concern. Despite the increasing number of cases each year, the majority of genes underlying these diseases are unknown. While efforts are being made to treat and develop therapies for these particular disorders, the methods for target discovery and identification are time and cost expensive. We propose a novel model to predict gene-autoimmune disease associations with Graph Neural Networks (GNNs).

Objectives

- Develop an accurate and computationally efficient model for predicting gene-autoimmune disease associations using heterogeneous graph neural networks.
- Train the model on data with known associations and deploy it to uncover potential new associations.
- The system aims to be adaptable for various deep learning applications, potentially advancing our understanding of the genetic basis of autoimmune diseases and identifying novel therapeutic targets for improved management and treatment.

Methods

In order to predict gene-autoimmune disease associations, we constructed a heterogeneous graph using a graph convolutional framework, with genes and autoimmune diseases as nodes and three types of edges representing their interactions. The data is sourced from the publicly available Gene & Autoimmune Disease Association Database (GAAD).



We load in 66 unique disease nodes, 4051 unique gene nodes, 183,346 gene-gene associations, 907 disease-disease associations, and 3295 known gene-autoimmune disorder associations from our datasets. Each disease node's features are based on their known pathophysiological pathways, and each gene's features are based on the number of known associations to ADs.

Methods cont.

We construct our graph using the Heterogeneous Data class available through the PyTorch library. Each node and edge is assigned a unique index for identification, and the connections are stored through tensors.

Following a convolutional framework, we update a node's features using an aggregation of its neighbor's features. For each hidden layer, the ReLu function is applied as the activation function.

$$x_i' = W_1 x_i + W_2 \cdot mean_{j \in N(i) x_j} \qquad RELU(x) = max(0, x)$$

We apply a final classifier function as the dot product between the vector of disease and gene node embeddings to obtain a measure (the prediction) of their relative "closeness".

Within the training loop, we calculate the loss for our model.

$$L = -\frac{1}{N}\left[\sum_{j=1}^{N}\left[t_j \log(p_j) + (1 - t_j)\log(1 - p_j)\right]\right]$$

Let N be the number of data points, t_j be the truth value of the edge, and p_j be the prediction given by the model.

Results

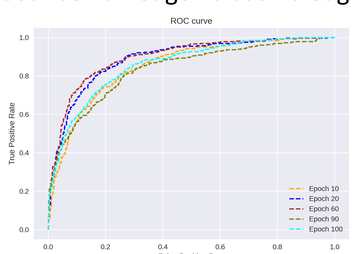
Measurements of performance:

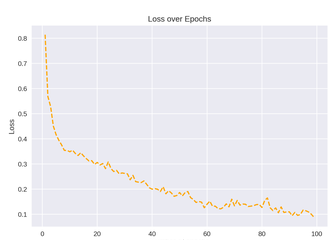
$$\text{Recall} = \frac{TP}{TP + FN} \qquad \text{Precision} = \frac{TP}{TP + FP}$$

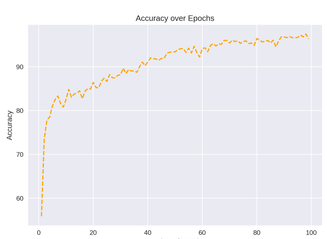
$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

Area Under the Receiving Operator Curve (ROC).

- The ROC curve is a metric to evaluate how well our model can distinguish between an edge versus no edge.

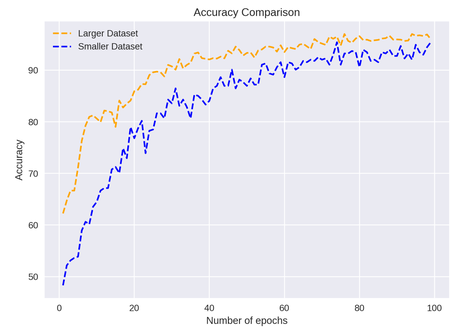


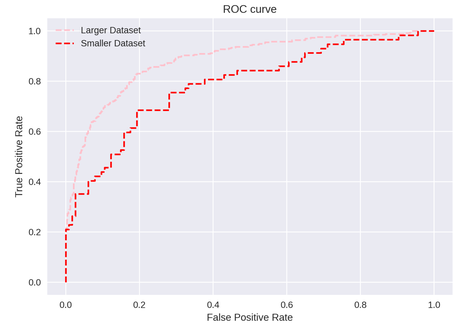




Results cont.

