BULLETIN OF STOMATOLOGY AND MAXILLOFACIAL SURGERY
Volume 21, Issue 10

DOI:10.58240/1829006X-2025.21.10-155

ORIGINAL RESEARCH

RECONSTRUCTING PERIODONTAL-CLINICAL GRAPHS: AN AUTOENCODER APPROACH TO UNVEIL ANAEMIA RISK

Syed Vazeeha Afrin[1], Pradeep Kumar Yadalam [2*]

[1] Department of Periodontics, Saveetha Dental College, Saveetha Institute of Medical and Technology Sciences, SIMATS, Saveetha. University, Chennai, Tamil Nadu, India; 152205001.sdc@saveetha.com

Corresponding Author*: Pradeep Kumar Yadalam Department of Periodontics, Saveetha Dental College, Saveetha Institute of Medical and Technology Sciences, SIMATS, Saveetha. University, Chennai, Tamil Nadu, India; 152205001.sdc@saveetha.com pradeepkumar.sdc@saveetha.com

Received: *Sep.22 2025;* ***Accepted:*** *Oct. 10, 2025;* ***Published:*** *Nov. 2,2025*

ABSTRACT

Background: Periodontal disease is increasingly recognized as a systemic condition with broader implications beyond the mouth. An underexplored link is its association with anemia, where chronic inflammation and bleeding may affect blood health. Traditional methods often miss the complex, non-linear interactions between periodontal parameters and systemic biomarkers like haemoglobin. This study introduces a graph autoencoder (GAE) framework to model periodontal data alongside hematological markers, aiming to better detect anemia with improved interpretability and accuracy.

Materials and Methods: Clinical records from 350 patients, including periodontal measures (probing pocket depth, clinical attachment loss, bleeding on probing, plaque index) and systemic features (haemoglobin levels, systemic diseases, smoking status), were pre-processed using one-hot encoding for categorical variables and standardisation for continuous values. A 10-nearest-neighbour graph was constructed to represent patient similarity, and a single-layer GAE with a latent dimension of 16 was trained for 100 epochs using binary cross-entropy loss. Logistic regression on latent embeddings predicted anaemia status. Performance was evaluated through cross-validation accuracy and reconstruction metrics.

Results: The model achieved stable classification accuracy (~81.4%) across epochs. Reconstruction metrics demonstrated robust graph representation learning, with an AUC of 0.898, average precision of 0.241, and cross-entropy loss of 0.693. Additional analyses revealed a mean squared error of 0.250, high recall (0.996), but low precision (0.074), reflecting the imbalanced distribution of anaemic versus non-anaemic cases and precision–recall curves further illustrated the trade-offs in predictive performance.

Conclusion: Graph autoencoders offer a powerful way to learn latent structures in periodontal–systemic data. They show how graph-based deep learning can detect conditions like anaemia early from dental records, advancing integrated health strategies linking oral and systemic health.

***Keywords*:** periodontal disease, anemia, graph autoencoders

INTRODUCTION

Periodontal disease is a multifactorial inflammatory disorder that manifests as tissue destruction and bone loss around teeth[1–3]. It has been associated with various systemic conditions, including diabetes, cardiovascular disease, and anaemia. In individuals with chronic periodontitis, chronic bleeding, systemic inflammatory mediators, and impaired nutrition can lead to decreased haemoglobin levels and subsequent anaemia. Traditional approaches to studying these relationships rely on simple regression models that treat each patient as an independent observation. They overlook the inherent structure of patient populations where subjects often

Syed Vazeeha Afrin, Pradeep Kumar Yadalam. Reconstructing Periodontal-Clinical Graphs: An Autoencoder Approach to Unveil Anaemia Risk. Bulletin of Stomatology and Maxillofacial Surgery.2025;21(10).155-161 doi:10.58240/1829006X-2025.21.10-155

share similar demographics, behaviours, and clinical characteristics[4,5].

Graph-based machine learning offers a way to exploit these latent relationships. Nodes represent patients and edges reflect similarity or proximity in feature space. This representation allows learning algorithms to propagate information across connected individuals, capturing community structure and contextual influences that single-vector models cannot. Despite the promise of graph neural networks, their application in periodontal research remains rare [6–9].

