ARTÍCULO ORIGINAL

Gac Méd Caracas 2025;133(2):399-408
DOI: 10.47307/GMC.2025.133.2.11

Leveraging LSGAN for synthetic gingival keratinization genomic data: Insights from drug interaction and gene ontology in an early fusion framework

Aprovechamiento de LSGAN para datos genómicos sintéticos de queratinización gingival: perspectivas de la interacción farmacológica y la ontología génica en un marco de fusión temprana

Pradeep Kumar Yadalam[1a*], Raghavendra Vamsi Anegundi[2b], Carlos M. Ardila[3a,c,*]

SUMMARY

Introduction: *Gingival keratinization, a vital process in oral health, involves the formation of a keratin-rich protective epithelial layer, providing resilience against mechanical stress, pathogens, and environmental factors.* **Objective:** *This study employs an early fusion omics approach with Least-Squares Generative Adversarial Networks (LSGAN) to generate synthetic genomic data, incorporating insights from drug interactions and geneontology annotations.* **Methods:** *Gene expression data from the NCBI GEO dataset (GSE182196) were analyzed to identify differentially expressed genes (DEGs) across diverse samples. Functional enrichment was performed using the Comparative Toxicogenomics Database (CTD) to explore chemical exposures and biological processes linked to DEGs. Outputs were standardized into TSV formats for downstream analyses. To ensure high-fidelity synthetic data generation, the LSGAN framework was optimized to minimize Mean Squared Error (MSE) and Mean Absolute Error (MAE).* **Results:** *The LSGAN model demonstrated robust performance, with low MSE and MAE values indicating a close resemblance between synthetic and real genomic data distributions. Additionally, reduced Wasserstein distances highlighted the enhanced similarity of synthetic data to the original dataset, confirming the model's reliability in preserving biologically relevant features.* **Conclusions:** *The LSGAN approach successfully generates high-quality synthetic genomic data for gingival keratinization, enabling advanced hypothesis testing, machine learning model training, and simulation of rare genomic conditions.*

Keywords: *Computational biology, epithelial tissue, gingiva, keratin, periodontium.*

DOI: https://doi.org/10.47307/GMC.2025.133.2.11

ORCID: https://orcid.org/0000-0003-4259-820X[1]
ORCID: https://orcid.org/0000-0002-8269-088X[2]
ORCID: https://orcid.org/0000-0002-3663-1416[3]

[a]Ph.D. Department of Periodontics, Saveetha Dental College, Saveetha Institute of Medical and Technology Sciences, SIMATS, Saveetha. University, Chennai, Tamil Nadu, India.
[b]MDS. Department of Periodontics, Saveetha Dental College, Saveetha Institute of Medical and Technology Sciences, SIMATS, Saveetha. University, Chennai, Tamil Nadu, India.
[c]Ph.D. Postdoctoral Researcher. Basic Sciences Department, Biomedical Stomatology Research Group, Universidad de Antioquia U de A, Medellín, Colombia.
[*] Author of Correspondence: Carlos M. Ardila, Universidad de Antioquia. Calle 70 No. 52-21, Medellín, Colombia. Phone: 57-4-2196700. FAX 057-4-2195332, E-mail: martin.ardila@udea.edu.co

Recibido: 22 de marzo 2025
Aceptado: 1 de mayo 2025

RESUMEN

Introducción: *La queratinización gingival, un proceso vital en la salud bucal, implica la formación de una capa epitelial protectora rica en queratina, proporcionando resistencia contra el estrés mecánico, patógenos y factores ambientales.* **Objetivo:** *Este estudio emplea un enfoque ómico de fusión temprana con Redes Generativas Adversarias de Mínimos Cuadrados (LSGAN) para generar datos genómicos sintéticos, incorporando perspectivas de interacciones farmacológicas y anotaciones de ontología génica.* **Métodos:** *Se analizaron datos de expresión génica del conjunto de datos NCBI GEO (GSE182196) para identificar genes diferencialmente expresados (DEGs) en diversas muestras. Se realizó un enriquecimiento funcional utilizando la Base de Datos de Toxicogenómica Comparativa (CTD) para explorar exposiciones químicas y procesos biológicos vinculados a los DEGs. Las salidas se estandarizaron en formatos TSV para análisis posteriores. El marco LSGAN se optimizó para minimizar el Error Cuadrático Medio (MSE) y el Error Absoluto Medio (MAE) para asegurar la generación de datos sintéticos de alta fidelidad.* **Resultados:** *El modelo LSGAN demostró un rendimiento robusto, con valores bajos de MSE y MAE que indican una estrecha semejanza entre las distribuciones de datos genómicos sintéticos y reales. Además, las distancias de Wasserstein reducidas resaltaron la similitud mejorada de los datos sintéticos con el conjunto de datos original, confirmando la fiabilidad del modelo en la preservación de características biológicamente relevantes.* **Conclusiones:** *El enfoque LSGAN genera con éxito datos genómicos sintéticos de alta calidad para la queratinización gingival, permitiendo pruebas de hipótesis avanzadas, entrenamiento de modelos de aprendizaje automático y simulación de condiciones genómicas raras.*

