# Role of Residual Graph Attention Networks in Predicting Gene-Drug Associations for Therapeutics in Vascularized Oral Cyst and Tumours

**Dr. Nausathkhan Ubayathulla [1], Dr. M.R. Muthusekar [2], Dr. Prathiba Ramani [3], Pradeep Kumar Yadalam [4], Dr. Subasree S [5]**

[1] *Phd Scholar, Saveetha Dental College and Research Institute, Saveetha University, SIMATS, Chennai, Tamil Nadu, India*
*Consultant Oral &Maxillofacial surgeon, EHS, Fujairah Specialized Dental Center and Hospital, Fujairah, UAE,*
*Email id : nausathkhan@gmail.com*
[2] *Program Director, Saveetha Dental College and Research Institute, Saveetha University, SIMATS, Chennai, Tamil Nadu, India, Email id : muthusekhar55@gmail.com*
[3] *HOD, Department of Oral and Maxillofacial Pathology, Saveetha Dental College and Research Institute, Saveetha University, SIMATS, Chennai, Tamil Nadu, India, Email id : pratibaramani@saveetha.com*
[4] *Department of Periodontology, Saveetha Dental College and Research Institute, Saveetha University, SIMATS, Chennai, Tamil Nadu, India, Email id : pradeepkumar.sdc@saveetha.com*
[5] *Assistant Professor, Department of Periodontology, Saveetha Dental College and Research Institute, Saveetha University, SIMATS, Chennai, Tamil Nadu, India Email id : subasrees.sdc@saveetha.com*
**Corresponding Author: Dr. M.R. Muthusekar MDS*

**KEYWORDS**
drugs, genes, graph neural networks, attention networks, VEGF, ameloblastoma, dentigerous cyst.

**ABSTRACT**

**Introduction:** Ameloblastoma and dentigerous cysts are aggressive odontogenic lesions that can cause tooth displacement or resorption. Vascular Endothelial Growth Factor (VEGF) is crucial for tumor growth and cyst expansion, with excessive angiogenesis leading to tumor growth. Gene mutations and stem cell markers SOX2 and OCT4 play roles in these lesions. The study uses Residual Graph Attention Networks (RGAN) to predict drug-gene angiogenic associations in ameloblastomas and dentigerous cysts. This helps understand growth patterns and recurrence risks, enabling targeted therapies to inhibit VEGF signaling pathways. The research promotes personalized medicine and multifaceted approaches to managing odontogenic lesions, aiming to predict gene-drug associations for therapeutics in vascularized oral cysts and tumors.

**Methods:** The study analyzed data on VEGF-associated drugs and genes using probe and drug sites, network analysis, and graph neural networks. Data was sourced from PubChem, ChEMBL, DrugBank, and gene databases like Ensembl and KEGG. A graph representing drugs and genes was constructed, with each drug and gene as a unique vertex. Network analysis used metrics like degree centrality, closeness centrality, and betweenness centrality to identify key nodes. Graph Neural Networks (GNNs) were used to analyze associations between drugs and genes. Cytoscape was used to import files and assign drugs and genes to analyze the interactome. At the same time, the CytoHubba plugin identified the top 20 hub drugs and genes using the maximum clique centrality method. It subjected them to residual attention graph neural networks for predictive modeling.

**Results:** The network, with 242 nodes and 333 edges, has a sparse structure with low clustering and density, moderate centralization, and an efficient computational process. The model's performance was assessed using various metrics, including accuracy, confusion matrix, precision, recall, F1-score, ROC-AUC, precision-recall curve, prediction confidence distribution, learning curves, and threshold-based accuracy. It demonstrated exceptional performance, with the best test accuracy of 97.53%, high precision, recall, and F1 scores.

**Conclusion:** The study of drug-gene associations, particularly the role of Vascular Endothelial Growth Factor (VEGF) in ameloblastoma and dentigerous cysts, has significant implications for understanding these conditions and refining clinical management strategies. Ameloblastomas are aggressive odontogenic tumors with a propensity for recurrence, making their management challenging. Targeting the VEGF signaling pathway could limit tumor growth and invasive tendencies, modifying the treatment landscape for ameloblastoma patients. The TARGETS framework has introduced a predictive model for tailored treatments based on unique tumor biology and genetic makeup.

## 1. Introduction

Ameloblastoma and dentigerous cysts are two odontogenic lesions that significantly impact oral health.[1,2]. Ameloblastomas are aggressive and require extensive surgical intervention, while dentigerous cysts are benign but can cause tooth displacement or resorption, necessitating differential diagnosis.[2,3]. Recent studies have spotlighted the pivotal role of vascular endothelial growth factor (VEGF) in angiogenesis, a critical component of tumor growth and cyst expansion. VEGF is a signaling protein that stimulates the formation of new blood vessels, a process essential for supplying nutrients and oxygen to rapidly proliferating cells within tumors, including ameloblastomas. This angiogenic process is crucial for tumor survival and aggressiveness, as enhanced vascularization facilitates local invasion and metastasis.[4,5].

Ameloblastoma and dentigerous cysts[4–6] Are aggressive odontogenic lesions with the potential for recurrence. Angiogenesis, forming new blood vessels, is crucial for tumor growth, invasion, and metastasis. Vascular Endothelial Growth Factor (VEGF) is a key regulator of angiogenesis, making it a prime target for therapeutic intervention. Excessive angiogenesis[7,8] This can lead to tumor growth and increased cystic expansion, contributing to their aggressive behavior. Odontogenic tumors are linked to gene mutations like Ameloblastin, KRAS, FHIT, and P53, while increased HIF1A expression is associated with odontogenic cysts. Elevated PIK3CA and low PTEN levels promote cyst formation, and 50% of dentigerous cysts show loss of heterozygosity. Gene polymorphisms correlate with odontogenic keratocyst[9,10]. Previous studies explore the role of stem cell markers, SOX2 and OCT4, in odontogenic lesions like ameloblastoma, odontogenic keratocyst, and ameloblastic carcinoma. SOX2 is crucial for tissue development and tumorigenesis, with distinct expression patterns in ACA. SOX2 immunostaining could help diagnose ACA and recognize malignant changes in AMBL. OCT4 is inconsistent and ineffective in distinguishing between OKC and AMBL. Further research is needed to understand their roles in malignant transformations and cellular reprogramming.[11].