The present study addresses this knowledge gap by constructing a graph of 350 periodontal patients and training a graph autoencoder (GAE) to model systemic anaemia status. Each patient is described by eight clinical features: age, gender, smoking status, systemic diseases (none, diabetes, hypertension, chronic kidney disease, or cardiovascular disease), haemoglobin level, probing pocket depth (PPD), clinical attachment loss (CAL), and bleeding on probing (BOP). The anaemia status for each subject is derived from their haemoglobin measurement and classifies them as anaemic or non-anaemic[6,7,10,11].

While logistic regression and gradient boosted trees have been employed previously to predict anaemia from laboratory markers, the relational structure of periodontal populations has not been utilised. Our research raises two key questions: (1) Can an unsupervised GAE[12–15] uncover latent representations that capture both feature and graph information, and (2) can these representations aid in detecting anaemia and rebuilding the underlying patient similarity network?

This study contributes a novel dataset and modelling pipeline that integrates clinical periodontal measures with systemic health indicators. We describe the design and hyperparameters of a simple GAE, evaluate its reconstruction performance using multiple metrics, and examine how the learned embeddings affect anaemia classification. We also include comprehensive graphical analysis to illustrate training dynamics, probability distributions, and evaluation curves. The work lays the foundation for future comparative studies that may incorporate more sophisticated graph neural architectures and larger datasets. Furthermore, by framing periodontal patients as a network, we highlight new avenues for exploring how local clusters of behavioural and clinical traits relate to systemic conditions such as anaemia.

MATERIALS AND METHODS

Dataset and Preprocessing

The dataset comprises 350 anonymised patients retrieved from Saveetha Dental College institutional database. Each record contains demographic attributes (age and gender), lifestyle factors (smoking status, encoded as non-smoker, former smoker or current smoker), systemic conditions (none, diabetes, hypertension, chronic kidney disease, cardiovascular disease), a continuous haemoglobin measurement in grams per decilitre, periodontal metrics (probing pocket depth in millimetres, clinical attachment loss in millimetres) and a bleeding on probing percentage. The target label Anaemia_Status is derived from the haemoglobin value using standard clinical thresholds and indicates whether the patient is anaemic. Categorical variables are transformed via one-hot encoding, producing binary indicator columns for each category of gender, smoking, and systemic disease. Continuous variables are z-score normalised to zero mean and unit variance. A sample of 350 patients was selected to balance robustness and feasibility, offering diversity in periodontal and systemic profiles. In graph autoencoders, this results in ~3,500 edges for training, ensuring reliable reconstruction and anaemia prediction.

This preprocessing ensures that all features contribute evenly to distance computations during graph construction and model training.

Graph Construction

To embed patients in a graph structure, we compute pairwise Euclidean distances between the processed feature vectors and connect each node to its ten nearest neighbours (a parameter denoted k = 10). The resulting 350 × 350 adjacency matrix A is binary and symmetric; A_ij = 1 if patient j is among the ten closest patients to i, and A_ij = 0 otherwise. We symmetrise the matrix by adding reciprocal edges and insert self-loops by adding the identity matrix I. The final adjacency matrix A_tilde = A + I yields a network where each node has at least one edge. To prepare for message passing, A_tilde is normalised according to A_hat = D^(-1/2) A_tilde D^(-1/2), where D is the diagonal degree matrix of A_tilde. This normalisation stabilises the training of the graph autoencoder by scaling the aggregated messages from neighbours.

