Scientific Research Report

# Variational Approaches for Drug-Disease-Gene Links in Periodontal Inflammation

*Pradeep Kumar Yadalam* [1], *Prabhu Manickam Natarajan* [2*], *Musab Hamed Saeed* [3], *Carlos M. Ardila* [4]

[1] *Department of Periodontics, Saveetha Dental College and Hospital, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai 600077, Tamil Nadu, India*

[2] *Department of Clinical Sciences, Center of Medical and Bio-allied Health Sciences and Research, College of Dentistry, Ajman University, Ajman, United Arab Emirates*

[3] *Associate Professor, Restorative Dentistry, College of Dentistry. Ajman University, Ajman, United Arab Emirates*

[4] *Carlos-M. Ardila. DDS. Periodontist. Ph.D in Epidemiology. Postdoc in Bioethics Titular Professor. Universidad de Antioquia U de A, Medellín, Colombia. Biomedical Stomatology Research Group, Universidad de Antioquia U de A, Medellín, Colombia*

ARTICLE INFO

*Article history:*
Received 13 August 2024
Received in revised form
9 September 2024
Accepted 19 September 2024
Available online 11 October 2024

*Keywords:*
periodontitis
Wnt Signaling Pathway
machine learning
quantum computing
Gene-Drug Interaction

ABSTRACT

*Introduction and Objectives:* Oral diseases, including gingivitis and periodontitis, are linked to the Wnt signaling pathway, vital for bone metabolism, cementum homeostasis, and mesenchymal stem cell differentiation. Advances in generative AI techniques, such as variational autoencoders (VAEs) and quantum variational classifiers (QVCs), offer promising tools for predicting gene associations between drugs and diseases. This study aims to compare the predictive performance of VAEs and QVCs in modeling drug-disease gene networks within the Wnt signaling pathway in periodontal inflammation.

*Methods:* Genes associated with Wnt-related periodontal inflammation were identified through comprehensive literature reviews and genomic databases. Their roles in various biological processes were evaluated using gene set enrichment analysis, employing tools like Enrichr, which integrates diverse gene sets from sources such as DSigDB, DisGeNET, and Lincs_l1000.drug. The study then applied VAEs and QVCs to predict gene-disease associations related to the Wnt signaling pathway.

*Results:* The analysis revealed an extensive network comprising 1738 nodes and 1498 edges, averaging 1.992 neighbors per node. The network exhibited a diameter of 2, a radius of 1, and a characteristic path length of 1.992, indicating limited interconnectivity. The VQA model demonstrated a high accuracy rate of 97.5%, although it only detected 50% of anomalies. The VQC model achieved a precision of 78%, with Class 1 samples showing improved recall and a balanced F1 score.

*Conclusion:* VQC and VAE models exhibit strong potential for discovering FDA-approved drugs by predicting gene-drug associations in periodontitis based on the Wnt signaling pathway.

*Clinical Relevance:* This study highlights the potential of VAEs and QVCs in predicting gene-drug associations for periodontal inflammation. This could lead to more targeted therapies for oral diseases like periodontitis, improving patient outcomes and advancing personalized treatment strategies in clinical practice.



* *Corresponding author*. Department of Clinical Sciences, Center of Medical and Bio-allied Health Sciences and Research, College of Dentistry, Ajman University, PO Box: 346, Ajman, United Arab Emirates.

E-mail addresses: pradeepkumar.sdc@saveetha.com (P.K. Yadalam), prabhuperio@gmail.com (P.M. Natarajan), m.saeed@ajman.ac.ae (M.H. Saeed), martin.ardila@udea.edu.co (C.M. Ardila).

Prabhu Manickam Natarajan: http://orcid.org/0000-0002-4780-0465

Musab Hamed Saeed: http://orcid.org/0000-0002-3564-184X

Carlos M. Ardila: http://orcid.org/0000-0002-3663-1416

https://doi.org/10.1016/j.identj.2024.09.025

## Introduction

Oral diseases, particularly inflammatory conditions such as gingivitis and periodontitis, present significant global health and economic challenges, affecting a large portion of the adult population. Periodontitis is a chronic condition characterized by a dysbiotic biofilm and a proinflammatory environment that disrupts cell signaling.[1] Wnt5a is key in periodontal ligament remodeling, while the Wnt/$\beta$-catenin pathway is crucial for alveolar bone formation.[2]

The Wnt signaling pathway consists of two primary branches: the canonical Wnt/$\beta$-catenin pathway and the noncanonical Wnt/$\beta$-catenin-independent pathway. In the canonical pathway, Wnt ligands (such as Wnt1 and Wnt3a) bind to Frizzled receptors, stabilizing $\beta$-catenin to promote gene expression.[3] This pathway is essential for processes like bone formation and cementum homeostasis. The noncanonical pathway, on the other hand, does not involve $\beta$-catenin and instead regulates processes like cell migration, behavior, and osteogenic differentiation.[4] Key modulators, such as R-spondins and cytokines, influence these pathways, adjusting cellular responses in periodontal ligament cells.[5]

Gingivitis and periodontitis are inflammatory diseases initiated by bacterial biofilms on dental surfaces. Gingivitis represents the initial immune response characterized by inflammation without tissue loss. If left untreated, it can progress to periodontitis, marked by irreversible tissue destruction and chronic inflammation. The progression of periodontal disease is influenced by local and systemic risk factors, with periodontopathic bacterial infections impacting the Wnt signaling pathway, which is essential for tissue maintenance.[6,7] Dysregulation of the Wnt signaling pathway is thought to exacerbate the inflammatory environment, disrupt tissue homeostasis, and impair bone regeneration, contributing directly to the progression of periodontal diseases. Studies have shown elevated Wnt levels in gingival tissues during stage 3 periodontitis, suggesting that inflammation may disrupt the Wnt/$\beta$-catenin signaling pathway, potentially contributing to the pathogenesis of periodontal disease.[1]

Wnt signaling is a fundamental pathway in embryonic development, tissue homeostasis, and stem cell regulation. It has been implicated in various diseases, including cancer, neurological disorders, and metabolic conditions. A previous study proposed using a network-based computational framework to identify network motifs in an integrated disease-drug-gene network from Semantic MEDLINE.[8] Understanding the Wnt signaling-associated gene network can provide insights into disease mechanisms, therapeutic targets, personalized medicine, and biomarkers for disease prediction and prognosis. By examining the specific molecular alterations within the Wnt signaling pathway, researchers can develop targeted treatment strategies, identify individuals at risk for systemic diseases, and assess the severity and progression of existing conditions, potentially leading to personalized drug targets and repurposed treatments.[6,7]