Targeting VEGF using various approaches can potentially inhibit angiogenesis and the growth and progression of these lesions. Promising strategies include targeting VEGF with VEGF inhibitors, gene therapy, combination therapies, and nanotechnology. These strategies can disrupt the angiogenic process, reduce vascularization, and enhance treatment outcomes. Further research and clinical trials are needed to establish the best strategies for integrating VEGF-targeted therapies into managing these odontogenic tumors, and one more previous study showed a deep learning model that predicts drug responses in tumors using genomic profiles.[12]. It uses mutation and expression data from cancer samples and a mutation encoder. Tested on 622 cancer cell lines, it predicts IC50 values for 265 drugs and identified new therapeutic options. This study uses deep neural network models to predict the impact of anticancer drugs on tumors using the half-maximal inhibitory concentration (IC50). The models combine biological and chemical data to understand genetic profiles and drug compounds. Two autoencoders are pre-trained with high-dimensional gene expression and mutation data of tumors. The models then predict the IC50 value, revealing that RSEM performs better in deep models, and CNNs provide better insight into these data types. The results show the effectiveness of the extracted deep representations in predicting IC50 values, with a mean squared error of 1.06[13,14].

Drug-gene associations[12,15,16] are crucial in personalized medicine and pharmacogenomics, as they help understand how an individual's genetic makeup influences their medication response. Understanding these associations allows healthcare providers to tailor medication choices to an individual's genetic profile, increasing treatment efficacy and minimizing adverse effects. It also improves efficacy by identifying genetic variants that affect drug metabolism, allowing clinicians to prescribe the right drug at the right dose[17]. It also reduces adverse effects by avoiding prescribing medications that could cause harmful reactions. Drug-gene association studies can inform drug development, improve cost-effectiveness, enhance disease mechanisms, guide drug repurposing, and impact public health. As genomics technology evolves, the importance of these associations will grow, paving the way for more individualized and effective patient care[18,19].

Residual Graph Attention Networks (RGATs)[20,21] They are powerful for predicting gene-drug associations in complex diseases like vascularized oral cysts and tumors. These networks combine the strengths of graph neural networks (GNNs) and attention mechanisms to provide a more nuanced understanding of biological interactions. RGATs capture complex relationships by treating genes and drugs as nodes in a graph, where edges represent associations or interactions. The attention mechanism allows the model to weigh the importance of different nodes more flexibly than traditional methods, enhancing the model's capability to focus on the most critical interactions. Residual connections[22,23] RGATs help mitigate the vanishing gradient problem in deep networks, enabling deeper models to train without loss of information. They can integrate various types of biological data, allowing for a more comprehensive analysis of gene-drug associations. RGATs can help identify potential therapeutic targets in vascularized oral cysts and tumors by accurately predicting gene-drug associations.[22]RGATs offer several advantages, including improved interpretability, handling sparsity, noise robustness, network architecture flexibility, and scalability, which are vital for personalized medicine.

Predicting drug-gene associations is crucial in pharmacogenomics and drug discovery for various reasons. Understanding these associations allows for personalized medicine, drug repurposing, mechanisms of action, biomarker discovery, and reducing time and costs.[24]. Machine learning and graph neural networks (GNNs) integrate data sources, create relevant features, and perform supervised and unsupervised learning. By accurately modeling complex biological relationships and leveraging rich biological datasets, these methodologies can facilitate more effective treatments, better patient outcomes, and innovative drug development strategies. Understanding VEGF's role in angiogenesis[25] In ameloblastomas and dentigerous cysts is crucial for managing these lesions. It provides insights into the factors influencing growth patterns and recurrence risks, allowing targeted therapies to inhibit VEGF signaling pathways. Recognizing angiogenesis's role in dentigerous cysts could lead to novel therapeutic strategies, minimizing recurrence rates and improving patient outcomes. This study promotes personalized medicine and multifaceted approaches to managing odontogenic lesions. Precision medicine requires innovative predictions of drug-gene interactions. Advanced computational techniques like graph neural networks help model complex biological processes. Residual Graph Attention Networks (RGAN) enhance prediction accuracy by integrating multiple layers of information. The study uses RGAN to predict drug-gene angiogenic associations, specifically targeting VEGF pathways in ameloblastoma and dentigerous cysts. A residual graph model is constructed to analyze gene interactions and pharmacological agents. No studies understand the drug-gene associations of vegf in dentigerous cysts and ameloblastoma. So, we aim to predict Gene-Drug associations Residual graph attention Networks for Therapeutics in Vascularized Oral Cysts and Tumors.

## 2. Methods

### Data Retrieval

The study retrieved data on VEGF-associated drugs and genes using probes and drug sites.[26] And analyzed them using network analysis and graph neural networks. The data was then pre-processed to normalize and remove missing and duplicate values. A comprehensive approach was taken from data acquisition to analysis, ensuring a comprehensive understanding of the subject matter.

Data was sourced from PubChem, ChEMBL, DrugBank, and gene databases like Ensembl and KEGG. VEGF-associated drugs and genes were identified through literature review and database queries targeting the VEGF pathway. Corresponding genes were retrieved from gene expression databases and genomic studies. A graph representing drugs and genes was constructed, with each drug and gene as a unique vertex. Edges represented interactions between drugs and genes, with the 'target_name' label as the identifying label. Data preprocessing involved normalization, handling missing values, and removing duplicates to prevent bias.

Network analysis uses basic metrics like degree centrality, closeness centrality, and betweenness centrality to identify key nodes in the network. Visualization using tools like NetworkX or Gephi provides an intuitive understanding of the relationships between drugs and genes, aiding in identifying clusters or modules that may indicate functional relationships. This study used Graph Neural Networks (GNNs) to analyze the associations between VEGF-related drugs and genes. The GNN architecture, such as Graph Convolutional Networks (GCNs) or Residual Graph Attention Networks (RGATs), was chosen based on the dataset's characteristics and analysis goals. The GNN was trained using the graph and evaluated using metrics like accuracy, precision, recall, and AUC-ROC. Hyperparameters like learning rate, epochs, batch size, and architecture parameters were optimized through cross-validation. The findings could inform future therapeutic strategies and research directions in VEGF-influenced diseases. (fig-1)

**Cytoscape and Cytohubba**

The study used Cytoscape.[27] To import files and assign drugs and genes to analyze the interactome. At the same time, the CytoHubba plugin identified the top 20 hub drugs and genes using the maximum clique centrality method. The study used Cytoscape, an open-source software platform, to visualize complex networks and integrate data. It imported biological data files, including drug and gene information, to create an interactome. Specific drugs and genes were assigned to the interactome for clearer visualization. The CytoHubba plugin was used to identify key players within the network, enhancing the analysis.

CytoHubba, a tool for measuring drug and gene centrality, identifies the top 20 hub drugs and genes. This method helps us understand the interactome dynamics, potential therapeutic interventions, and biological processes, contributing to our understanding of the studied biological system and its underlying mechanisms.

