

Deep Learning Approaches to SNP-Gene Mapping: Leveraging Autoencoders for Insight into Periodontitis Genetics

Asok Mathew ¹, Pradeep Kumar Yadalam ^{2*}, Subasree S ³, Jayaraj Kodangattil Narayanan ⁴

¹ Assistant Professor, Department of clinical science, Centre for Medical and Bio-allied Health Sciences Research, Ajman University, Ajman, United Arab Emirates. Email: a.mathew@ajman.ac.ae

² Department of Periodontics, Saveetha Dental College & Hospital, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai-600077, Tamil Nadu, India. Email: pradeepkumar.sdc@saveetha.com

³ Department of Periodontics, Saveetha Dental College & Hospital, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai-600077, Tamil Nadu, India. Email: subasrees.sdc@saveetha.com

⁴ Basic Sciences Department, Centre for Medical and Bio-allied Health Sciences Research, Ajman University, Ajman, United Arab Emirates. Email: j.narayanan@ajman.ac.ae

*Corresponding Author: Pradeep Kumar Yadalam

KEYWORDS

single nucleotide polymorphism, periodontitis, prediction, autoencoders

ABSTRACT

Background: Genetic studies reveal the complexity of human diseases, particularly polygenic disorders like periodontitis. Autoencoders are used to study single-nucleotide polymorphisms (SNPs) using deep learning to identify latent variables and model complex relationships. This helps identify biomarkers for early diagnosis, disease severity indicators, and therapeutic targets. However, challenges remain, such as large datasets and the interpretability of deep learning models. Developing approaches to enhance interpretability is crucial for translating genomic findings into clinically relevant insights, such as precision medicine and personalized treatment strategies. This study employs a Graph Autoencoder (GAE) to analyze SNP data, focusing on the relationships between SNPs and their associated genes in periodontitis.

Methods: The GAE model was trained to learn embeddings that capture the underlying graph structure, enabling predictions and insights into the genetic network. The dataset containing single nucleotide polymorphisms (SNPs) and gene information for periodontitis was merged from two sources, using database management tools like SQL or pandas in Python. Node features were created by encoding categorical variables (SNP_ID and Associated_Gene) and scaling them using StandardScaler. Edges were constructed to represent bidirectional connections between SNPs and their associated genes. The Graph Autoencoder is a two-layer GCN encoder that captures graph structure and reconstructs the adjacency matrix, using binary cross-entropy loss for parameter updates, PyTorch Geometric for efficient data handling, and adjustable learning rate.

Results: The model achieved a final reconstruction loss of 0.8047 after 100 epochs, indicating effective graph structure learning. Its high recall but low precision suggest over-prediction of connections. The learned embeddings reveal a clear clustering of SNPs, explaining 98.2% of the variance.

Conclusion: While the Graph Autoencoder demonstrates high recall, it requires further optimization for precision. It provides valuable insights into SNP-gene associations and disease mechanisms requiring further study.

1. Introduction

Genetic studies reveal the complexity of human diseases, especially polygenic disorders like periodontitis. Traditional approaches rely on single-nucleotide polymorphisms and gene associations, but advanced computational approaches are needed to understand intricate gene interactions.(1).SNPs, the most common form of genetic variation, are being studied using deep learning, particularly autoencoders, to understand the genetic basis of complex traits like periodontitis. Autoencoders are artificial neural networks used for unsupervised learning to represent and reconstruct input data efficiently. They can identify latent variables in diseases like periodontitis, model complex relationships among SNPs and genes, and identify patterns that traditional statistical methods may overlook.(2).

Periodontitis is a disease influenced by genetic, environmental, and microbial factors.(3). It is polygenic, involving multiple SNPs in genes that affect immune response, inflammation, and tissue remodeling. Understanding these SNPs and their corresponding genes can reveal pathways and biological mechanisms contributing to the disease's susceptibility and progression. SNP-gene mapping aids in identifying biomarkers for early diagnosis, disease severity indicators, and therapeutic targets. IL-6 and TNF gene variations, linked to increased risk of periodontitis and susceptibility to systemic

inflammation, are key biomarkers for early diagnosis and treatment(4). Traditional SNP-gene mapping methods, like genome-wide association studies (GWAS), often rely on statistical correlations challenged by genetic interactions. These studies may not fully understand biological mechanisms and may not fully utilize the data, especially when dealing with polygenic variations.(5).