Variational autoencoders (VAEs)[9] are generative AI techniques that predict drug-disease gene associations within the Wnt pathway. VAEs encode drugs, diseases, and genes into a lower-dimensional latent space, extracting relevant features to predict their associations. OntoVAE, a novel VAE, incorporates ontologies to allow direct interpretability and predictive modeling for genetic or drug-induced perturbations. VAEs were chosen for their ability to handle high-dimensional data and capture latent structures in complex biological systems, making them suitable for identifying potential associations that may not be apparent through traditional methods. VAEs can generate potential associations by sampling from the learned latent space, identifying new associations, and guiding further research.[10,11]

Using quantum computing, the quantum variational classifier (QVC)[12] is another promising tool for predicting drug-gene-disease associations within the Wnt pathway. Quantum feature mapping transforms input data into a high-dimensional space, enabling the exploration of complex relationships between drugs, genes, and diseases and predicting associations with probabilities. QVCs were selected due to their unique capability to manage large and intricate datasets by leveraging quantum feature mapping, which allows for the detection of subtle and complex patterns in biological data, offering computational advantages over classical machine learning models. Quantum variational classifiers offer computational advantages over classical machine learning algorithms, with larger quantum computers and advanced quantum algorithms enhancing their scalability and potential.[13]

Variational Quantum Circuits (VQCs) encode data into quantum states using a feature map, process it with a quantum circuit, optimize parameters using machine learning algorithms, classify new data points, and predict the highest probability class.VQCs[14] offers advantages over traditional machine learning algorithms, including improved performance in tasks such as image classification and natural language processing and the ability to learn complex relationships through quantum entanglement. However, challenges remain, including the need for access to advanced quantum computers and the computational expense of training on these systems.

VAEs and QVCs are generative models that capture complex data distributions and reveal latent structures in high-dimensional datasets. Their potential lies in predicting drug-disease-gene networks, particularly within the Wnt signaling pathway in periodontal inflammation. These models were chosen for their capacity to efficiently process high-dimensional data and uncover hidden relationships in biological datasets, which are essential for making accurate predictions in drug-disease-gene networks.

The efficacy of these technologies in biomedical applications is underexplored, and the unique capabilities of quantum computing in handling complex biological data are not yet fully realized. This study compares the predictive performance of VAEs and QVCs in modeling drug-disease-gene networks within the Wnt signaling pathway. This is the first study to predict drug-disease gene associations in the Wnt pathway related to periodontal inflammation.

## Materials and methods

### *Dataset retrieval and preparation*

Genes associated with Wnt-related periodontal inflammation were collected from literature sources and genomic

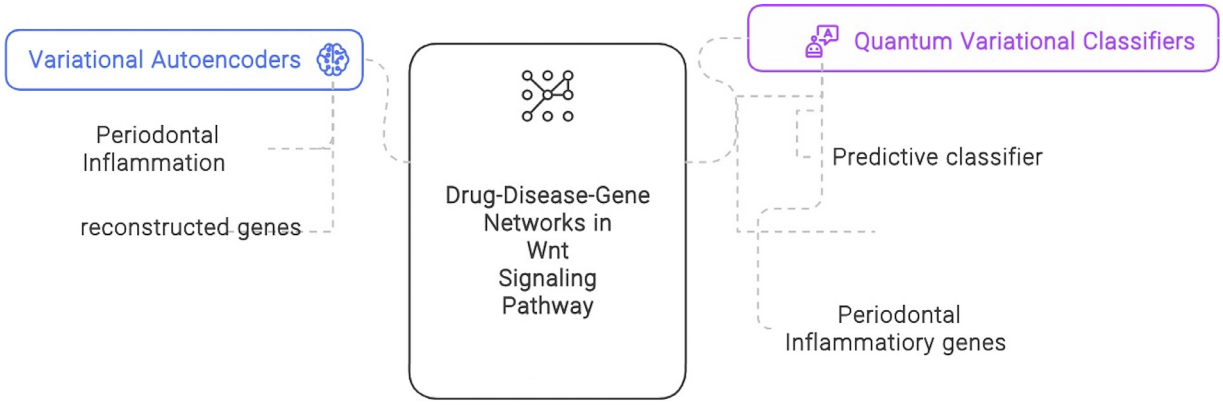


**Fig. 1 – Overall model architecture.**

databases.[1,2] Their involvement in biological processes and their interactions with drugs and other genes was identified using the Enrichr tool.[2]

Based on the search results, key Wnt-related genes implicated in periodontal inflammation were identified, including WNT3, WNT3A, WNT5A, WNT8A, WNT9B, WNT11, DKK1, SOST, LRP5, AXIN2, LEF1, TCF7, CTNNB1, GSK3B, and FZD.

### *Enrichr*

Enrichr is a user-friendly, web-based tool designed for gene set enrichment analysis and functional annotation.[15] It integrates gene sets from various sources, such as Gene Ontology and KEGG Pathways, allowing researchers to analyze and interpret gene lists. Enrichr was utilized to analyze the retrieved genes, focusing on gene ontology, DSigDB, DisGeNET, and Lincs_l1000 datasets.

### *DSigDB (drug signatures database)*

DSigDB is a comprehensive resource within Enrichr that provides gene expression signatures linked to drug treatments or exposures.[16] From enrichr, disgdb was used to identify drugs targeting wnt-associated genes.

### *DisGeNET*

DisGeNET is another crucial resource integrated into Enrichr, offering a comprehensive database of gene-disease associations.[17] DisGeNET from enrichr, all associated diseases were identified in wnt genes.

### *Lincs_l1000*

The Lincs_l1000 dataset in Enrichr provides gene expression signatures from the Library of Integrated Network-based Cellular Signatures (Lincs) project, focusing on small molecules and drugs.[18] Lincs_l1000 signatures were identified for periodontal wnt genes, gaining insights into drug-gene interactions, facilitating drug discovery, and repurposing efforts.

All datasets from the Drug Signatures Database, gene-disease associations, and gene expression signatures from chemical perturbations were preprocessed to remove outliers, missing data, and duplicate values. The cleaned data were merged, and genes were set as targets. The merged dataset contains gene IDs, associated diseases, drugs, p-values, odds ratios, and gene names with 1499 entries (Figure 1).

### *Cytoscape*

Cytoscape is an open-source software platform used for visualizing and analyzing biological networks.[19] Cytoscape was employed to investigate the interactome networks of drugs, diseases, and chemical perturbations in this study. Merged data were imported to Cytoscape, and drug-gene-disease networks were constructed and studied for network properties.