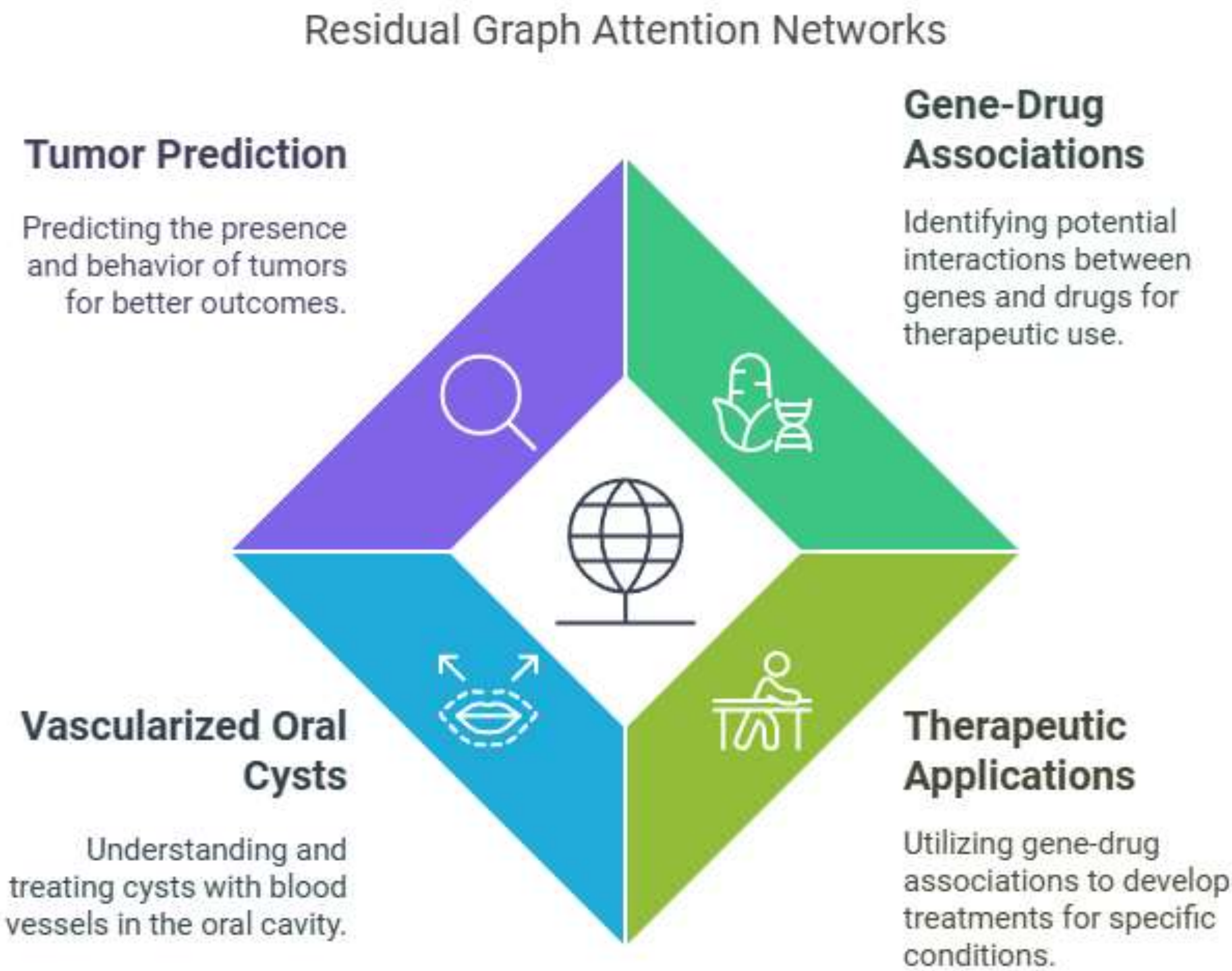


Figure 1 shows the workflow model.

**Model Architecture and Training of ResidualAttentionGNN**

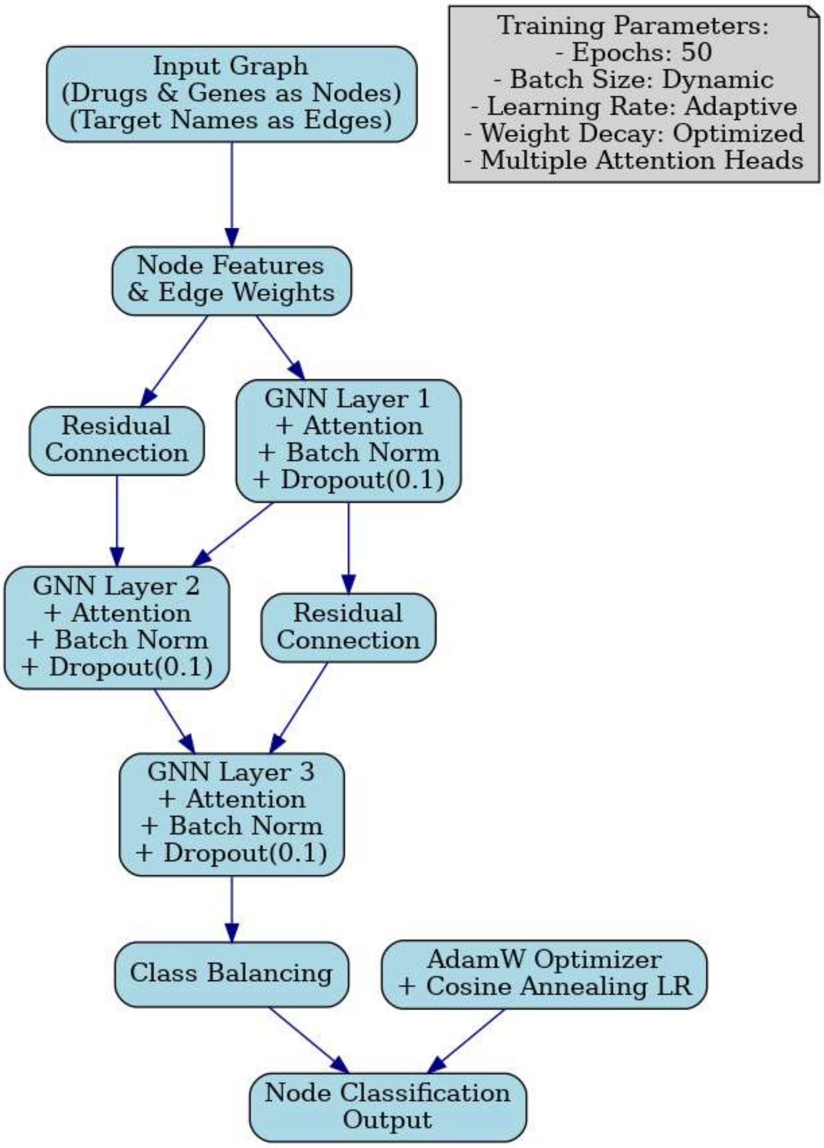


Figure 2 shows the architecture of the model.