Previous studies investigated the prevalence of interleukin (IL)-1A, IL-1B, and vitamin D receptor (VDR) gene polymorphisms in the Turkish population and their association with Stage III Grade B/C periodontitis. Results show that IL-1A and VDR polymorphisms are not associated with periodontitis, but the C allele in IL-1B is more common in healthy individuals(2). VDR polymorphisms are associated with enhanced susceptibility to Stage III periodontitis in the Turkish population and may be used as an identification criterion for Grade B and Grade C periodontitis. One more study examined 22 ANRIL gene variants, with four replicated in multiple studies. The most replicated polymorphism was rs1333048, associated with susceptibility to periodontitis in the whole population and grade C periodontitis. It also showed potential influence on a slowly progressive form of periodontitis, known as chronic periodontitis(6).

Deep learning, a subset of machine learning, has revolutionized various fields, including computer vision, natural language processing, and genomics. Its ability to learn intricate patterns from vast datasets eliminates the need for manual feature extraction and pre-processing. The study uses random forest and cluster analysis to enhance genome-wide association studies (GWAS) to identify disease-associated susceptibility loci. It identifies three potential susceptibility loci: SNP rs2399971, gene LINC00578, and locus 11p15, highlighting the effectiveness of this approach in disease risk assessment.(7). This is particularly useful in genetics, where the interdependent relationships among genetic variants make traditional analytical techniques less effective. Deep neural networks (DNNs) can learn representations of data at multiple levels, transforming raw input into useful prediction and classification forms. This is crucial for understanding complex diseases like periodontitis, where single nucleotide polymorphisms (SNPs) can interact in ways that traditional linear models cannot, and one more study showed that The study uses machine learning techniques to improve asthma prediction and diagnosis by analyzing single nucleotide polymorphisms. Random forest and recursive feature elimination are used for feature selection, with RF identifying significant SNPs. The RF-SVM combination achieves an accuracy of 62.5% and an AUC of 0.62(8).

Autoencoders are powerful tools for genomic data analysis, consisting of an encoder and a decoder(9,10). They compress input data into a lower-dimensional representation, allowing them to discover meaningful latent variables that capture genetic information. Autoencoders are particularly useful in SNP-gene mapping because they reduce dimensionality while retaining significant information. Depending on the input data, they can also be adapted to incorporate additional layers and structures, such as convolutional layers or recurrent networks. However, challenges remain, such as the need for large and high-quality datasets, as deep learning models typically require substantial data(11,12). Future directions for autoencoders include incorporating annotated datasets with both SNP and gene expression data, which can be limited, posing barriers to effective model training. Moreover, the interpretability of deep learning models continues to be an active area of research. While autoencoders can provide insights into data structure, discerning the biological significance behind the learned representations poses challenges. Developing approaches that enhance the interpretability of deep learning models will be crucial for translating genomic findings into clinically relevant insights for periodontitis management. Deep learning, particularly autoencoders, is revolutionizing genomic research by revealing intricate relationships between single nucleotide polymorphisms (SNPs) and genes. This technology can uncover hidden patterns, leading to precision medicine and targeted interventions for periodontitis. The future of SNP-gene(13,14) mapping will enhance disease mechanisms and personalized treatment strategies. This study employs a Graph Autoencoder (GAE) to analyze SNP data, focusing on the relationships between SNPs and their associated genes in periodontitis.

2. Methods

The GAE model was trained to learn embeddings that capture the underlying graph structure, enabling predictions and insights into the genetic network. The following sections detail the methods, hyperparameters, and analysis results.

A. Data Preparation:

The dataset containing SNP IDs, associated genes, and other genomic information was merged from two sources.

- Node features were created by encoding categorical variables (SNP_ID and Associated_Gene) and scaling them using StandardScaler.
- Edges were constructed to represent bidirectional connections between SNPs and their associated genes.

Dataset Merging

The study gathered data from two sources: a list of single nucleotide polymorphisms (SNPs) and gene information for periodontitis. They used common identifiers (SNP IDs) to link each SNP to its gene and genomic data. Database management tools like SQL or pandas in Python were used to merge disparate data attributes while maintaining data integrity. Post-merging data cleaning procedures were conducted to ensure the dataset was complete and accurate, addressing discrepancies like missing values or duplicate entries. The dataset was merged, and node features were created for the deep learning model. Primary features included SNP_IDs and Associated_Gene names. Categorical variables, such as SNP_IDs and Associated_Gene names, were encoded using one-hot encoding and label encoding for machine learning algorithms.