### *Variational autoencoder architecture*

The Variational Autoencoder (VAE)[20] is a generative model that utilizes an encoder-decoder structure to learn a latent input data representation. The VAE, as a generative model, encodes input data into a lower-dimensional latent space and decodes it back to the original space. The encoder maps the input data into a latent space, and the latent variables are sampled from the output parameters, allowing the VAE to generate new samples. The decoder then reconstructs the original input data with the intention of minimizing the reconstruction error.

During training, the VAE was applied to the provided dataset, and the process was monitored through loss curves. The training and validation losses decreased rapidly in the initial epochs, stabilizing after approximately 20-30, indicating successful model convergence. The final training loss was very low, suggesting that the VAE effectively learned to reconstruct the input data. Although the validation loss was slightly higher, it remained low, indicating good generalization to unseen data.

The latent space was visualized in a 2D representation, with each point corresponding to a data sample. The colors represented the values of the first feature in the original data space, and the clustering of similar colors suggested that the VAE captured meaningful data representations in the latent space, effectively encoding key features and relationships.

The encoder maps input data to a latent space, starting with an input layer with a dimension determined by the input dataset's shape. The first layer is dense with seven units, using the ReLU activation function to reduce data dimensionality. The output is fed into two dense layers, each with 2 units, representing the mean and log variance of the latent distribution. The sampling layer is a crucial part of the VAE, generating new data points in the latent space. It uses a custom sampling function that generates random noise from a standard normal distribution and combines it with mean and log variance to create a latent vector.

The decoder structure involves a dense layer with 7 units and ReLU activation, mirroring the encoder's structure. The output is then passed to a dense layer with input_dim units and a sigmoid activation function, producing a continuous approximation of the original data. The VAE model connects encoder and decoder components, encoding and decoding input data. The loss function consists of reconstruction loss and Kullback-Leibler divergence loss. The reconstruction loss measures the decoded output's accuracy, while the KL divergence loss ensures the learned distribution is close to a standard normal distribution. The total loss is the sum of these two terms, averaged over the batch. The VAE model is trained using the Adam optimizer, a popular choice for deep learning due to its adaptive learning rate capabilities. A custom loss function is used during training. The VAE was trained for 100 epochs on the dataset.

Input_dim: The parameter 'dimensionality' refers to the dimensionality of the input data, influenced by the dataset's shape and the number of features or columns, determining the input layer size. latent_dim: This parameter specifies the dimensionality of the latent space, which is the lower-dimensional space where the input data is encoded. The 2-dimensional latent space facilitates easy visualization and interpretation of the learned representations.

Intermediate_dim: The parameter defines the number of units in the encoder and decoder's intermediate layer, acting as a bottleneck to reduce input data dimensionality before mapping to latent space.

The intermediate layer has 7 units, with 100 epochs representing a complete pass through the dataset. The batch-size parameter determines the number of samples processed before updating the model's internal parameters, with a 32-sample batch size indicating weight updates after processing 32 samples.

The VAE class encapsulates the encoder, decoder, and other necessary components. The model is then compiled using the compile method, preparing it for training by specifying the optimizer and loss function. The Adam optimizer is utilized for training deep learning models due to its adaptive learning rate capabilities. The loss function, binary cross-entropy, is used for reconstruction tasks with binary input data to improve the model's reconstruction accuracy during training.

A grid search was performed using various hyperparameter combinations, and the model was trained using the 'train_vae' function. After training, anomalies were detected on the test set using a 'detect_anomalies' function, which provided a list of anomaly labels and mean square error (MSE) scores. The average precision score was computed using the 'sklearn.metrics.average_precision_score' function. The optimal model was selected based on the highest average precision score, saved, and visualized with a scatter plot where each point represented the MSE score.

### *Quantum variational classifier architecture*

The Variational Quantum Classifier (VQC)[21] is a machine learning algorithm that leverages quantum computing principles for classification tasks. The specific architecture and hyperparameters of a VQC can vary depending on the implementation. VQCs use an input encoding process to transform data into quantum states, which are then processed by a parameterized quantum circuit. Measurements are performed on the quantum states to obtain classical data, which is used to make predictions or update the circuit's parameters. Key hyperparameters include the number of layers, qubits, learning rate, and regularization techniques tailored to the problem at hand. Its architecture includes several key components: the feature map, which encodes classical data into a quantum state; the ansatz, a parameterized quantum circuit that applies adjustable quantum gates; and the COBYLA optimizer, a classical algorithm that fine-tunes the ansatz parameters to minimize the cost function. Additionally, the sampler measures the final quantum state, converting it into classical information for classification.

In this study, gene data was loaded, and the features were scaled to ensure a mean of 0 and a standard deviation of 1. The target variable, 'Combined Score,' was converted into binary classes based on its median value. The dataset was split into training and testing sets using an 80-20 split. A VQC was then constructed using Qiskit's machine learning library, which included a feature map and an ansatz.

The COBYLA optimizer was employed to fine-tune the parameters of the quantum circuit, making it particularly effective for quantum machine learning applications. The Variational Quantum Classifier (VQC) was trained on the training dataset, with iterative adjustments to minimize classification error. After training, the VQC was tested on the validation dataset, where its performance was measured using metrics such as accuracy, precision, recall, and F1-score.

## Results

The intricate relationships between drugs, genes, and diseases associated with the Wnt signaling pathway are visualized in Figure 2, which reveals potential therapeutic targets and biomarkers.

### *Network analysis results*

The network analysis identified a large network comprising 1738 nodes and 1498 edges. The average number of neighbors is 1.992, reflecting relatively strong node connectivity. The network has a diameter of 2, a radius of 1, and a characteristic path length of 1.992, suggesting short paths between most node pairs. However, the clustering coefficient is 0.000, indicating a lack of significant interconnections within the