**Residual attention graph neural network architecture**

The model uses an input graph with nodes representing drugs and genes, edges representing target names, and edge weights derived from activity_biochemical values. Features are processed to create initial representations. The GNN layers feature an attention mechanism, batch normalization for stability, and a 0.1 rate dropout to prevent overfitting. Two residual connections prevent vanishing gradients. Training components include Class Balancing, AdamW Optimizer, 50 epoch training, dynamic batch sizing, adaptive learning rate, optimized weight decay, and multiple attention heads. After combining all processed information, the model generates predictions about node classes in the final layer. (fig-2)

The Residual Attention GNN model utilizes advanced techniques to enhance node classification performance in complex graph datasets, incorporating deep learning techniques and key hyperparameters during training. The model uses residual connections to bypass certain layers, improving gradient flow and addressing the vanishing gradient problem. Attention mechanisms prioritize important nodes and features, enhancing the model's ability to capture complex patterns in graph data. Batch normalization stabilizes the training process, while dropout regularization tackles overfitting by randomly setting each neuron to zero with a 10% dropout rate. Class balancing prevents bias towards the majority class and improves the performance of minority classes. The

ResidualAttentionGNN model combines residual connections, attention mechanisms, batch normalization, dropout regularization, and class balancing. It includes the AdamW optimizer and a dynamic learning rate adjustment strategy through cosine annealing to enhance performance on complex graph-based datasets. The learning rate, weight decay, dropout rate, batch size, number of epochs-50, and attention heads are all set to optimize the model's performance.

## 3. Results

The network comprises 242 nodes and 333 edges, indicating a sparse structure with low clustering and density. The average number of neighbors is 2.736, suggesting that most nodes have few direct connections. The network's diameter is 4, suggesting a limited overall reach when connecting nodes through paths. The radius indicates that from the most central nodes, you can reach other nodes within a distance of 2 edges. The average shortest path length between pairs of nodes is approximately 3.142 edges, indicating a moderately connected structure. The clustering coefficient 0 indicates no tightly-knit groups of interconnected nodes, suggesting a more tree-like structure. The network density is very low, indicating sparse connectivity. The network heterogeneity is higher, suggesting a disparity in node connectivity. The network centralization is moderate, with some key nodes serving as critical hubs. The network has only one connected component, indicating all nodes are part of a single unified structure without isolated nodes. The analysis time is 0.089 seconds, indicating an efficient computational process.

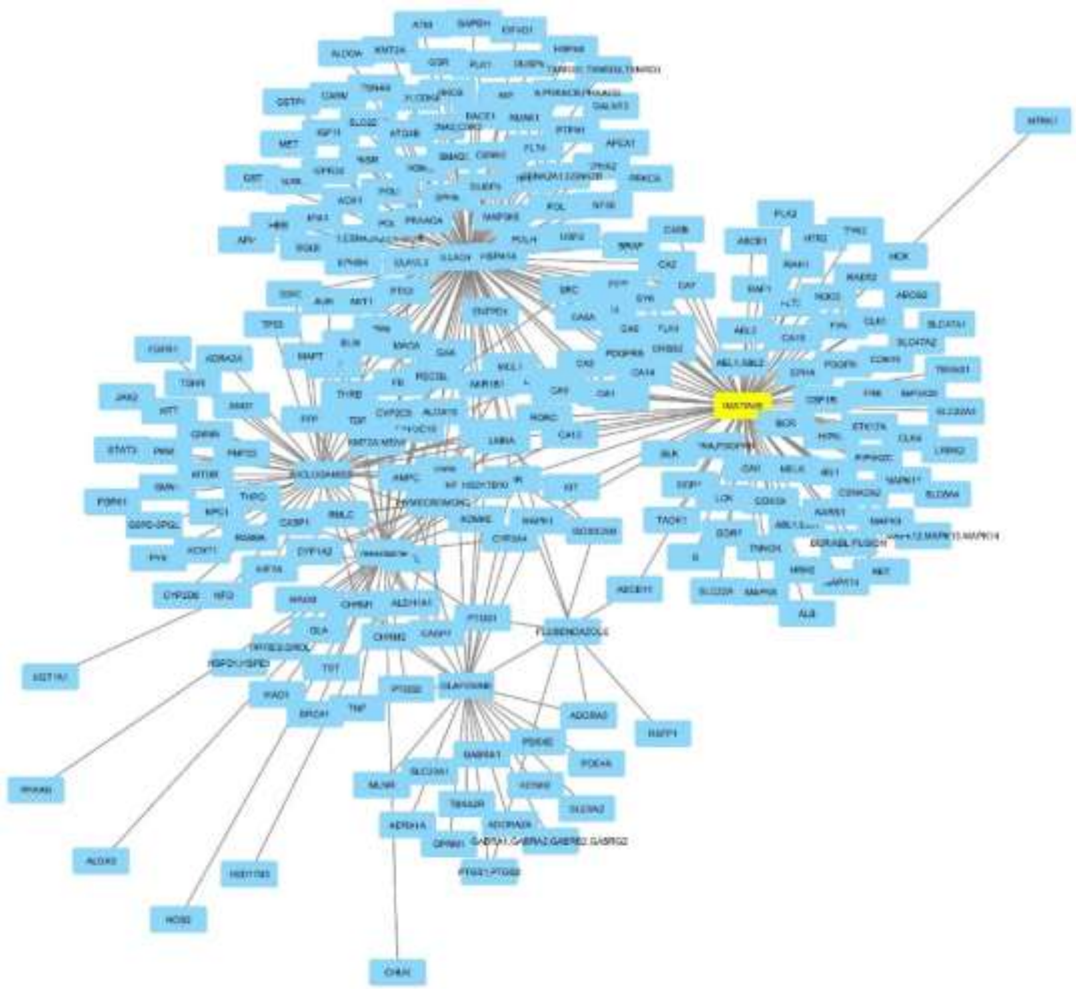

Figure 3 shows the interactome of the drugs and genes associated with VEGF FAMILY.