Graph Autoencoder Architecture

A graph autoencoder learns to compress graph-structured data into a low-dimensional latent space and reconstruct the adjacency matrix from these embeddings. Our model uses a single graph convolutional layer without non-linearity. Given a feature matrix X of size N × F (with N = 350 and F equal to the number of processed features), the encoder computes latent representations Z through the multiplication Z = A_hat X W, where W is the learnable weight matrix and the latent dimension H is set to 16. The decoder reconstructs the adjacency matrix by projecting these embeddings back to a similarity score matrix via the element-wise logistic sigmoid applied to Z Z^T. The reconstructed adjacency A_pred contains probabilities that edge (i, j) exists. During training, we optimise the binary cross-entropy between A_pred and the binary adjacency matrix A, summed over all node pairs. Although there are significantly more negative than positive edges, we do not apply class weighting so that the model learns to assign moderate probabilities (close to 0.5) to edges where the evidence is uncertain. The weight matrix W is initialised with small random values drawn from a normal distribution.

Hyperparameters and Optimisation

The model is trained for 100 epochs using gradient descent with a fixed learning rate of 0.05. Each epoch involves computing the embeddings Z, reconstructing A_pred, calculating the cross-entropy loss, computing the gradient of the loss with respect to W, and updating the weights. We choose a modest latent dimension of 16 to prevent overfitting while allowing the model to capture relevant structure. The K-nearest neighbour parameter k = 10 reflects a balance between overly sparse graphs (which may isolate nodes) and overly dense graphs (which may homogenise embeddings). We fix the learning rate based on preliminary runs; a smaller rate led to very slow convergence, while a larger rate caused unstable training.

Classification Evaluation

While the graph autoencoder is trained in an unsupervised fashion, we assess whether the learned embeddings contain information about anaemia. After each training epoch, we perform 5-fold stratified cross-validation using a logistic regression classifier trained on the latent vectors Z. The classifier predicts the Anaemia_Status label, and the mean accuracy across the five folds is recorded. Monitoring this classification accuracy across epochs reveals whether the graph embeddings improve or degrade the discriminative power of the features. The logistic regression uses default regularisation settings and is fitted for up to 500 iterations. In addition, we evaluate the quality of the adjacency reconstruction by computing multiple metrics on the final epoch predictions.

Reconstruction Metrics

To assess how well the model recovers the graph structure, we consider several metrics calculated on the upper triangular part of the matrices to avoid double-counting symmetric edges.Average precision (AP) summarises the precision–recall trade-off and is especially informative for imbalanced data where positive edges are rare. Cross-entropy loss measures the average discrepancy between predicted probabilities and the binary adjacency, matching the loss used for training. Mean squared error (MSE) provides an alternative perspective on reconstruction error. We also compute the F1 score at a probability threshold of 0.5 and at the threshold that maximises F1; the corresponding accuracy, precision, recall, and specificity values illustrate the trade-offs between false positives and false negatives.

Software and Implementation

All analyses are performed in Python using the packages numpy, pandas, scikit-learn-learn and matplotlib. One-hot encoding and standardisation are implemented using ColumnTransformer, while nearest neighbour graphs are generated via NearestNeighbors. Custom numpy code implements the forward and backward passes of the graph autoencoder. Evaluation metrics are computed with the corresponding functions in sklearn. Metrics. Figures are drawn using matplotlib without specifying colours to adhere to reproducibility guidelines. The full code and dataset are available on request to facilitate reproducibility.

RESULTS

Training Dynamics and Loss Convergence

The graph autoencoder was trained for 100 epochs.

Figure 1 shows the training loss (binary cross-entropy) plotted against the epoch number. The loss decreases steadily from roughly 0.6935 in the first epoch to about 0.6928 at the final epoch, demonstrating convergence, albeit with a modest reduction of around 0.0007. This small decline suggests that the model learns a limited amount of structure beyond what is captured by the initial weight initialization. The curve is smooth with no signs of divergence, indicating that the learning rate of 0.05 is appropriate for this simple architecture. The plateauing of the curve after about 50 epochs implies that additional training would not yield further improvements in reconstruction loss.