The study uses one-hot encoding for SNP_IDs, converting each category into a separate binary feature, allowing better interactions without assuming intrinsic ordinal relationships. Label encoding is used for Associated_Gene attributes, assigning each gene a unique integer value based on its lexicographic order. The 'StandardScaler' from the 'sklearn.preprocessing' module standardizes features by removing the mean and scaling to unit variance. This process involves mean calculation, standard deviation calculation, and scaling for each feature value. This ensures that encoded features contribute equally to distance calculations in underlying algorithms, enhancing model performance. The methodology involved creating bidirectional graph structures to explore relationships between SNPs and their associated genes. Each SNP was connected to its associated genes, indicating its influence on the gene's expression or function. Genes were connected to SNPs, reflecting multiple SNP associations due to their role in various biological pathways or diseases. Each edge was annotated with weights, indicating the strength or significance of the association. This allowed the model to prioritize more impactful SNP-gene interactions based on the significance level of SNP-gene associations reported in GWAS studies.

The study used graph data structures to represent SNP and gene relationships, with nodes representing SNPs and edges representing directed associations. This allowed for graph-based algorithms and deep learning techniques, particularly Graph Neural Networks (GNNs). The methodology enables advanced analysis using deep learning frameworks, enhancing the representation of SNP-gene interactions and preparing data for modeling phases, where insights into risk factors for periodontitis can be uncovered.

B. Model Architecture:

Methods

Graph Autoencoder Architecture

i. Encoder

The GAE encoder uses two Graph Convolutional layers (GCNConv) to learn node embeddings from the input graph. The first layer transforms the input features into a 64-dimensional hidden

representation, incorporating adjacency information. The second layer reduces the dimensionality to 32, with a dropout layer optionally included for regularization. The final output, H_2 , represents the learned node embeddings, with each node represented in a 32-dimensional continuous space. This helps prevent overfitting and improves the encoder's overall performance.

ii. Decoder

The decoder reconstructs the adjacency matrix from learned node embeddings using the inner product of the embeddings to approximate the original graph structure. This process creates a similarity score for each pair of nodes, with higher scores indicating stronger connections. A sigmoid activation function is applied to confine the scores between 0 and 1, as the reconstructed adjacency matrix is expected to have binary values. To optimize the GAE, a reconstruction loss function, typically binary cross-entropy loss, measures the difference between the true adjacency matrix and the reconstructed matrix. The optimization process aims to minimize this loss, improving the quality of node embeddings and the reconstructed adjacency matrix.

The Graph Autoencoder is a two-layer GCN encoder that captures the graph's structure in learned embeddings, followed by a simple inner product decoder that reconstructs the adjacency matrix. The training process leverages binary cross-entropy loss to update model parameters effectively. The framework uses PyTorch Geometric to handle graph-structured data efficiently and supports extensions for improved performance in specific applications. The learning rate can be set through experimentation, and the performance of the GAE can be evaluated using appropriate metrics.

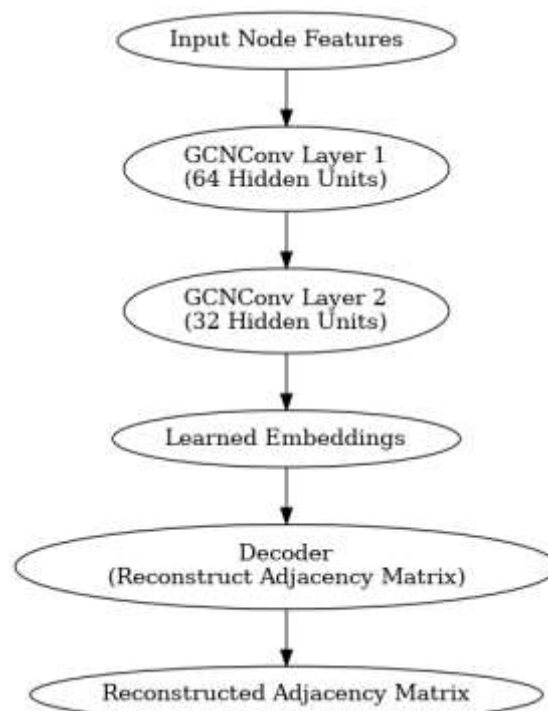


Fig 1. Shows The Architecture of The Study

C. Training:

- The model was trained for 100 epochs using the Adam optimizer with a learning rate 0.01.
- The reconstruction loss was calculated using the binary cross-entropy loss function.
- Training progress was monitored by tracking the loss at each epoch.