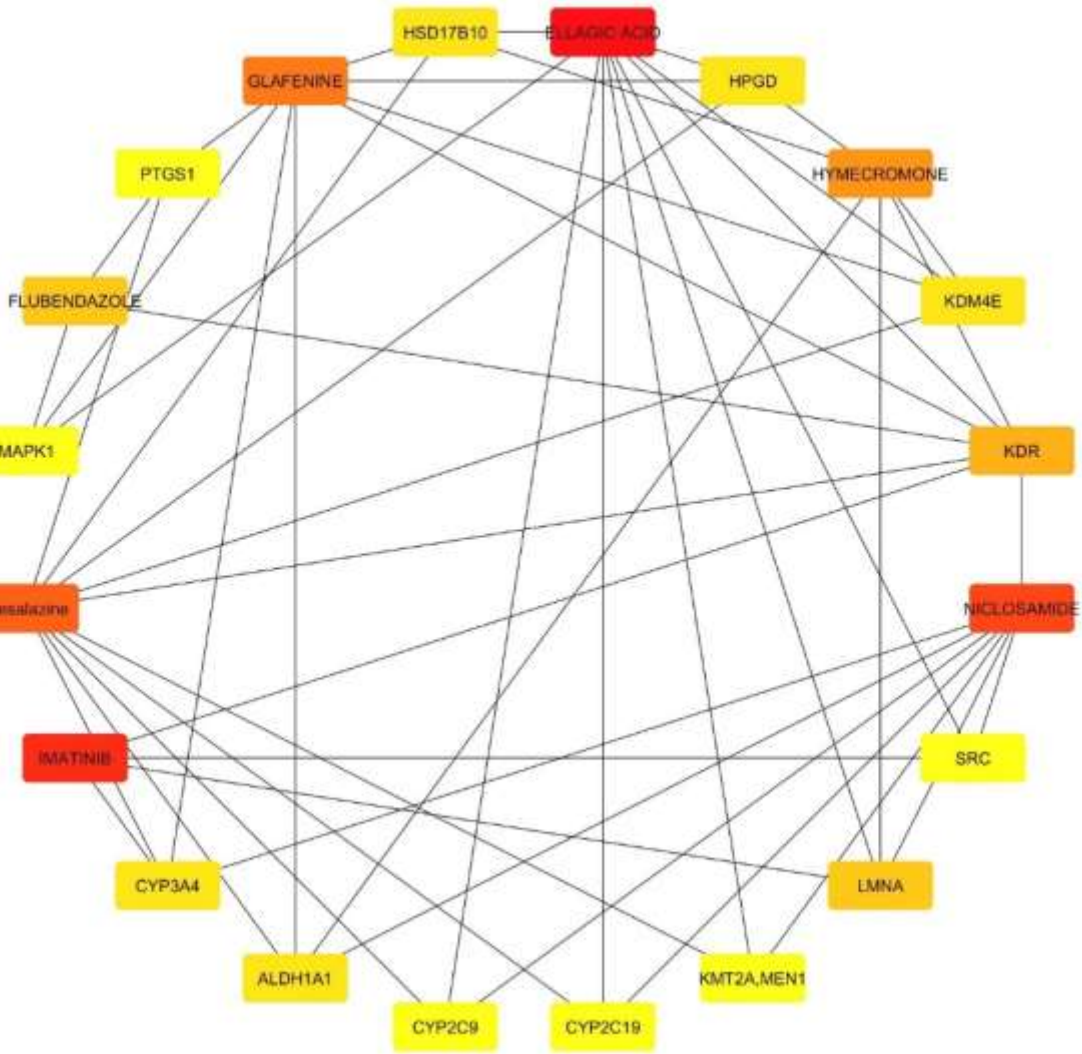


Figure 4 shows the top ten hub drugs and genes associated with VEGF are MAPK1, FLUBENDAZOLE, PTGS1, KDR, KDM4E, LMNA, CYP2C19, CYP2C9, ALDH1A1, CYP3A4, IMATINIB, MESALAZINE

## 2. Evaluation Metrics

The model's performance was evaluated using accuracy, confusion matrix, precision, recall, F1-score, ROC-AUC, precision-recall curve, prediction confidence distribution, learning curves, and threshold-based accuracy. These metrics provided insights into the model's performance for each class, highlighting the trade-off between precision and recall, predicting confidence levels, and monitoring overfitting or underfitting. The model demonstrated exceptional performance, with a best test accuracy of 97.53%. It correctly classified 162 instances of Class 0 and 75 instances of Class 1, with only six misclassifications. The model achieved high precision, recall, and F1 scores, with a ROC-AUC score of 0.9959 and an average precision of 0.9923. The model maintained perfect accuracy at higher confidence thresholds, but coverage decreased. (fig-3,4)

The ResidualAttentionGNN model demonstrated advanced performance on the node classification task, achieving high accuracy, precision, recall, and F1 scores. Its robustness and reliability make it suitable for real-world applications. Further improvements could involve exploring additional architectures, feature engineering, or fine-tuning hyperparameters. The model's performance is highlighted through ROC, precision-recall curves, prediction confidence, learning curves, and accuracy progression.

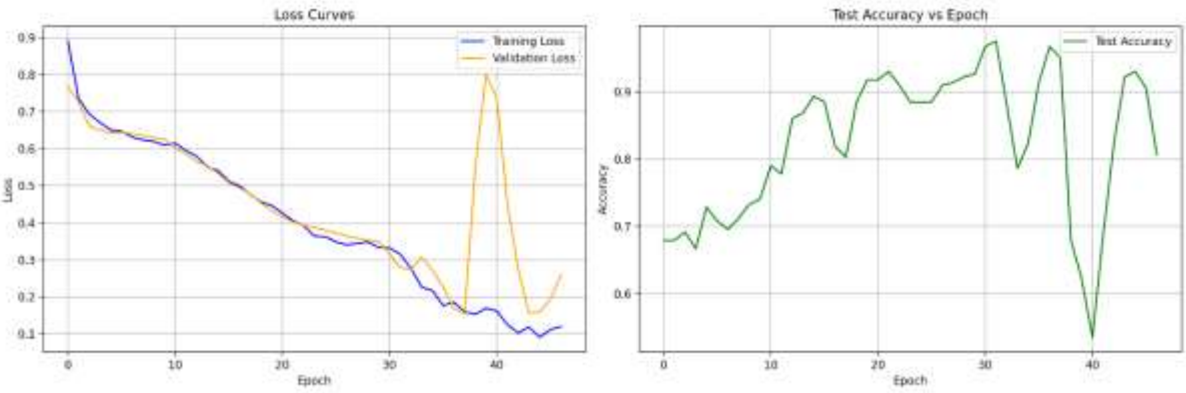


Figure 5

FIG-5 shows the two graphs illustrating the performance metrics of a machine learning model during training and evaluation. The training loss curve decreases over epochs, indicating good learning, while the validation loss curve shows fluctuation, potentially overfitting, and does not decrease as reliably. The graph shows test accuracy, which fluctuates but generally trends upwards. It shows high values at certain epochs, suggesting good performance on unseen data. Training loss decreases consistently, but validation loss suggests overfitting. Test accuracy shows variability, suggesting potential instability or overfitting that affects robustness. Monitoring both loss and accuracy metrics helps assess model performance and guides decisions regarding potential adjustments to the training process, such as using regularization techniques or altering the learning rate.