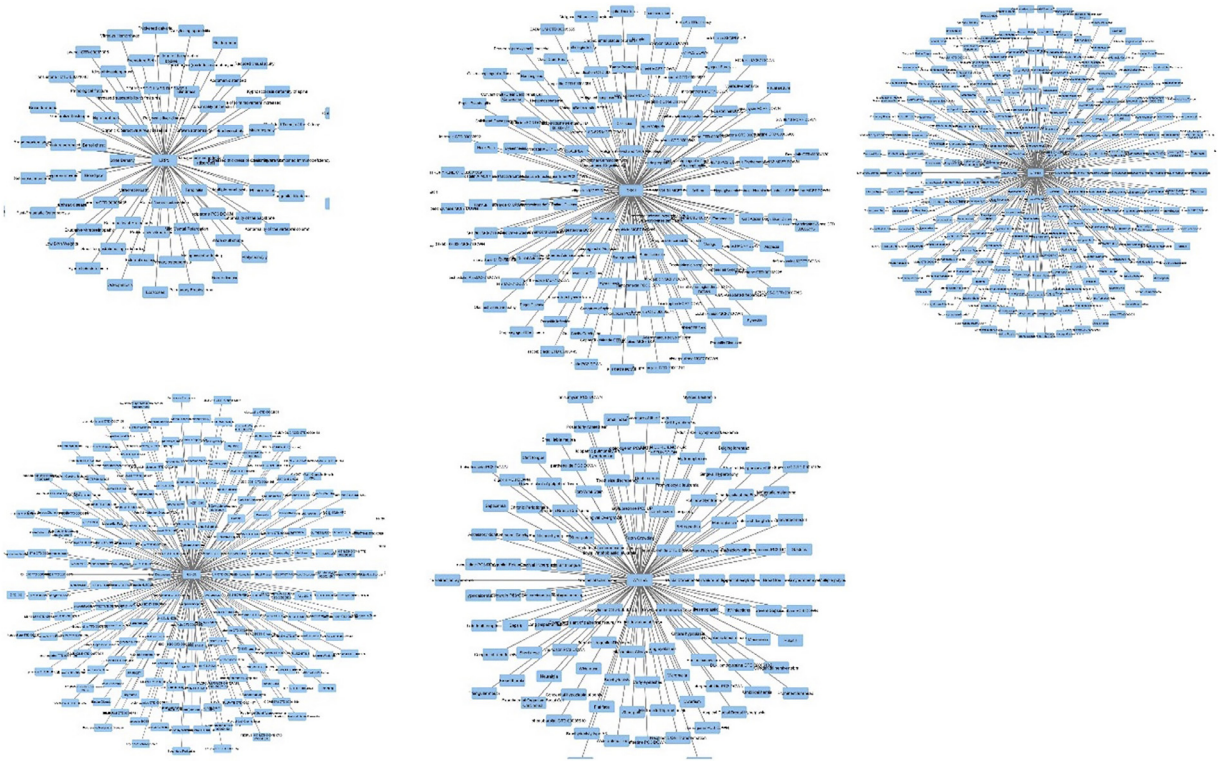

**Fig. 2 – Network of drug-gene-disease interactions in the Wnt pathway.**

network. The network density, calculated at 0.008, suggests the network is sparse.

Additionally, the network heterogeneity is 7.778, indicating a high variation in the degree of neighboring nodes. The network centralization is 1.000, highlighting the presence of a highly central node playing a crucial role. The presence of 244 connected components indicates the existence of multiple subnetworks within the larger network.

### *Variational autoencoder (VAE) training results*

The VAE model was trained over 200 epochs, with a decreasing average loss indicating successful convergence. The model correctly classified 1900 samples as normal, flagged 50 as anomalous, and failed to detect 50 anomalous samples with no false positives at the specified threshold. The VAE demonstrated a high accuracy rate of 97.5% at a certain threshold, though it only identified half of the anomalies. The model exhibited 100% precision for anomalies, accurately flagging samples as anomalous. However, a trade-off remains between detecting all anomalies and maintaining high precision. These results stem from a simulated binary classification task, with reconstruction error as the decision metric and the threshold selection influenced by application requirements.

### *Additional metrics*

The following additional metrics were calculated to assess the VAE model's performance:

1. Mean Squared Error (MSE): 1.3932

The MSE, which represents the average squared difference between predicted and actual values, reflects the model's effectiveness in accurately reconstructing input data.

2. Mean Absolute Error (MAE): 0.9183

The MAE measures the average absolute difference between predictions and actual values, indicating the model's improved reconstruction accuracy, similar to MSE.

3. Average KL Divergence Loss: 0.0026

This metric assesses the alignment of the learned latent space with a predefined prior distribution. A lower KL Divergence Loss value indicates a better approximation.

The study concluded that lower MSE and MAE values signify enhanced data reconstruction. In contrast, a lower KL Divergence Loss reflects a closer match between the learned latent space and its prior distribution. These metrics underscore the VAE model's capability to compress and reconstruct data accurately.

Figure 3 presents a comprehensive analysis of our model's performance, showcasing the latent space visualization and training dynamics, which collectively demonstrate the effectiveness of our approach.

These visualizations provide a detailed perspective on the model's performance within a simulated classification context, showcasing its strengths and areas needing improvement. They offer valuable insights that can guide further refinement and practical applications.

### *Reconstruction error analysis*

The reconstruction error analysis identified the top 10 entities —drugs, diseases, genes, and related terms—with the highest

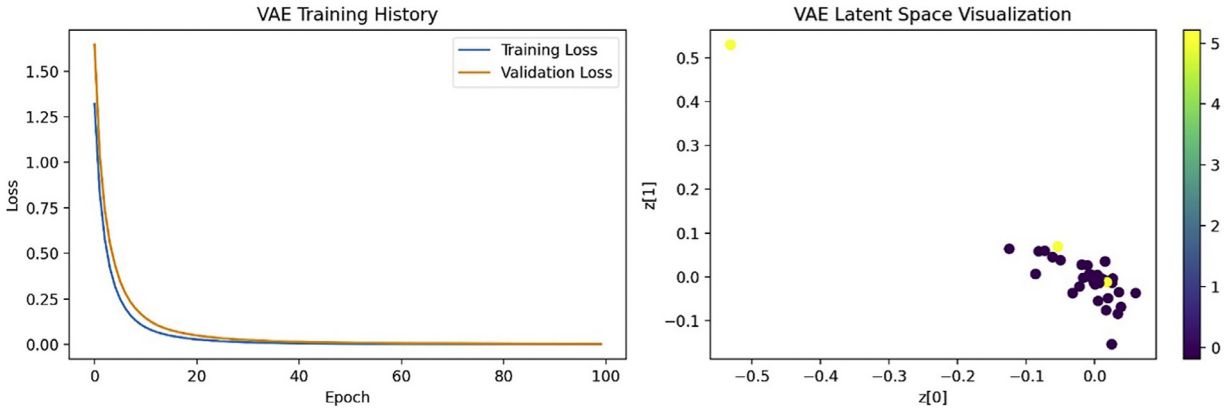


**Fig. 3 – Analysis of latent space and training dynamics. Left plot: This scatter plot visualizes the latent space, with each point representing a data sample projected into a two-dimensional latent space. The colors correspond to the values of the first feature in the original data space. Right plot: This plot depicts the training and validation loss across epochs, illustrating the model's learning progress.**

reconstruction errors, including WNT5A, LEF1, GSK3B, kaempferol, SOST, GSK3B (again), transient cerebral ischemia, CTNNB1, deferoxamine, and Zoledronic acid. These terms exhibit a reconstruction error of 0.9979, suggesting potential issues in the reconstruction process. Specifically, DKK1 (Metabolic Bone Disorder) had a reconstruction error of 0.9930, while GSK3B, TCF7, and CTNNB1 (AH 23848 CTD 00002080) had a reconstruction error of 0.9912.