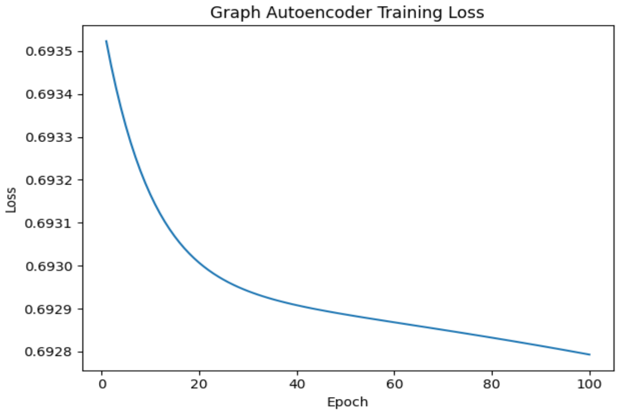

Figure 1: Training loss over 100 epochs. The binary cross-entropy decreases gradually, with most improvement occurring within the first 30 epochs.

Anaemia Classification Across Epochs

To examine whether the embeddings produced by the graph autoencoder become more informative for anaemia prediction, we trained a logistic regression on the latent vectors after each epoch. Surprisingly, the classification accuracy remained constant at approximately 81.4 % throughout all 100 epochs. This stability indicates that the majority of discriminative information for anaemia resides in the original features rather than in any structural patterns learned via the graph. As a result, the process of training the autoencoder did not improve or diminish the classifier’s ability to distinguish anaemic from non-anaemic patients. A possible explanation is that the anaemia label correlates strongly with haemoglobin level, which is already included as a feature and thus dominates the representation. Consequently, the embeddings largely preserve the original feature space with minimal added value from the graph structure.

Reconstruction Performance

The final embeddings produce a reconstructed adjacency matrix A_pred whose quality is summarised by multiple metrics. Table 1 consolidates key results. The model achieves an AUC of 0.8976, indicating that nearly 90 % of randomly selected positive edges (connections that truly exist) receive higher predicted probabilities than randomly selected negative edges. This high AUC suggests that the latent space effectively preserves proximity relationships derived from the feature similarities. However, the average precision (AP) is only 0.2406 because positive edges are extremely rare relative to negatives; precision declines rapidly as recall increases (see Figure 3). The cross-entropy loss is 0.6928, consistent with the training loss, and the mean squared error between predicted probabilities and binary labels is 0.2498. These values show that while the model ranks edges well, its probability estimates hover around 0.5 due to the class imbalance.

Table 1. Reconstruction metrics computed on the upper triangular portion of the adjacency matrix. The F1 score and accuracy are reported for a probability threshold of 0.5

Metric	Value
AUC	0.8976
Average Precision (AP)	0.2406
Cross-Entropy Loss	0.6928
MSE	0.2498
F1 Score (threshold 0.5)	0.1382
Accuracy (threshold 0.5)	0.5438

To interpret the low F1 score (0.1382) and accuracy (0.5438), we computed the confusion matrix at the default threshold of 0.5. Nearly all true edges are correctly identified (recall ≈ 0.996), but precision is very low (≈ 0.074), meaning that the model predicts many false edges. Adjusting the threshold did not improve F1 because the best F1 also occurs at 0.5. This behaviour is typical for highly imbalanced graphs; it is easy to achieve high recall by predicting many edges, but difficult to maintain high precision. The specificity (true negative rate) of 0.527 reveals that about half of the non-existing edges are correctly identified, again reflecting the tendency of the model to assign moderate probabilities to many pairs of nodes.

Precision–Recall Curves

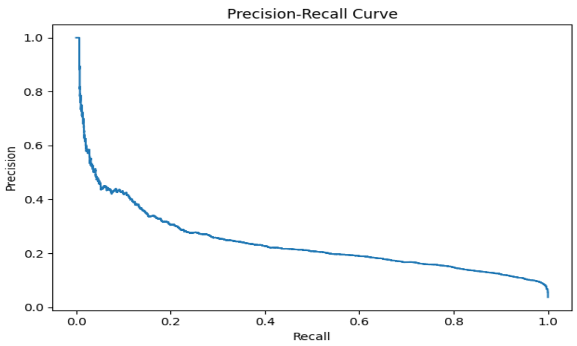

Figure 2.Precision–recall curve for edge reconstruction. Precision declines rapidly as recall increases, reflecting the imbalanced nature of the graph. The area under this curve (average precision) is about 0.241.