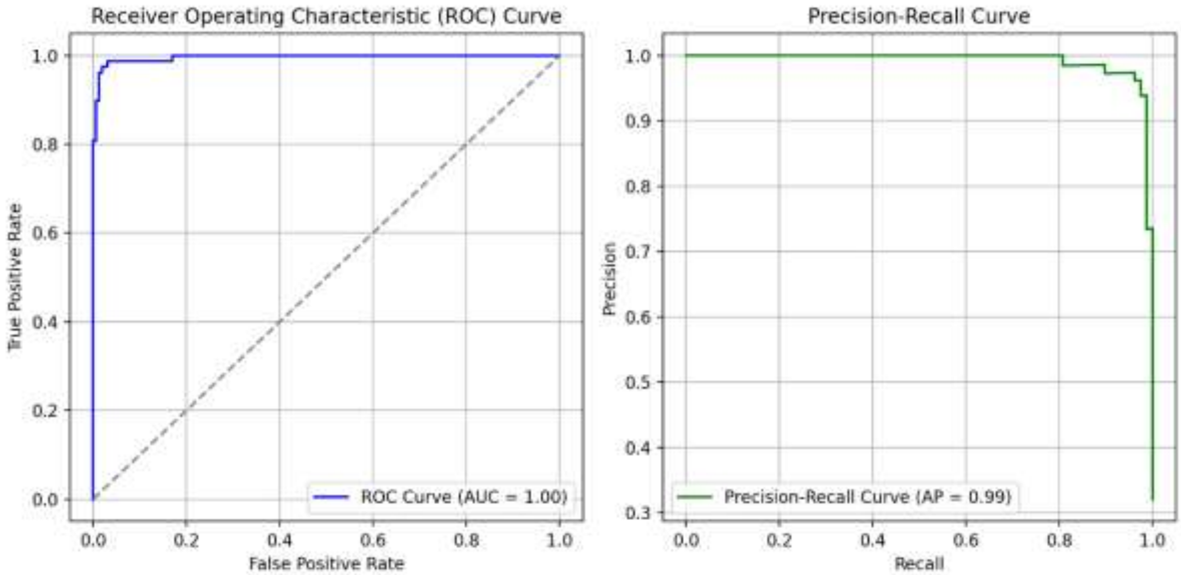


Figure 6

Fig-6 shows the Receiver Operating Characteristic (ROC) and Precision-Recall curves, crucial evaluation curves for binary classification models. The ROC curve, represented by a solid blue line, indicates a low false positive rate and a high true positive rate. The dashed diagonal line represents random classifier performance. The AUC is 1.00, indicating perfect classifier performance. The Precision-Recall curve shows a solid green line with minimal fluctuation, indicating high precision even as recall increases. The model's average precision (AP) is 0.99, indicating excellent performance, especially in imbalanced datasets. Overall, the model shows outstanding performance in both graphs. These curves provide a comprehensive view of the model's ability to balance sensitivity and specificity, showcasing its robustness in making accurate predictions.

**Additional Evaluation Metrics:**

The study's evaluation results show high accuracy and confidence in the model, with an average confidence of 0.3822 for correct predictions and 0.5065 for incorrect ones. The model's performance metrics on the test dataset demonstrated its effectiveness in classifying target classes, as detailed in the following results. The model's high ROC-AUC score of 0.9959 indicates excellent classification ability. In contrast, its Average Precision score of 0.9923 demonstrates strong performance, especially in class imbalance scenarios, indicating excellent accuracy across various classification thresholds. The model's performance metrics for Class 0 show high accuracy with a precision of 0.9818, recall of 0.9818, an F1-score of 0.9818, and support of 165, indicating robust sensitivity and a substantial representation against which the model is evaluated.

The model's Class 1 performance is highly reliable, with a precision score of 96.15%, a recall score of 96.15%, and a good balance between precision and recall. However, the total number of instances for Class 1 is lower than Class 0, indicating potential class imbalance. The model's confidence statistics show low confidence for correct predictions (38.22%) and higher confidence for incorrect predictions (50.65%). The confidence standard deviation (0.2780) indicates inconsistencies in the model's confidence levels across different instances, indicating potential caution in its predictions. The model outperforms the dataset with high ROC-AUC and Average Precision scores. However, the average confidence for correct predictions is low, suggesting further model confidence calibration may be beneficial.

## 4. Discussion

Ameloblastomas are benign, invasive, and aggressive tumors that can exhibit aggressive behavior due to reduced oxygen concentrations and hypoxia. These tumors have been linked to hypoxia-induced angiogenesis, leading to tumor progression and increased invasiveness. This study investigated the expression of HIF-1α, MMP-2, VEGF, and VEGFR-2 proteins in 24 ameloblastoma cases, 10 calcifying odontogenic cysts, and nine dental follicles using immunohistochemistry. Results showed higher expression of these proteins in ameloblastomas than in dental follicles and cysts, suggesting potential roles in the tumor's biological behavior. A previous study examined 45 cases of keratocystic odontogenic tumors, dentigerous cysts, and ameloblastomas to assess their immunohistochemical expression of vascular endothelial growth factor (VEGF) and argyrophilic nucleolar organizer regions (AgNORs)[1,2].

Vascular Endothelial Growth Factor (VEGF) is a key gene and drug involved in various biological processes, including angiogenesis, cell signaling, and inflammation. The top ten hub genes and genes associated with VEGF include MAPK1, KDR, and KDM4E. MAPK1 promotes cell proliferation and differentiation, while KDR mediates VEGF's pro-angiogenic effects in endothelial cells[1,3]. Binding VEGF to KDR activates downstream signaling pathways, promoting endothelial cell survival, proliferation, and migration. KDM4E, involved in the demethylation of histone proteins, plays a role in chromatin remodeling and gene expression. These genes and drugs play critical roles in VEGF's regulation.[28,29].

PTGS1, also known as COX-1, converts arachidonic acid to prostaglandins, mediators of inflammation. These prostaglandins stimulate the release of VEGF, contributing to angiogenesis and inflammation. LMNA, a nuclear envelope protein, maintains nuclear integrity and gene regulation. Alterations in lamins can affect gene expression in angiogenesis, including VEGF. ALDH1A1, involved in retinoic acid metabolism, regulates angiogenesis by affecting VEGF expression. CYP2C19, involved in drug metabolism, may influence metabolic pathways of drugs that modulate angiogenesis and indirectly affect VEGF. CYP2C9, similar to CYP2C19, plays a role in drug metabolism, activation, and detoxification and may impact drug metabolism related to VEGF-modulating therapies[9,10]. CYP3A4, an enzyme in drug metabolism, affects the pharmacokinetics of many drugs and is associated with VEGF. Drugs that affect VEGF signaling could be metabolized by CYP3A4, influencing their therapeutic effects. Hub drugs include flubendazole, an anti-parasitic agent that inhibits angiogenesis; imatinib, a tyrosine kinase inhibitor for cancer treatment; and mesalazine, an anti-inflammatory drug used in inflammatory bowel diseases. These drugs modulate the VEGF pathway, promoting or inhibiting angiogenesis and disease processes. Understanding these associations can help develop targeted therapies and improve clinical outcomes in abnormal blood vessel growth conditions.[30].