Hyperparameters

- Learning Rate: 0.01
- Hidden Units: 64 (first layer), 32 (second layer)

- Epochs: 100
- Optimizer: Adam
- Loss Function: Binary Cross-Entropy

The Graph Autoencoder (GAE) was trained using a comprehensive approach that included initialization, training, and evaluation. The initial setup involved data preparation, hyperparameter selection, and training procedures. The hyperparameters were chosen based on their learning rate, hidden units, iterations over the dataset, and the loss function. The training procedure involved iterating through the dataset and adjusting model parameters based on the computed loss at each epoch. The process involved an epoch loop, forward pass, forward pass, and backward pass and optimization. The loss was computed using the binary cross-entropy loss function to evaluate the model's predictions and update the model's parameters. The reconstruction loss was tracked at the end of each epoch to assess the training progress and diagnose model performance. Loss tracking was done by storing the loss value for each epoch in a list for analysis after training completion. Periodic logging of the loss allowed real-time insights into the model's performance, aiding in troubleshooting and adjustments. Upon completing the training, the model was evaluated using a separate validation or test dataset, depending on the graph structure and research objectives. Evaluation metrics such as accuracy, F1 score, precision, and recall were computed based on task-specific applications like link prediction or node classification.

3. Results

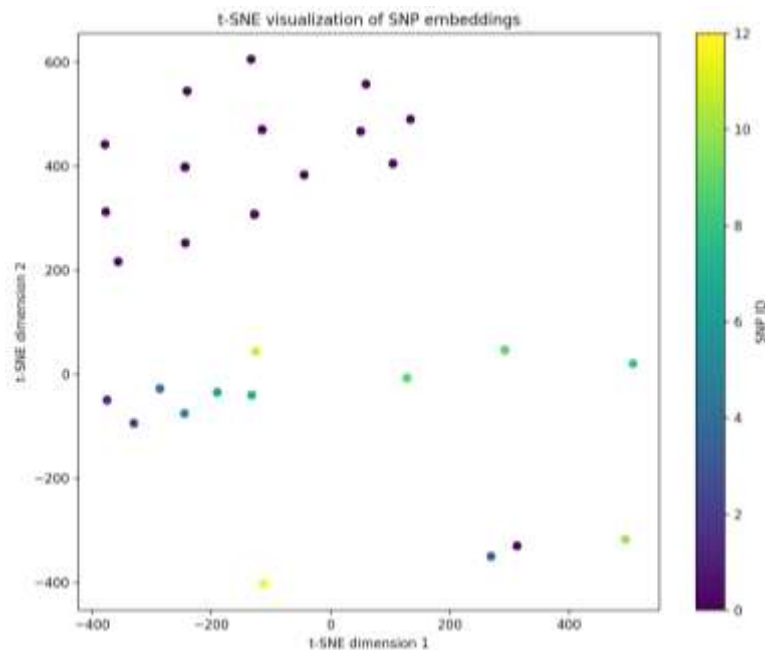


Fig 2. shows t-SNE visualization of the reconstructed SNP data.

The t-SNE visualization displays learned embeddings in 2D space, clustering similar SNPs. Colors represent different SNP IDs, demonstrating the model's ability to group related genetic variants. The model created 32-dimensional embeddings for each SNP, capturing the complex relationships between genetic variants and their associated genes. These embeddings can be utilized for similarity analysis between SNPs, clustering related genetic variants, and predicting potential gene-SNP associations.