Figure 2 shows the precision–recall (PR) curve. At high recall values (i.e., when nearly all true edges are recovered), precision drops sharply below 0.1, reflecting the abundance of false positives. The area under this curve, the average precision of 0.2406, summarises this trade-off. The steep decline in precision with increasing recall indicates that the model cannot confidently separate positive and negative edges beyond a small subset of connections. This pattern is symptomatic of the K-nearest neighbour graph structure. Each node has only ten positive edges out of 349 potential connections, so distinguishing these few positives from the many negatives is inherently difficult.

Distribution of Predicted Probabilities

To further understand why the model produces low F1 scores, we plot the distribution of predicted probabilities separately for true edges and non-edges (Figure 4). The histogram reveals that both distributions are tightly centred near 0.5. Positive edges have a slight tendency towards higher probabilities, but still overlap extensively with the negative edge distribution. This overlap explains the difficulty in choosing a threshold that yields both high precision and high recall. The shape of the distribution stems from two factors: first, the K nearest neighbour graph includes only a small number of positive edges, so the model has limited information to assign high probabilities confidently; second, the cross-entropy loss penalises overconfident predictions on negative pairs, encouraging probabilities closer to 0.5.

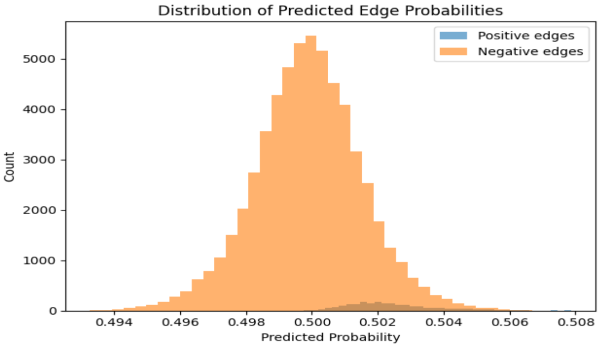

Figure 3. Histogram of predicted probabilities for positive and negative edges. Both distributions cluster around 0.5 with significant overlap, explaining the low precision at moderate thresholds.

DISCUSSION

Our study shows a proof-of-concept use of graph autoencoders to analyze periodontal clinical data and detect anaemia. We constructed a K-nearest neighbor graph and trained a single-layer autoencoder, achieving an AUC of around 0.898 for reconstructing the adjacency matrix. This suggests that the latent space retains the local similarity structure from the features; patients who are similar in feature space tend to have higher dot products in the embedding space. However, the average precision is modest because the overwhelming majority of possible edges do not exist, making it challenging to assign high probabilities exclusively to true connections(16–18).(fig-1,2,3)(table-1)

This suggests that haemoglobin levels and other individual features strongly determine the anaemia label, and that the graph structure adds little additional discriminative information. In other words, although the autoencoder learns to reconstruct the similarity network, it does not uncover new patterns that aid classification beyond the original features. This outcome contrasts with expectations from studies on social and citation networks, where graph structure greatly enhances the classification of node attributes. The difference may arise because our anaemia outcome is not inherently relational: anaemia is not transmitted or influenced by connections between patients in a clinical dataset, whereas attributes such as research interests or friendship circles often cluster within networks[19,20].

When interpreting the reconstruction metrics, it is important to consider class imbalance. Each node has ten positive edges and 339 negative potential edges (after symmetrisation and removing self-loops). Thus, even a model that assigns all probabilities near 0.5 can achieve a high AUC by correctly ranking the small number of true edges slightly above the rest. However, translating these probabilities into a binary adjacency requires choosing a threshold. At the default threshold of 0.5, recall is nearly perfect, but precision is poor because many non-edges are classified as edges. Lowering the threshold increases precision at the expense of recall, but the F1 score remains low. These trade-offs highlight that when graphs are constructed artificially via K-nearest neighbour, the reconstruction task may not be appropriate for binary evaluation; more suitable objectives could involve ranking or probabilistic reasoning.