The ResidualAttentionGNN model demonstrated exceptional performance, with a best test accuracy of 97.53%. (fig-3,4,5,6) It correctly classified Class 0 instances and Class 1 instances with six misclassifications. The model achieved high precision, recall, and F1 scores, with a ROC-AUC score of 0.9959 and an average precision of 0.9923. Further improvements could involve additional architectures, feature engineering, or fine-tuning hyperparameters. The study evaluates a high accuracy and confidence model, with an average confidence of 0.3822 for correct predictions and 0.5065 for incorrect ones. The model's performance metrics show excellent classification ability, particularly in class imbalance scenarios. Class 0 performance is high, with precision and recall scores of 0.9818 and 0.9818, respectively. Class 1 performance is reliable, but low confidence for correct predictions and higher confidence for incorrect ones. Further model confidence calibration may be beneficial, similar to this study investigating the complex relationship between microbes and pharmaceuticals, including drug sourcing, microbial degradation, drug resistance, and microbial community involvement. Researchers developed a predictive model called GCGACNN, which uses data from databases to create association and similarity matrices. The model, which uses a multi-layer Graph Neural Network and a two-dimensional Convolutional Neural Network, shows superior predictive performance, and computer-aided drug-drug interaction prediction (DDI)[18,19] models rely on drug-associated features or networks, neglecting potential information from drug-related biological entities like targets and genes. An attention-based cross-domain graph neural network (ACDGNN) is proposed to address these limitations. ACDGNN [31]considers drug-related entities and propagates information through cross-domain operation, eliminating heterogeneity between entities. It can be used in both transductive and inductive settings.

Future directions[17–19,32] For the model including improving model calibration, addressing the class imbalance, incorporating additional genetic and environmental factors, conducting longitudinal studies, integrating clinical data, exploring other machine learning algorithms, and collaborating across multiple research centers[13,14,33]. These steps aim to improve the model's accuracy and predictive capabilities, address potential class imbalance, and enhance its predictive power. Additionally, the study should explore other machine learning techniques to compare performance and ensure diverse datasets for more robust conclusions.

The study has limitations, including a limited dataset size, potential overfitting, and lack of model interpretability. It also focuses on specific genes related to VEGF, which could be expanded to include a broader range of genes. The data may not reflect dynamic patient conditions or treatment responses over time, necessitating adaptive models. Additionally, the model's performance needs validation using independent cohorts to ensure applicability across different populations.

## Clinical importance

The study on drug-gene associations of VEGF in ameloblastomas and dentigerous cysts provides valuable insights into the aggressive behavior of these tumors. The higher expression of VEGF proteins in ameloblastomas compared to dentigerous cysts suggests a direct link between hypoxia-induced angiogenesis mediated by VEGF and tumor progression. This understanding enhances our knowledge of the biological mechanisms underlying the invasive nature of ameloblastomas and potentially guiding future therapeutic strategies.

The study identifies genes and pathways linked to VEGF, including MAPK1, KDR, and KDM4E, which are crucial for angiogenesis and tumorigenesis. This could lead to targeted therapies inhibiting angiogenic processes and restricting tumor growth, potentially improving treatment options for ameloblastomas. The study's findings can inform treatment strategies incorporating agents targeting the VEGF pathway in conjunction with standard therapies. The predictive model introduced in the study (TARGETS) brings potential clinical utility for improving patient selection for therapeutic interventions. Future research and collaboration may include larger, multi-institutional cohorts to validate predictive models and expand understanding of gene-drug interactions in ameloblastomas and related conditions.

## 5. Conclusion

The study of drug-gene associations, particularly the role of Vascular Endothelial Growth Factor (VEGF) in ameloblastoma and dentigerous cysts, has significant implications for understanding these conditions and refining clinical management strategies. Ameloblastomas are aggressive odontogenic tumors with a propensity for recurrence, making their management challenging. The study found that VEGF and its associated proteins are significantly elevated in ameloblastomas compared to dentigerous cysts, indicating the role of VEGF in fostering aggressive tumors. Targeting the VEGF signaling pathway could limit tumor growth and invasive tendencies, modifying the treatment landscape for ameloblastoma patients. Identifying genes and pathways essential to angiogenesis reinforces the potential for developing targeted therapies. Exploring existing drugs targeting the VEGF pathway opens up possibilities for repurposing agents like imatinib, mesalazine, and flubendazole for ameloblastoma treatment. The study also has significant implications for biomarker development, as VEGF could be a potential biomarker for diagnosing ameloblastomas and distinguishing them from less aggressive lesions like dentigerous cysts. The exploration of the VEGF pathway in ameloblastoma provides a deeper understanding of aggressive tumor behavior and opens up new therapeutic possibilities. Future research should focus on larger cohorts, gene-drug interactions, and collaborations for innovation. Prioritizing personalized medicine can improve therapeutic efficacy and prognosis for ameloblastoma patients.

## References:

[1] Ghai S. Ameloblastoma: An Updated Narrative Review of an Enigmatic Tumor. Cureus 2022;14:e27734. https://doi.org/10.7759/cureus.27734.

[2] Oh K-Y. Treatment options for advanced ameloblastoma in the era of precision medicine: A brief review. Oral Oncol 2023;146:106585. https://doi.org/10.1016/j.oraloncology.2023.106585.

[3] Marín-Márquez C, Kirby J, Hunter KD. Molecular pathogenesis of ameloblastoma. J Oral Pathol Med 2024;53:277–93. https://doi.org/10.1111/jop.13538.

[4] Nahajowski M, Hnitecka S, Antoszewska-Smith J, Rumin K, Dubowik M, Sarul M. Factors influencing an eruption of teeth associated with a dentigerous cyst: a systematic review and meta-analysis. BMC Oral Health 2021;21:180. https://doi.org/10.1186/s12903-021-01542-y.

[5] Rajae EG, Karima EH. Dentigerous cyst: enucleation or marsupialization? (a case report). Pan Afr Med J 2021;40:149. https://doi.org/10.11604/pamj.2021.40.149.28645.

[6] Lai P-T, Li C-Y, Wu Y-C, Chiang C-P. Glandular odontogenic cyst in a dentigerous relationship. J Dent Sci 2022;17:1058–60. https://doi.org/10.1016/j.jds.2021.12.011.

[7] Toprani SM. DNA damage and repair scenario in ameloblastoma. Oral Oncol 2020;108:104804. https://doi.org/10.1016/j.oraloncology.2020.104804.

[8] Shi HA, Ng CWB, Kwa CT, Sim QXC. Ameloblastoma: A succinct review of the classification, genetic understanding, and novel molecular targeted therapies. Surgeon 2021;19:238–43. https://doi.org/10.1016/j.surge.2020.06.009.