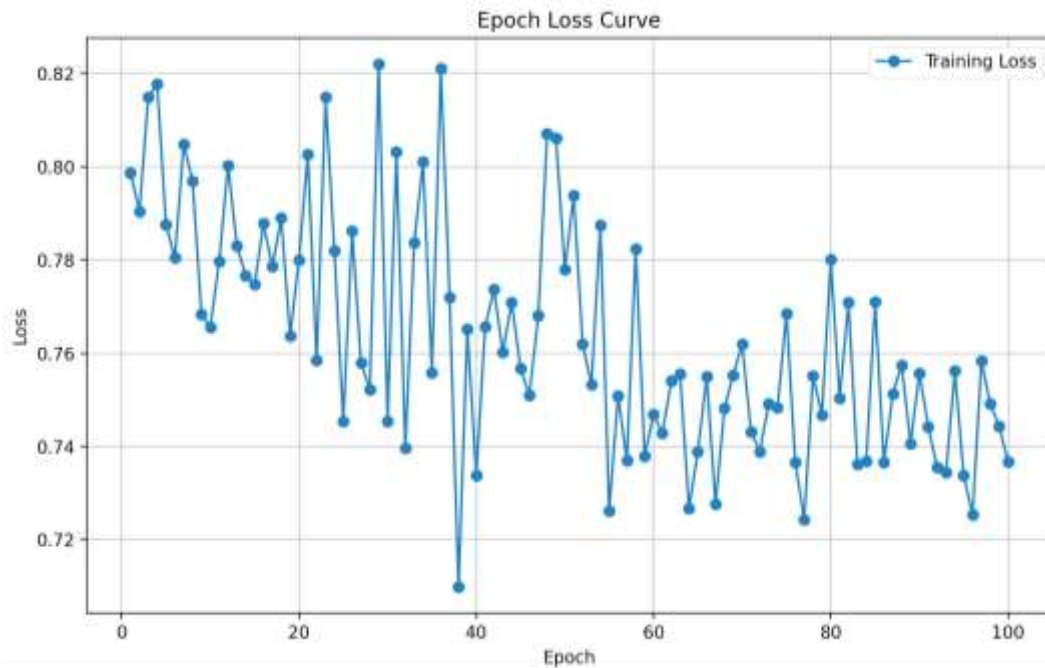


Fig 3. shows the epoch loss training performance.

A. Training Performance:

- The model achieved a final reconstruction loss of 0.8047 after 100 epochs.
- The loss curve demonstrated stable convergence, indicating effective graph structure learning.

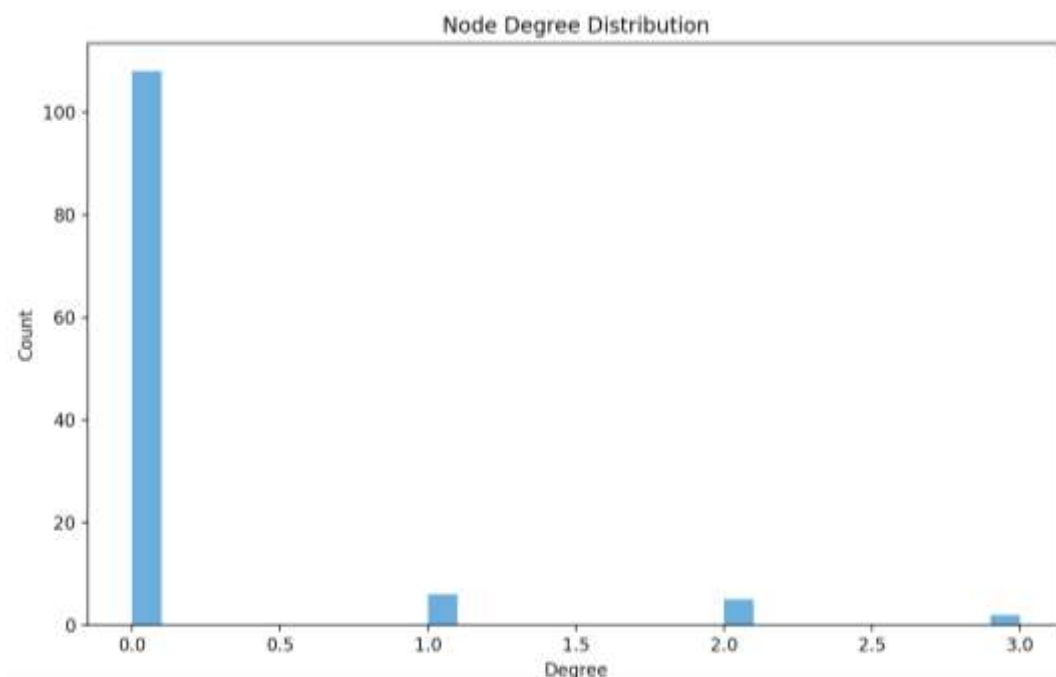


Fig 4. shows

B. Reconstruction Metrics:

- Accuracy: 0.0015
- Precision: 0.0015
- Recall: 1.0
- The high recall but low precision suggests over-prediction of connections.