The dataset's top 10 list includes seven unique genes based on their reconstruction errors, indicating their significant influence due to their unique or complex patterns. GSK3B appears multiple times, linked to various biological processes. SOST and DKK1 are associated with bone metabolism, highlighting their importance. WNT5A, LEF1, and CTNNB1 are part of the Wnt signaling pathway, which plays a crucial role in development, cell proliferation, and cancer.

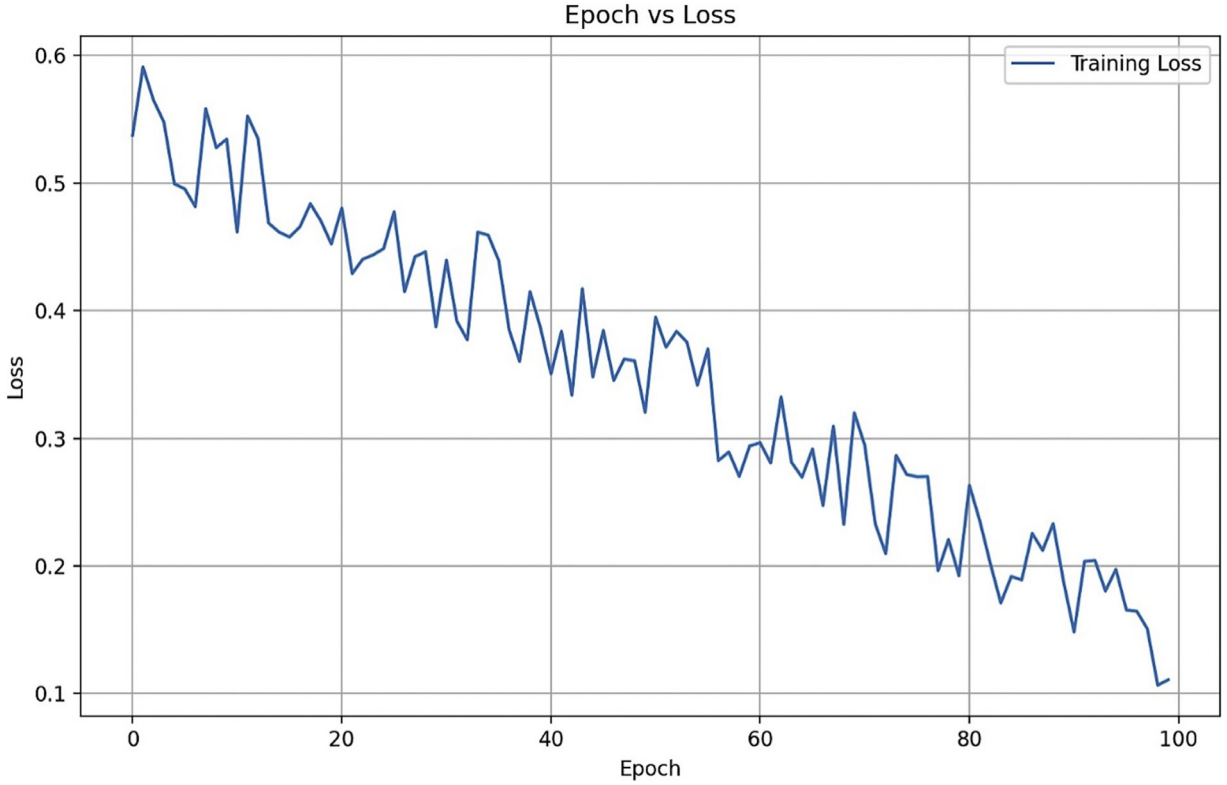


**Fig. 4 – Epoch loss curve. This plot illustrates the simulated training loss over 100 epochs. The loss decreases progressively throughout the training process, indicating that the model is learning and enhancing its performance over time.**