Palabras clave: *Biología computacional, tejido epitelial, encía, queratina, periodonto.*

INTRODUCTION

Gingival keratinization is a vital protective mechanism in which epithelial tissues develop a keratin-rich outer layer. This specialized adaptation serves as a crucial barrier against mechanical stress, microbial invasion, and environmental challenges in the oral cavity (1,2). The keratinized epithelium, predominantly found in the attached gingiva and hard palate, comprises multiple stratified cell layers that collectively provide mechanical protection while supporting local immune responses. The integrity of this barrier is clinically significant, as compromised keratinization increases susceptibility to inflammation, infection, and periodontal disease progression. Multiple factors influence this process, including genetic predisposition, oral hygiene practices, hormonal fluctuations, local trauma, and systemic health conditions (3).

Keratin proteins form an essential cytoskeletal network that maintains epithelial cell structure and function. These intermediate filaments, particularly the type I and II keratin heteropolymers, participate in fundamental cellular processes ranging from proliferation and differentiation to intracellular transport and cell-cell signaling. Despite their biological importance, the complete architecture of cytokeratin networks in oral tissues remains inadequately characterized, especially when studying ex vivo specimens where native tissue organization may be altered (2,4). This knowledge gap underscores the need for advanced investigative approaches.

In clinical periodontology, enhancing gingival keratinization represents a key therapeutic objective. Various interventions, including soft tissue grafting, periodontal surgical techniques (5,6), antimicrobial therapies, and comprehensive patient education programs, aim to optimize keratinized tissue dimensions and quality. Successful outcomes depend on careful case selection, adjunctive therapies, and consistent postoperative monitoring, highlighting the importance of individualized treatment planning and long-term maintenance (7).

The emergence of omics technologies has revolutionized our capacity to investigate complex biological processes like keratinization. By integrating genomic, transcriptomic, proteomic, and metabolomic datasets, researchers can now examine keratinocyte differentiation through multidimensional lenses. Early fusion omics approaches are particularly valuable, as they synthesize information across biological scales to reveal the interconnected regulatory networks governing epithelial specialization (8). This system-level perspective provides unprecedented opportunities to decode the molecular basis of oral barrier function.

Recent advances in computational biology have introduced innovative tools for keratinization research, including generative adversarial networks (GANs). The Least-Squares GAN (LSGAN) framework has demonstrated particular promise in synthesizing high-quality omics data, enabling researchers to model gene expression patterns, predict treatment responses, and explore disease mechanisms with enhanced flexibility (9). Unlike conventional GANs, LSGAN's modified loss function improves training stability while minimizing common artifacts, yielding more biologically plausible synthetic datasets.

The unique advantages of LSGAN architecture make it especially suitable for biomedical applications. Its robust performance stems from the least-squares loss function, which effectively addresses vanishing gradient problems while maintaining sample diversity. Clinical validations of this approach, including successful implementations in glioblastoma imaging studies (10-14), confirm its potential for generating reliable synthetic biological data. When combined with gene ontology frameworks, LSGAN-generated datasets can systematically categorize genes by their functional roles in keratinization processes (11,15).

Despite these advancements, applying LSGAN technology to keratinization research remains a significant challenge. Current limitations include the need for larger, higher-quality training datasets, more sophisticated model interpretability tools, and standardized validation protocols. Addressing these hurdles through multidisciplinary collaboration will be essential for realizing the full potential of synthetic data approaches in oral health research and clinical applications.

MATERIALS AND METHODS

Data Collection and Preparation

Data Retrieval from NCBI GEO (GSE182196):

Raw sequencing data counts from the NCBI GEO dataset GSE182196 were retrieved and analyzed (16,17). Differential gene expression (DEG) analysis was conducted across all experimental groups to identify upregulated and downregulated genes. Raw counts were normalized using the DESeq2 package (v1.38.3) with default parameters, and DEGs were selected based on predefined statistical cut-off criteria (adjusted p-value < 0.05 and log2 fold change$| > 1$).

Functional Annotation Using Comparative Toxicogenomics Database (CTD)

DEG data were imported into the CTD for functional analysis (18). Gene ontology (GO) terms were assigned to identify associated biological processes and molecular functions. Additionally, gene-chemical interactions were extracted to explore potential chemical exposures linked to the DEGs. The CTD facilitated the annotation of toxicological effects and pathway associations with relevant chemicals. Only high-confidence interactions (inference score $\geq$ 40) were retained to ensure biological relevance.

File Conversion and Integration

Results were exported as Tab-Separated Value (TSV) files to standardize formatting and enable compatibility across various analysis pipelines. Standardizing data into TSV format involves organizing tabular data in a plain text file where each line represents a record and values within each record are separated by tab characters. Data from different sources were integrated using cosine similarity via early fusion methods, combining gene expression profiles with GO annotations and chemical associations. Normalization procedures ensured consistency in the datasets, making them suitable for machine learning applications.