Initially, we input 49 autoimmune disorders, 3618 gene nodes, 174,513 gene-gene associations, 907 disease-disease associations, and 575 disease-gene associations. Looking to improve the accuracy of our model, we performed data augmentation to obtain our final model.





Conclusion & Future Directions

We built a heterogeneous graph neural network with a convolutional framework which accurately predicts novel gene-autoimmune disease associations. We discovered that increasing the amount of data points improves the model significantly. We hope to extend our model to include even more genes which we can relate to our 67 autoimmune disorders. We would like to benchmark our model in order to compare it to existing heterogeneous GNNs.

Acknowledgements

This work was funded by the National Science Foundation through Research Experience for Undergraduates (REU) award, with additional support from the National Institute of Computational Sciences at the University of Tennessee, Knoxville. In addition, we would like to thank Dr. Kwai Wong for his mentorship and advice.

References:

1. Azadifar, S., & Ahmadi, A. (2022). A novel candidate disease gene prioritization method using deep graph convolutional networks and semi-supervised learning. BMC Bioinformatics, 23(1). https://doi.org/10.1186/s12859-022-04954-x . 2. Cai, H., Zheng, V. W., & Chang, K. C. (2018). A Comprehensive Survey of Graph Embedding: Problems, Techniques, and Applications. IEEE Transactions on Knowledge and Data Engineering, 30(9), 1616-1637. 3. Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., Ferrero, E., Agapow, P., Zietz, M., Hoffman, M. M., ... & Greene, C. S. (2018). Opportunities and obstacles for deep learning in biology and medicine. Journal of the Royal Society Interface, 15(141). 4. Goodsell, D. S. (2001). The Molecular Perspective: The Immune System. Stem Cells, 19(5), 453-455. 5. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In Proceedings of the International Conference on Learning Representations (ICLR). 6. Lerner, A., Jeremias, P, & Matthias, T. (2016). The World Incidence and Prevalence of Autoimmune Diseases is Increasing. International Journal of Celiac Disease, 3(4), 151–155. https://doi.org/10.12691/ijcd-3-4-8 . 7. Li, Y., Kuwahara, H., Yang, P., Song, L., & Gao, X. (2019). PGCN: Disease gene prioritization by disease and gene embedding through graph convolutional neural networks. https://doi.org/10.1101/532226 . 8. Lu, G., Hao, X., Chen, W.-H., & Mu, S. (2018). GAAD: A Gene and Autoimmune Disease Association Database. Genomics, Proteomics & Bioinformatics, 16(4), 252–261. https://doi.org/10.1016/j.gpb.2018.05.001 . 9. Mahler, E. R., & Bentzen, S. M. (2016). Emerging Clarity on the Mechanistic Aspects of Autoimmune Diseases. Journal of Autoimmunity, 70, 1-9. 10. McCarthy, S., & Sullivan, P. F. (2020). The Role of Genome-Wide Association Studies (GWAS) in Complex Trait Genetics. Trends in Genetics, 36(6), 398-409. 11. Rojas-Villarraga, A., Amaya-Amaya, J., Mantilla, R. D., & Anaya, J. M. (2010). The Importance of Comorbidity Indices in Autoimmune Diseases. Rheumatology, 49(1), 15-27. 12. Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Computer Science, 2(3), 1–21. Springer. https://doi.org/10.1007/s42979-021-00592-x . 13. Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., & Welling, M. (2018). Modeling Relational Data with Graph Convolutional Networks. In Proceedings of the 15th European Semantic Web Conference (pp. 593-607). 14. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Philip, S. Y. (2021). A Comprehensive Survey on Graph Neural Networks. IEEE Transactions on Neural Networks and Learning Systems, 32(1), 4-24. 15. Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. Insights into Imaging, 9(4), 611–629. https://doi.org/10.1007/s13244-018-0639-9