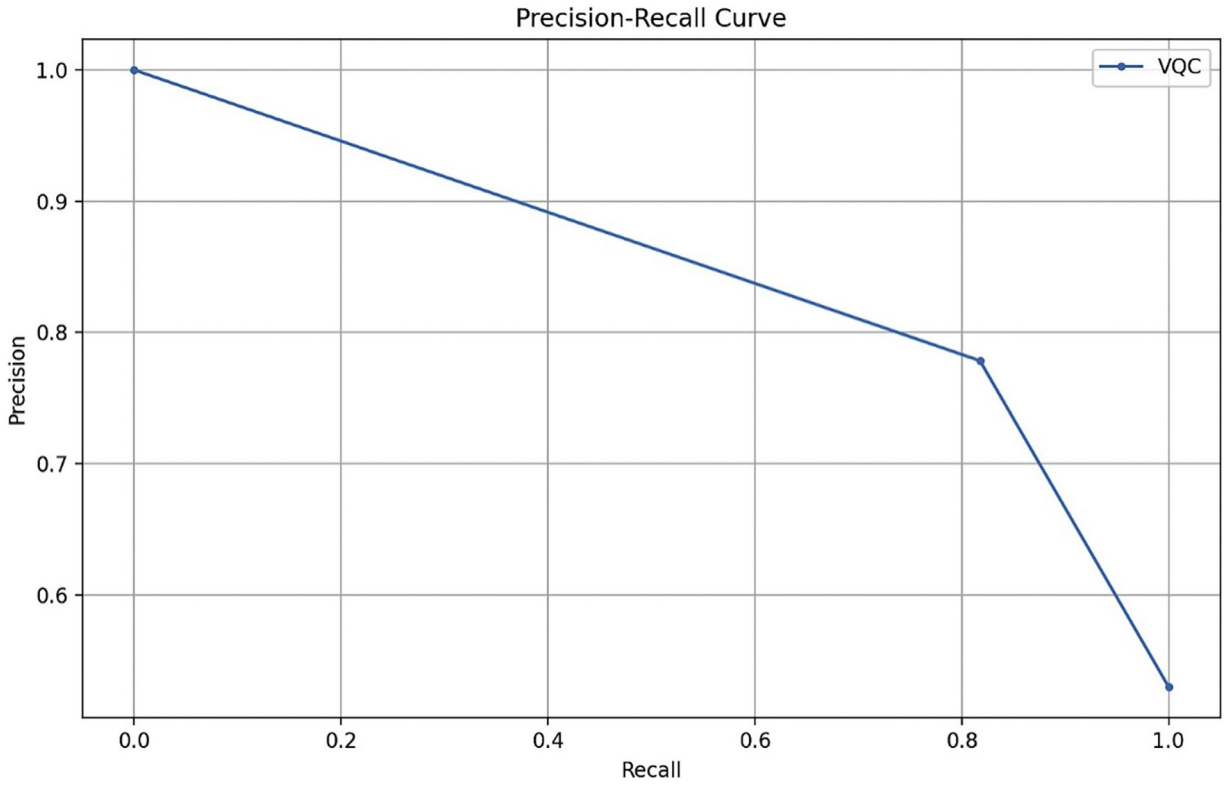


**Fig. 5 – Precision-recall curve. The precision-recall curve depicts the trade-off between precision and recall, with larger areas under the curve signifying better model performance. The curve varies with changes in the decision threshold.**

Some terms relate to specific compounds or drugs, pointing to potential gene-drug interactions.

### *Variational quantum classifier (VQC) performance*

The VQC model achieved an accuracy of 78% on the test set, correctly classifying 78% of the samples. The model performed slightly better on Class 1 samples compared to Class 0. Precision was 0.78 for both classes, with Class 1 showing higher recall. The F1-score, which balances precision and recall, was 0.76 for Class 0 and 0.80 for Class 1. The loss curve over 100 epochs consistently decreased, indicating improvement in learning and performance. The precision-recall curve illustrated a trade-off between precision and recall, with a higher area under the curve indicating better overall performance.

The VQC model demonstrated promising results on the gene data, maintaining a good balance between precision and recall. However, there is potential for improvement, particularly in reducing false negatives for Class 0 and false positives for Class 1. Future work could explore different quantum circuit architectures, feature maps, and optimizers to enhance performance. Additionally, addressing any class imbalance in the training data and implementing feature selection strategies could improve classification outcomes.

The model's training dynamics are illustrated in Figure 4, which demonstrates a consistent decrease in loss over 100 epochs, indicating effective learning and performance improvement.

Figure 5 presents the Precision-Recall Curve, illustrating the model's performance across different decision thresholds, where a larger area under the curve indicates superior accuracy.

Figure 6 presents the Confusion Matrix, providing a detailed breakdown of the model's classification performance across all classes, highlighting accuracy, precision, and recall for each category.

The confusion matrix reveals that our model performs slightly better on Class 1 (high Combined Score) than Class 0 (low Combined Score), correctly identifying 130 out of 159 Class 1 samples compared to 104 out of 141 Class 0 samples. Additionally, the matrix indicates 104 true negatives and 29 false positives.

These visualizations underscore that while our model performs reasonably well overall, there is potential for improvement, particularly in reducing false negatives for Class 0 and positives for Class 1. The precision-recall curve illustrates the relationship between precision and recall across various threshold levels for making predictions. Precision measures the accuracy of positive predictions, whereas recall assesses the model's ability to identify all relevant instances. A larger area under the curve indicates a more effective model, and the curve demonstrates how alterations in the threshold influence both precision and recall.

## Discussion

Research indicates that classical and nonclassical Wnt signaling pathways are essential for bone development and

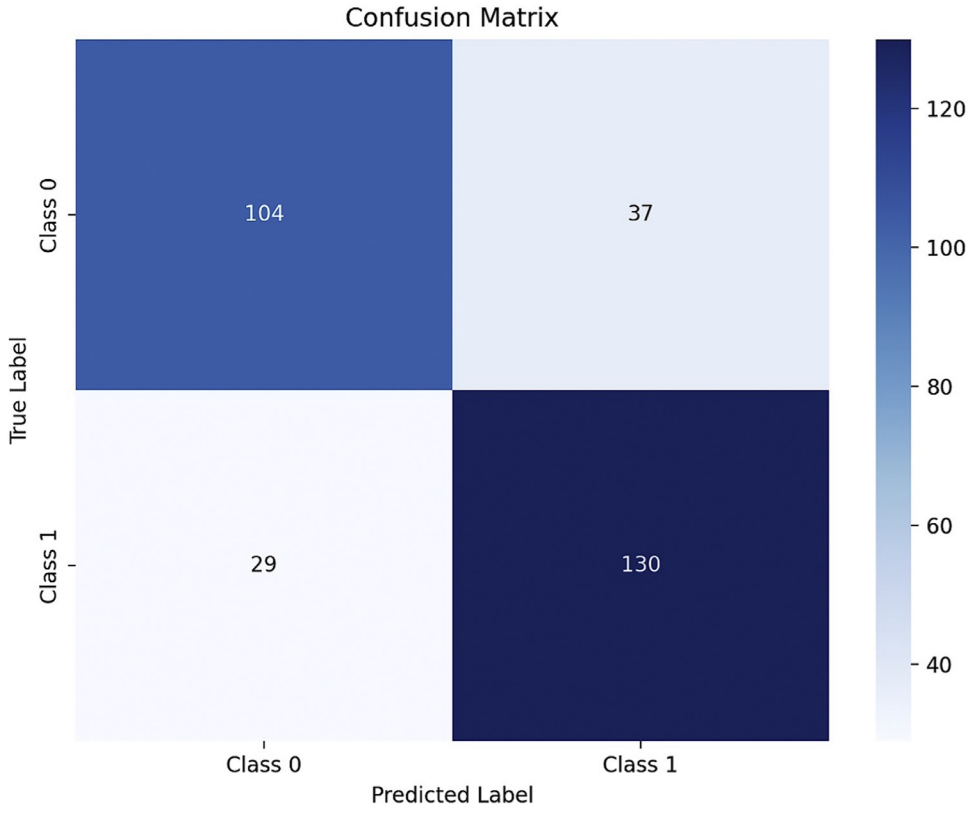


**Fig. 6 – Confusion matrix for all the classes.**

regeneration. Inflammatory environments can disrupt this balance, and targeting $\beta$-catenin in the classical Wnt pathway may enhance osteogenic differentiation in periodontal ligament stem cells (PDLSCs), potentially improving bone repair and regeneration. Although the mechanisms and interactions involved in activating the Wnt signaling pathway are not fully understood, gaining insight into these processes could help identify potential therapeutic targets for treating periodontitis.[1] The top reconstructed genes, diseases, and drugs identified through variational autoencoders (Figures 1 and 2) include WNT5A, GSK3B, and LEF1, which are crucial molecules in periodontal inflammation and influence osteogenic differentiation, tooth development, and root formation, thereby affecting tissue regeneration.[3]