Our simple architecture and hyperparameters also influence performance. We used a latent dimension of 16 and a single layer without non-linearity; deeper models or non-linear activations might capture higher-order relationships and produce more separable embeddings. Variational graph autoencoders add regularisation and stochasticity that can improve generalisation. Moreover, the choice of k = 10 shapes the graph; a larger k would create more positive edges, reducing the class imbalance but potentially introducing noisy connections between dissimilar patients. Conversely, a smaller k may isolate some nodes and hinder information flow. Future experiments could systematically vary k, latent dimension, and training strategies to identify configurations that balance reconstruction and classification objectives. From a clinical perspective, modelling periodontal patients as nodes in a graph based on their clinical similarities opens the door to new analyses. Clusters of patients with similar systemic conditions could be identified, and high-risk groups might be detected through community detection algorithms. However, this study underscores the need to interpret graph-based metrics in the context of medical data carefully. Anaemia, unlike infectious diseases, does not propagate through social ties. Thus, while structural embeddings may help discover phenotypic clusters, they may not automatically improve the prediction of systemic outcomes. Combining graph autoencoders with semi-supervised learning, where the anaemia label is incorporated into the loss function, may align the embeddings more closely with the prediction task. Additionally, integrating longitudinal data or including inflammatory biomarkers could enrich the feature set and reveal temporal or biochemical correlations with graph structure(21,22).Larger, multi-centre datasets and longitudinal records could enhance generalisability and enable temporal prediction of anaemia risk. Including multi-omics or imaging features might further boost accuracy. The study is limited by a moderate sample size, class imbalance between anaemic and non-anaemic groups, and no external validation, which could affect applicability to broader populations.

CONCLUSION

This study used graph autoencoders on clinical data to model anemia. By creating a K-nearest neighbor graph of 350 patients and training a simple autoencoder, we achieved high performance (AUC ≈ 0.898) in reconstructing the adjacency matrix. However, anemia classification did not surpass 81% accuracy, and reconstruction precision was low due to class imbalance. Analysis showed the model assigns moderate probabilities to many edges and struggles to distinguish true connections. These results suggest graph autoencoders capture similarity but have limited usefulness for predicting systemic conditions like anemia, which depend on individual features. Future research should explore semi-supervised graph models with labels, test deeper or variational architectures, and consider alternative graph construction methods reflecting clinical interactions. Including biochemical markers, long-term data, or environmental exposures may enhance graph relevance. Combining graph-based learning with traditional methods could provide a comprehensive approach to understanding periodontal disease and systemic health.

DECLARATION

Conflicts of Interest

There are no conflicts of interest.

Financial support and sponsorship

Nil.

Funding

The work was not funded.

Ethical Approval

Not applicable

REFERENCES

1. Malaiappan S, P T P, Niveditha S. Green Synthesis and Characterization of Zinc Oxide Nanoparticles Using Catharanthus roseus Extract: A Novel Approach. Cureus [Internet]. 2024;16(5):e60407. http://dx.doi.org/10.7759/cureus.60407
2. Bajpai D, Malaiappan S, Shanmugam R. Evaluation of Anti-inflammatory and Antimicrobial Properties of Mustard Seed Extract-Based Hydrogel: An In Vitro Study. Cureus. 2023 Sep 13;15.
3. Swarna Meenakshi P, Sankari M, Rajeshkumar S. Formulation and evaluation of a novel herbal trio gel containing flax seed extract, carbopol and carboxymethyl cellulose. Bioinformation. 2023;19(5):540–5.
4. Chi AC, Neville BW, Krayer JW, Gonsalves WC. Oral manifestations of systemic disease. Am Fam Physician. 2010 Dec;82(11):1381–8.
5. Sari A, Ilhan G, Akcali A. Association between periodontal inflamed surface area and serum acute phase biomarkers in patients with sickle cell anemia. Arch Oral Biol. 2022 Nov;143:105543.
6. Perdoncini NN, Furquim CP, Bonfim CMS, Soares GMS, Torres-Pereira CC. Self-perception of periodontal health status among individuals with Fanconi anemia. Hematol Transfus Cell Ther. 2021;43(4):453–8.
7. Show S, Bagchi S, Dey AK, Boyapati R, Pal PC, Tejaswi KS. Effect of nonsurgical periodontal therapy and smoking status on hematological variables related to anemia of chronic disease in chronic periodontitis patient: a case-control study. Journal of Yeungnam medical science. 2022 Jul;39(3):244–9.
8. Velliyagounder K, Chavan K, Markowitz K. Iron Deficiency Anemia and Its Impact on Oral Health-A Literature Review. Dent J (Basel). 2024 Jun;12(6).
9. da Silva BB, Pontes AEF, Lemos CAA, Ortega RM. Association between sickle cell disease and periodontal disease: A systematic review. J Periodontal Res. 2023 Aug;58(4):679–86.
10. Panneerselvam S, Theyagarajan R, Sekhar V, Mani E, Krishnamurthi I, Saketharaman P. Evaluation of Systemic Markers Related to Anemia in Aggressive Periodontitis Patients before and after Phase I Periodontal Therapy: An Interventional Study. J Contemp Dent Pract. 2021 Dec;22(12):1413–6.