[9] Sadri D, Farhadi S, Nourmohamadi P. Angiogenesis in odontogenic keratocyst and dentigerous cyst: Evaluation of JunB and VEGF expression. Dent Res J (Isfahan) 2019;16.

[10] Martiny-Baron G, Marmé D. VEGF-mediated tumor angiogenesis: a new target for cancer therapy. Curr Opin Biotechnol 1995;6:675–80. https://doi.org/https://doi.org/10.1016/0958-1669(95)80111-1.

[11] Su S, Liu M, Zhou J, Zhang J. GCGACNN: A Graph Neural Network and Random Forest for Predicting Microbe–Drug Associations. Biomolecules 2024;14. https://doi.org/10.3390/biom14080946.

[12] Chiu Y-C, Chen H-IH, Zhang T, Zhang S, Gorthi A, Wang L-J, et al. Predicting drug response of tumors from integrated genomic profiles by deep neural networks. BMC Med Genomics 2019;12:18. https://doi.org/10.1186/s12920-018-0460-9.

[13] Harris J, Yadalam PK, Anegundi RV, Arumuganainar D. Comparing Graph Sample and Aggregation (SAGE) and Graph Attention Networks in the Prediction of Drug-Gene Associations of Extended-Spectrum Beta-Lactamases in Periodontal Infections and Resistance. Cureus 2024;16:e68082. https://doi.org/10.7759/cureus.68082.

[14] Wang Y, Qi J, Chen X. Accurate Prediction of Epigenetic Multi-Targets with Graph Neural Network-Based Feature Extraction. Int J Mol Sci 2022;23. https://doi.org/10.3390/ijms232113347.

[15] Li Y, Umbach DM, Krahn JM, Shats I, Li X, Li L. Predicting tumor response to drugs based on gene-expression biomarkers of sensitivity learned from cancer cell lines. BMC Genomics 2021;22:272. https://doi.org/10.1186/s12864-021-07581-7.

[16] Abbasi M, Carvalho FG, Ribeiro B, Arrais JP. Predicting drug activity against cancer through genomic profiles and SMILES. Artif Intell Med 2024;150:102820. https://doi.org/https://doi.org/10.1016/j.artmed.2024.102820.

[17] Zhao C, Wang H, Qi W, Liu S. Toward drug-miRNA resistance association prediction by positional encoding graph neural network and multi-channel neural network. Methods 2022;207:81–9. https://doi.org/10.1016/j.ymeth.2022.09.005.

[18] Yu H, Li K, Dong W, Song S, Gao C, Shi J. Attention-based cross domain graph neural network for prediction of drug-drug interactions. Brief Bioinform 2023;24. https://doi.org/10.1093/bib/bbad155.

[19] Zhang Z-R, Jiang Z-R. GraphDPA: Predicting drug-pathway associations by graph convolutional networks. Comput Biol Chem 2022;99:107719. https://doi.org/https://doi.org/10.1016/j.compbiolchem.2022.107719.

[20] Wang H, Dai C, Wen Y, Wang X, Liu W, He S, et al. GADRP: graph convolutional networks and autoencoders for cancer drug response prediction. Brief Bioinform 2023;24. https://doi.org/10.1093/bib/bbac501.

[21] Wang S, Qiao J, Feng S. Prediction of lncRNA and disease associations based on residual graph convolutional networks with attention mechanism. Sci Rep 2024;14:5185. https://doi.org/10.1038/s41598-024-55957-y.

[22] Zhang L, Wang C-C, Zhang Y, Chen X. GPCNDTA: Prediction of drug-target binding affinity through cross-attention networks augmented with graph features and pharmacophores. Comput Biol Med 2023;166:107512. https://doi.org/10.
1016/j.compbiomed.2023.107512.

[23] Aly Abdelkader G, Ngnamsie Njimbouom S, Oh T-J, Kim J-D. ResBiGAAT: Residual Bi-GRU with attention for protein-ligand binding affinity prediction. Comput Biol Chem 2023;107:107969. https://doi.org/10.1016/j.compbiolchem.2023.
107969.

[24] Jang HY, Song J, Kim JH, Lee H, Kim I-W, Moon B, et al. Machine learning-based quantitative prediction of drug exposure in drug-drug interactions using drug label information. NPJ Digit Med 2022;5:88. https://doi.org/10.1038/ s41746-022-00639-0.

[25] Bagherian M, Sabeti E, Wang K, Sartor MA, Nikolovska-Coleska Z, Najarian K. Machine learning approaches and databases for prediction of drug–target interaction: a survey paper. Brief Bioinform 2020;22:247–69. https://doi.org/10.
1093/bib/bbz157.

[26] Skuta C, Popr M, Muller T, Jindrich J, Kahle M, Sedlak D, et al. Probes & Drugs portal: an interactive, open data resource for chemical biology. Nat Methods 2017;14:759–60. https://doi.org/10.1038/nmeth.4365.

[27] 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. https://doi.org/10.1101/gr.1239303.

[28] Khodabakhsh F, Merikhian P, Eisavand MR, Farahmand L. Crosstalk between MUC1 and VEGF in angiogenesis and metastasis: a review highlighting roles of the MUC1 with an emphasis on metastatic and angiogenic signaling. Cancer Cell Int 2021;21:200. https://doi.org/10.1186/s12935-021-01899-8.

[29] Alsafadi R, Almohareb M. The Importance of Vascular Endothelial Growth Factor (VEGF) in Aggressiveness of Odontogenic Lesions. JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH 2019;13. https://doi.org/10.7860/

JCDR/2019/40465.12810.

[30] Chen Z, Zhao P, Li C, Li F, Xiang D, Chen Y-Z, et al. iLearnPlus: a comprehensive and automated machine-learning platform for nucleic acid and protein sequence analysis, prediction and visualization. Nucleic Acids Res 2021;49:e60. https://doi.org/10.1093/nar/gkab122.

[31] Rassil A, Chougrad H, Zouaki H. Augmented Graph Neural Network with hierarchical global-based residual connections. Neural Netw 2022;150:149–66. https://doi.org/10.1016/j.neunet.2022.03.008.

[32] Yadalam PK, Natarajan PM, Mosaddad SA, Heboyan A. Graph neural networks-based prediction of drug gene association of P2X receptors in periodontal pain. J Oral Biol Craniofac Res 2024;14:335–8. https://doi.org/10.1016/j.jobcr.2024.
04.008.

[33] Zhu Y, Ouyang Z, Chen W, Feng R, Chen DZ, Cao J, et al. TGSA: protein-protein association-based twin graph neural networks for drug response prediction with similarity augmentation. Bioinformatics 2022;38:461–8. https://doi.org/10.
1093/bioinformatics/btab650.