Wnt5a mRNA expression levels are upregulated in gingival tissues of patients with periodontitis and positively correlated with clinical parameters of periodontitis.[1,2] Wnt5a is a key molecule in the innate immune response to periodontal pathogen invasion, expressed in gingival epithelial cells and monocyte/macrophage-lineage cells in response to *Porphyromonas gingivalis*-derived lipopolysaccharide.[1] PDL cells are sensitive to Wnt/$\beta$-catenin pathway activation, potentially increasing PDL mineralization. Wnt5a stimulation suppresses osteogenic markers and increases POSTN, FBN1, and COL1A1, suggesting PDL remodeling. These periodontal-Wnt constructed genes from this generative AI model play an important role in inflammation and regeneration.[1,2]

Kaempferol, SOST, GSK3B, and CTNNB1 are flavonoids and proteins that may modulate the Wnt signaling pathway, potentially inhibiting cancer cell proliferation. SOST inhibits the Wnt pathway by binding to LRP5/6 receptors, which impacts bone formation. GSK3B phosphorylates $\beta$-catenin,[3,22] thereby inhibiting the pathway. Transient cerebral ischemia can also influence the Wnt signaling pathway, potentially producing neuroprotective effects. CTNNB1 stabilizes and accumulates in the cytoplasm, activating Wnt target genes. Deferoxamine, an iron chelator, may provide neuroprotection by modulating oxidative stress and Wnt/$\beta$-catenin signaling. Zoledronic acid, a bisphosphonate, impacts bone metabolism and may interact with the Wnt signaling pathway. Understanding these associations among drugs, genes, and diseases could facilitate the identification of novel drug-repurposing strategies for periodontal inflammation and systemic disease connections.[5,6,23-26]

The study demonstrates the efficacy of the Variational Autoencoder (VAE) model, which achieved a high accuracy rate of 97.5% at a specific threshold. Figure 3 presents two plots of a Variational Autoencoder model, showing a significant decrease in training and validation losses over 100 epochs and stabilizing near zero after 20 epochs, indicating effective training and good generalization (Figure 3). However, it only detects half of the anomalies at this threshold. The Variational Quantum Classifier (VQC) model achieved an accuracy of 78%, accurately classifying 78% of samples. The model performed slightly better on Class 1 samples than on Class 0 samples, exhibiting higher recall for Class 1 (Figures 5 and 6). The precision-recall curve for VQC starts with high precision and low recall, and as recall increases, precision gradually decreases, indicating a trade-off between coverage and accuracy. The F1-score balanced precision and recall,

yielding values of 0.76 for Class 0 and 0.80 for Class 1, which is consistent with previous studies, such as those evaluating Tybalt, a VAE approach,[14,21] on simulated single-cell RNA sequencing (scRNA-seq) data, where challenges with larger datasets were noted.

The VQC and VAE models for drug-gene-disease association prediction in the Wnt pathway demonstrate potential for further improvement in periodontal inflammation. Strategies to address class imbalance, such as oversampling or undersampling, should be considered. VAE may miss anomalies, leading to false negatives, while VQC trades coverage and accuracy, indicating the need for further improvements. Additionally, exploring different quantum circuit architectures, optimizing feature maps and hyperparameters, and reducing false negatives and positives could enhance model performance

While our study demonstrates the potential of VAE and VQC models in predicting drug-disease-gene associations, we acknowledge that these models are not without limitations. Specifically, the VAE model struggles to detect anomalies in the data, which may lead to biased results. Furthermore, the class imbalance issue in our dataset poses a challenge for the VQC model, potentially affecting its performance. These limitations may be attributed to the complexity of the data and the choice of hyperparameters. Future studies could explore techniques such as data preprocessing, regularization methods, and ensemble learning to address these limitations and improve model performance. By acknowledging and addressing these challenges, we can further enhance the accuracy and reliability of our predictions and ultimately contribute to the development of more effective therapeutic strategies. Moreover, addressing the limitations of simulated data, considering additional evaluation metrics, performing feature selection, experimenting with ensemble methods, ensuring interpretability, and leveraging large-scale quantum computing are also crucial avenues for future research. These efforts enhance the models' performance and provide valuable insights for drug discovery and precision medicine, particularly in Wnt pathway-associated periodontal inflammation.

Future research could build upon our study by addressing the class imbalance issue through techniques such as oversampling, undersampling, or generating synthetic samples. Additionally, exploring different quantum circuit architectures, like Variational Quantum Eigensolver (VQE) or Quantum Approximate Optimization Algorithm (QAOA), could provide insights into their effectiveness for drug-disease-gene association prediction. Furthermore, leveraging large-scale quantum computing resources could enable the simulation of more complex biological systems, allowing for a deeper understanding of the underlying mechanisms. Alternative evaluation metrics, such as F1-score or Area Under the Receiver Operating Characteristic Curve (AUC-ROC), could also be considered to provide a more comprehensive assessment of model performance. By pursuing these avenues, future studies can further advance the application of quantum machine learning in biomedical research and potentially uncover new therapeutic targets.

To validate these findings, we propose a multistep approach. First, to collaborate with clinicians to identify relevant real-world datasets and ensure that these models are tailored to address actual clinical needs. Next, to test these models using these datasets, evaluating their performance and identifying potential biases. Based on these results, it is possible to refine these models, exploring techniques such as transfer learning, domain adaptation, and ensemble methods to improve their accuracy and reliability. Finally, it is necessary to work with clinicians to integrate these models into existing workflows, ensuring that they can be used to inform actual clinical decision-making.

By tailoring treatment strategies to an individual's unique needs, clinicians can improve patient outcomes and reduce the risk of adverse reactions. Furthermore, these models could be integrated into existing electronic health record systems, enabling clinicians to access personalized treatment recommendations at the point of care.