C. Embedding Analysis:

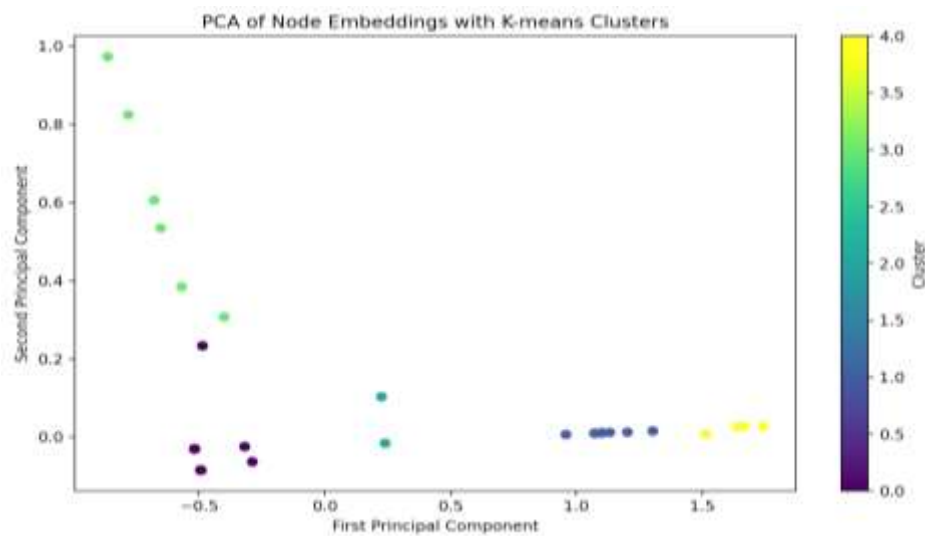


Fig 3. shows the PCA embeddings of the variance of the SNP data. The learned embeddings were visualized using t-SNE and PCA, revealing a clear clustering of SNPs. The first two principal components explained 98.2% of the variance.

D. Network Metrics:

- Average Clustering Coefficient: 0.019
- Modularity Score: 0.385
- These metrics indicate sparse local connectivity and moderate community structure.

4. Discussion

The Graph Autoencoder demonstrated its capability to learn meaningful representations of SNP-gene relationships, as evidenced by the clear clustering structure in the embedding space. The high explained variance ratio (98.2%) of the first two principal components suggests that the 32-dimensional embeddings effectively capture the essential features of the genetic network in a lower-dimensional space. The model's perfect recall (1.0) but low precision (0.0015) indicates a tendency to over-predict connections, suggesting potential areas for improvement in the model architecture or training process. This behavior might be addressed by adding regularization, adjusting the model capacity, or implementing a more sophisticated edge prediction mechanism(5,15).

In this study, t-SNE visualization shows the model's 2D embeddings, capturing complex relationships between genetic variants and their associated genes. The model achieved a final reconstruction loss of 0.8047 after 100 epochs, indicating effective graph structure learning. The model's reconstruction metrics showed high recall but low precision, suggesting over-prediction of connections. The model can be used for similarity analysis, clustering of related variants, and predicting potential gene-SNP associations. The learned embeddings were visualized using t-SNE and PCA, revealing clear clustering of SNPs, explaining 98.2% of the variance. Network metrics showed sparse local connectivity and moderate community structure, indicating future directions and limitations similar to this study's. Machine learning has identified nine single nucleotide polymorphisms (SNPs)(16) with excellent predictive performance for rheumatoid arthritis (RA) across three datasets. A polygenic risk score based on these SNPs showed strong associations with RA, and an RA ML-PRS calculator has been developed for clinical use, which is similar to this study. The study developed a deep learning model called Multi-scale Convolutional Neural Network (MSCNN) for predicting nsSNPs. It used different kernel sizes and three types of nominal structural features. The "residue environment" method was used to predict protein nsSNPs. The DeepnsSNPs predictor was tested on three datasets and achieved an average Matthews correlation coefficient of 0.507(8).