1 11. Han Y, Luo Z, Yue ZG, Miao LL, Xv M, Chang S, et al. The tendency of anemia of inflammation in periodontal diseases. Clin Sci (Lond). 2023 Feb;137(3):251–64.

1 12. Dong B, Sun W, Xu D, Wang G, Zhang T. DAEMDA: A Method with Dual-Channel Attention Encoding for miRNA-Disease Association Prediction. Biomolecules. 2023 Oct;13(10).

1 13. Yang Y, Sun Y, Li F, Guan B, Liu JX, Shang J. MGCNRF: Prediction of Disease-Related miRNAs Based on Multiple Graph Convolutional Networks and Random Forest. IEEE Trans Neural Netw Learn Syst. 2023 Jul;PP.

1 14. Ning Q, Zhao Y, Gao J, Chen C, et al. AMHMDA: attention aware multi-view similarity networks and hypergraph learning for miRNA-disease associations identification. Brief Bioinform. 2023;24(2).

1 15. Jia C, Wang F, Xing B, Li S, Zhao Y, Li Y, et al. DGAMDA: Predicting miRNA-disease association based on dynamic graph attention network. Int J Numer Method Biomed Eng. 2024 May;40(5):e3809.

1 16. Ai N, Liang Y, Yuan H, Ouyang D, Xie S, Liu X. GDCL-NcDA: identifying non-coding RNA-disease associations via contrastive learning between deep graph learning and deep matrix factorization. BMC Genomics. 2023 Jul;24(1):424.

1 17. Xie GB, Yu JR, Lin ZY, Gu GS, Chen RB, Xu HJ, et al. Prediction of miRNA-disease associations based on strengthened hypergraph convolutional autoencoder. Comput Biol Chem. 2024 Feb;108:107992.

1 18. Ma Z, Kuang Z, Deng L. NGCICM: A Novel Deep Learning-Based Method for Predicting circRNA-miRNA Interactions. IEEE/ACM Trans Comput Biol Bioinform. 2023;20(5):3080–92.

1 19. Xie R, Wen J, Quitadamo A, Cheng J, Shi X. A deep auto-encoder model for gene expression prediction. BMC Genomics [Internet]. 2017;18(9):845. Available from: https://doi.org/10.1186/s12864-017-4226-0

2 20. Ahmed KT, Sun J, Cheng S, Yong J, Zhang W. Multi-omics data integration by generative adversarial network. Bioinformatics. 2021 Dec;38(1):179–86.

2 21. Nußberger J, Boesel F, Lenz S, Binder H, Hess M. Synthetic observations from deep generative models and binary omics data with limited sample size. Brief Bioinform. 2021 Jul;22(4).

2 22. Liu D, Huang Y, Nie W, Zhang J, Deng L. SMALF: miRNA-disease associations prediction based on stacked autoencoder and XGBoost. BMC Bioinformatics [Internet]. 2021;22(1):219. Available from: https://doi.org/10.1186/s12859-021-04135-2