## Conclusion

The Variational Quantum Classifier (VQC) and Variational Autoencoder (VAE) models are promising in predicting drug-gene-disease associations. The VQC model demonstrates superior accuracy in classifying drug-gene pairs, while the VAE model effectively detects anomalies. These models can potentially transform FDA-approved drug discovery by pinpointing potential therapeutic targets for periodontal inflammation. Nevertheless, several areas for improvement remain, including addressing class imbalance, optimizing quantum circuit architectures, enhancing feature maps, validating findings with real-world data, and ensuring interpretability for large-scale applications.

## Declaration of competing interest

he authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability statement

The datasets used and/or analysed during the current study are available from the corresponding author upon reasonable request.

## Funding

None.

## Ethical approval

Not required.

## Authorship contribution statement

Prabhu Manickam Natarajan, Musab Hamed Saeed, Pradeep Kumar Yadalam, and Carlos M. Ardila contributed to the

conception, analysis, interpretation of data, and drafting of the manuscript.

Prabhu Manickam Natarajan, Musab Hamed Saeed, Pradeep Kumar Yadalam, and Carlos M. Ardila: Conceptualization

Prabhu Manickam Natarajan, Musab Hamed Saeed, Pradeep Kumar Yadalam, and Carlos M. Ardila: Methodology

Prabhu Manickam Natarajan, Musab Hamed Saeed, Pradeep Kumar Yadalam, and Carlos M. Ardila: Data curation

Pradeep Kumar Yadalam, and Carlos M. Ardila: Writing-Original draft preparation.

Prabhu Manickam Natarajan, Musab Hamed Saeed, Pradeep Kumar Yadalam, and Carlos M. Ardila: Visualization

Pradeep Kumar Yadalam, and Carlos M. Ardila: Investigation.

Prabhu Manickam Natarajan, Musab Hamed Saeed, Pradeep Kumar Yadalam, and Carlos M. Ardila Validation

Pradeep Kumar Yadalam, and Carlos M. Ardila: Writing-Reviewing and Editing.

## Acknowledgements

We would like to thank the Center of Medical and Bioallied Health Sciences and Research, Ajman University, Ajman, UAE

## REFERENCES


1. González-Quintanilla D, Abásolo N, Astudillo P. Wnt signaling in periodontal disease. Front Dent Med 2021;2:763308.
2. Güney Z, Kurgan Ş, Önder C, Tayman MA, Günhan Ö, Kantarci A, et al. Wnt signaling in periodontitis. Clin Oral Investig 2023;27:6801–12.
3. Baron R, Kneissel M. WNT signaling in bone homeostasis and disease: from human mutations to treatments. Nat Med 2013;19:179–92.
4. Huang HC, Klein PS. The Frizzled family: receptors for multiple signal transduction pathways. Genome Biol 2004;5:234.
5. MacDonald BT, He X. Frizzled and LRP5/6 receptors for Wnt/$\beta$-catenin signaling. Cold Spring Harb Perspect Biol 2012;4(12):a007880.
6. Ahn VE, Chu MLH, Choi HJ, Tran D, Abo A, Weis WI. Structural basis of Wnt signaling inhibition by Dickkopf binding to LRP5/6. Dev Cell 2011;21:862–73.
7. Boudin E, Fijalkowski I, Piters E, Van Hul W. The role of extracellular modulators of canonical Wnt signaling in bone metabolism and diseases. Semin Arthritis Rheum 2013;43:220–40.
8. Chow YL, Singh S, Carpenter AE, Way GP. Predicting drug polypharmacology from cell morphology readouts using variational autoencoder latent space arithmetic. PLoS Comput Biol 2022;18:e1009888.
9. Zhang Y, Hu Y, Li H, Liu X. Drug-protein interaction prediction via variational autoencoders and attention mechanisms. Front Genet 2022;13:1032779.
10. Jia P, Hu R, Pei G, Dai Y, Wang YY, Zhao Z. Deep generative neural network for accurate drug response imputation. Nat Commun 2021;12:1740.
11. Doncevic D, Herrmann C. Biologically informed variational autoencoders allow predictive modeling of genetic and drug-induced perturbations. Bioinformatics 2023;39:btad387.
12. de Oliveira NM, Park DK, Araujo IF, da Silva AJ. Quantum variational distance-based centroid classifier. Neurocomputing 2024;576:127356.
13. Zeguendry A, Jarir Z, Quafafou M. Quantum machine learning: a review and case studies. Entropy (Basel) 2023;25(2):287.
14. Park S, Park DK, Rhee JKK. Variational quantum approximate support vector machine with inference transfer. Sci Rep 2023;13:3288.
15. Kuleshov M V, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 2016;44(1):W90-7.
16. Yoo M, Shin J, Kim J, Ryall KA, Lee K, Lee S, et al. DSigDB: drug signatures database for gene set analysis. Bioinformatics 2015;31:3069–71.
17. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, Ronzano F, Centeno E, Sanz F, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res 2020;48(D1):D845–55.
18. Han HW, Hahn S, Jeong HY, Jee JH, Nam MO, Kim HK, et al. LINCS L1000 dataset-based repositioning of CGP-60474 as a highly potent anti-endotoxemic agent. Sci Rep 2018;8:14969.
19. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498–504.
20. Pravalphruekul N, Piriyajitakonkij M, Phunchongharn P, Piyayotai S. De novo design of molecules with multiaction potential from differential gene expression using variational autoencoder. J Chem Inf Model 2023;63:3999–4011.
21. Bermejo P, Orús R. Variational quantum and quantum-inspired clustering. Sci Rep 2023;13:13284.
22. Goldring SR, Goldring MB. Eating bone or adding it: the Wnt pathway decides. Nat Med 2007;13:133–4.
23. Ke J, Harikumar KG, Erice C, Chen C, Gu X, Wang L, et al. Structure and function of Norrin in assembly and activation of a Frizzled 4-Lrp5/6 complex. Genes Dev 2013;27(21):2305–19.
24. Hua Y, Yang Y, Li Q, He X, Zhu W, Wang J, et al. Oligomerization of Frizzled and LRP5/6 protein initiates intracellular signaling for the canonical WNT/$\beta$-catenin pathway. J Biol Chem 2018;293(51):19710–24.
25. Wo D, Peng J, Ren DN, Qiu L, Chen J, Zhu Y, et al. Opposing roles of Wnt inhibitors IGFBP-4 and Dkk1 in cardiac ischemia by differential targeting of LRP5/6 and $\beta$-catenin. Circulation 2016;134(24):1991–2007.
26. Bao J, Zheng JJ, Wu D. The structural basis of DKK-mediated inhibition of Wnt/LRP signaling. Sci Signal 2012;5(224):pe22.