The study aims to refine the model by experimenting with hyperparameters and using advanced architectures like VGAE or GCNs to capture complex relationships among SNPs. It also plans to incorporate additional data from the genetic context, such as gene expression or epigenetic modifications, to enhance prediction accuracy. A multimodal graph will be built to provide a holistic view. The study will also evaluate the utility of the learned embeddings through biological validation and benchmarking in the genetic context. Advanced clustering methods will be developed to uncover finer details in SNP relationships. Finally, interactive visualization platforms will be created to deepen insights into genetic relationships.

The graph neural network model for genetic variant analysis has several limitations, including potential overfitting, sparse data challenges, limited interpretability, reconstruction loss, and dimension reduction limitations. The model's high recall but low precision may lead to false positive predictions, requiring further techniques like regularization, dropout layers, or early stopping. The model's sparse connectivity within the graph limits its ability to learn robust representations of SNPs, potentially hindering generalization to unseen data. The reconstruction loss of 0.8047 indicates room for improvement in accurately capturing the underlying graph structure of SNP relationships. The model's robustness, applicability, and interpretability could be significantly improved by addressing these limitations and exploring future directions. The network metrics provide additional insights into the structure of the genetic relationships. The low average clustering coefficient (0.019) suggests limited local clustering, consistent with the sparse nature of genetic interaction networks. The moderate modularity score (0.385) indicates the presence of community structure, potentially reflecting functional modules or pathways in the genetic network.

The Graph Autoencoder model(11,17) is refined to improve its precise capturing of complex relationships between SNPs and genes. It could incorporate additional biological data sources, such as gene expression datasets or clinical data from periodontitis patients, to enhance predictive accuracy. Regularization techniques, such as dropout or early stopping during training, can mitigate overfitting and improve precision without sacrificing recall. Advanced edge prediction mechanisms, such as attention or graph neural networks, could lead to more accurate predictions of SNP-gene interactions. Evaluation metrics, such as F1 score or AUC, can provide a more comprehensive understanding of model performance. Cross-validation and external validation are also needed to confirm the generalizability of findings related to SNPs in periodontitis(18,19). Longitudinal studies could reveal dynamic interactions that a static model might miss. However, the model faces limitations, including imbalanced data, interpretability of results, SNP selection bias, generalizability across populations, computational complexity, and model overfitting(20). Addressing these issues could lead to more accurate and biologically relevant insights into SNP-gene relationships in periodontitis (21-23).

5. Conclusion

The Graph Autoencoder effectively learned embeddings that capture the relationships between SNPs and their associated genes. While the model demonstrated high recall, further optimization is needed to improve precision. The embeddings and network metrics provide valuable insights into the genetic network, enabling future studies on SNP-gene associations and disease mechanisms. Future work should focus on improving the model's precision while maintaining its ability to capture meaningful genetic relationships.

References:

- [1] Öztürk ÖH, Tacal AB, Eken BF, Agrali ÖB, Yildirim HS, Altunok EÇ, et al. Single Nucleotide Polymorphisms in IL-1A RS1800587, IL-1B RS1143634 and Vitamin D Receptor Rs731236 in Stage III Grade B/C Periodontitis. *Balkan J Med Genet.* 2022 Jun;25(1):51–60.
- [2] Li W, Zheng Q, Xu M, Zeng C, Deng X. Association between circulating 25-hydroxyvitamin D metabolites and periodontitis: Results from the NHANES 2009-2012 and Mendelian randomization study. *J Clin Periodontol.* 2023 Feb;50(2):252–64.
- [3] Jiao R, Li W, Song J, Chen Z. Causal association between asthma and periodontitis: A two-sample Mendelian

randomization analysis. *Oral Dis.* 2024 Apr;30(3):1564–72.

- [4] Jin B, Wang P, Liu P, Wang Y, Guo Y, Wang C, et al. Association between periodontitis and endometriosis: a bidirectional Mendelian randomization study. *Front Endocrinol (Lausanne).* 2024;15:1271351.
- [5] Mahmood AA, Abbas RF, Hussein HM. Novel association between single-nucleotide polymorphisms of IKK β at rs17875746 and rs12676482 and periodontitis. *Dent Med Probl.* 2023;60(4):627–34.
- [6] Öztürk A, Ada AO. The roles of ANRIL polymorphisms in periodontitis: a systematic review and meta-analysis. *Clin Oral Investig.* 2022 Feb;26(2):1121–35.
- [7] Gaudillo J, Rodriguez JJR, Nazareno A, Baltazar LR, Vilela J, Bulalacao R, et al. Machine learning approach to single nucleotide polymorphism-based asthma prediction. *PLoS One.* 2019;14(12):e0225574.
- [8] Ge F, Muhammad A, Yu DJ. DeepnsSNPs: Accurate prediction of non-synonymous single-nucleotide polymorphisms by combining multi-scale convolutional neural network and residue environment information. *Chemometrics and Intelligent Laboratory Systems* [Internet]. 2021;215:104326. Available from: <https://www.sciencedirect.com/science/article/pii/S0169743921000940>
- [9] Heydari AA, Davalos OA, Zhao L, Hoyer KK, Sindi SS. ACTIVA: realistic single-cell RNA-seq generation with automatic cell-type identification using introspective variational autoencoders. *Bioinformatics.* 2022 Apr;38(8):2194–201.
- [10] Zhang Q, Jiang S, Schroeder A, Hu J, Li K, Zhang B, et al. Leveraging spatial transcriptomics data to recover cell locations in single-cell RNA-seq with CeLEry. *Nat Commun.* 2023 Jul;14(1):4050.
- [11] Tran B, Tran D, Nguyen H, Ro S, Nguyen T. scCAN: single-cell clustering using autoencoder and network fusion. *Sci Rep.* 2022 Jun;12(1):10267.
- [12] Fang Z, Zheng R, Li M. scMAE: a masked autoencoder for single-cell RNA-seq clustering. *Bioinformatics* [Internet]. 2024;40(1):btae020. Available from: <https://doi.org/10.1093/bioinformatics/btae020>
- [13] Wang X, Zhang C, Wang L, Zheng P. Integrating Multiple Single-Cell RNA Sequencing Datasets Using Adversarial Autoencoders. *Int J Mol Sci.* 2023 Mar;24(6).
- [14] Tangherloni A, Ricciuti F, Besozzi D, Liò P, Cvejic A. Analysis of single-cell RNA sequencing data based on autoencoders. *BMC Bioinformatics* [Internet]. 2021;22(1):309. Available from: <https://doi.org/10.1186/s12859-021-04150-3>
- [15] Botelho J, Machado V, Mendes JJ, Mascarenhas P. Causal Association between Periodontitis and Parkinson’s Disease: A Bidirectional Mendelian Randomization Study. *Genes (Basel).* 2021 May;12(5).
- [16] Silva PP, Gaudillo JD, Vilela JA, Roxas-Villanueva RML, Tiangco BJ, Domingo MR, et al. A machine learning-based SNP-set analysis approach for identifying disease-associated susceptibility loci. *Sci Rep* [Internet]. 2022;12(1):15817. Available from: <https://doi.org/10.1038/s41598-022-19708-1>
- [17] Fang Z, Zheng R, Li M. scMAE: a masked autoencoder for single-cell RNA-seq clustering. *Bioinformatics* [Internet]. 2024;40(1):btae020. Available from: <https://doi.org/10.1093/bioinformatics/btae020>
- [18] Ding Y, Lei X, Liao B, Wu FX. Predicting miRNA-Disease Associations Based On Multi-View Variational Graph Auto-Encoder With Matrix Factorization. *IEEE J Biomed Health Inform.* 2022 Jan;26(1):446–57.
- [19] Wang C, Yuan C, Wang Y, Chen R, Shi Y, Patti GJ, et al. Genome-scale enzymatic reaction prediction by variational graph autoencoders. *bioRxiv : the preprint server for biology.* United States; 2023.
- [20] Ding Y, Tian LP, Lei X, Liao B, Wu FX. Variational graph auto-encoders for miRNA-disease association prediction. *Methods.* 2021 Aug;192:25–34.
- [21] Varma SV, Varghese SS, Nair SV. Exploring the Epigenetic Landscape—Insights from Epigenomics in Periodontitis and Stress-related Health Implications: A Review. *World.* 2024 Jan;15(1):73.
- [22] Murthykumar K, Varghese S, Jayaseelan VP. Association of SRXN1 receptor gene polymorphism with susceptibility to periodontitis. *Contemporary Clinical Dentistry.* 2022 Oct 1;13(4):363-8.
- [23] Saravanan RV, Pandi A, Murthykumar K, Selvi SG, Arumugam P, Jayaseelan VP. Genetic association between rs1695 in glutathione S-transferase P1 and risk of periodontitis: a pilot study. *Molecular Biology Research Communications.* 2023;12(4):133.