Data Preprocessing

Datasets were preprocessed by aligning gene expression data with associated statistical, chemical, and GO annotations based on the `GeneID` column. Expression data were normalized to comparable scales using min-max scaling, enabling reliable downstream analysis. Cosine similarity metrics were employed

to evaluate the similarity between vectors representing gene expression profiles, GO terms, and chemical associations.

Cosine Similarity

Cosine similarity was used as a metric to compare gene expression profiles across samples, providing a quantitative measure of the similarity in expression patterns. This method calculates the cosine of the angle between two non-zero vectors, yielding a value between -1 and 1. A threshold of ≥ 0.7 was applied to define significant similarity, as empirically validated in prior studies (16-19).

Least Squares Generative Adversarial Network Architecture

Model Architecture

The LSGAN framework was designed to generate synthetic gene expression data by transforming random noise into structured outputs resembling real data.

Generator

A neural network with an input dimension of 100 (representing random noise) and an output dimension matching the six features in the gene expression data. The architecture includes three fully connected layers (512, 256, and 128 units, respectively) with ReLU activations for non-linearity and a final Tanh activation to scale outputs within the range $(-1.1)(-1.1)$. The choice of layer sizes was based on empirical testing, balancing model complexity and computational efficiency.

Discriminator

A neural network acting as a binary classifier with an input dimension corresponding to the six features of the expression data. It includes three fully connected layers (128, 256, and 512 units, respectively) utilizing Leaky ReLU activations (slope $=0.2$), outputting a single value indicating the likelihood of the input being real or generated.

Both networks were trained using a learning rate of 0.0002 (optimized via grid search), a batch size of 64, and over 200 epochs. The Mean Squared Error (MSE) loss function was used to quantify the differences between real and generated data, iteratively updating the weights of both the generator and discriminator. Early stopping was implemented if validation loss plateaued for 20 consecutive epochs.

Evaluation Metrics

The performance of the LSGAN model was evaluated using the following metrics: Mean Squared Error (MSE): Quantifies the average squared difference between real and generated data distributions; Mean Absolute Error (MAE): Measures the average magnitude of errors without considering their direction; Wasserstein Distance (Earth Mover's Distance): Assesses the overall similarity between real and generated data distributions.

Data Visualization

To validate model outputs, several visualization techniques were employed, including: Density plots and correlation heatmaps to evaluate distributional similarities; Q-Q (quantile-quantile) plots to assess normality and variance; Violin plots to illustrate the distribution and density of generated data in comparison with real data.

RESULTS

The results of this study provide insights into the performance of the LSGAN model in generating synthetic gene expression data that closely resembles real data. Key findings include metrics evaluating distributional similarity, cosine similarity, error values and detailed visualizations that compare real and generated datasets. The analyses demonstrate the efficacy of the LSGAN model in genomic data applications.

Figure 1 displays a volcano plot showing the log2 fold change (log2FC) on the x-axis and the negative log10 of the adjusted p-value (-log10(Padj)) on the y-axis. Each point corresponds to a gene, where red dots represent upregulated genes and blue dots indicate downregulated genes between the two conditions.

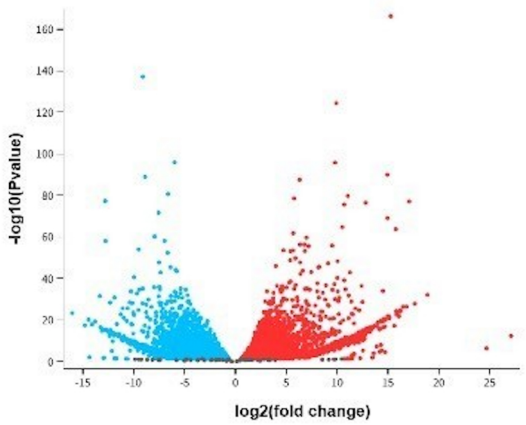

Figure 1. Volcano Plot of Differential Gene Expression.

Cosine Similarity Analysis

Cosine similarity was used to evaluate the similarity of gene expression profiles across samples. The results indicate a high similarity between the generated and real data, suggesting that the LSGAN model effectively captures the directionality of real data vectors. The trained LSGAN model achieves a balance between generating realistic samples and distinguishing between real and synthetic data. Low MSE and MAE values further support the close resemblance between generated and real data distributions. Additionally, low Wasserstein distances, overlapping distributions, and correlation structures validate the model's capability for bioinformatics applications.

Figure 2 shows the dynamics of training losses for the generator and discriminator over 200 epochs. The generator's loss consistently remains lower than the discriminator's, demonstrating that the model is effectively learning the underlying data distribution. The steady decline in both loss curves indicates convergence and model stability.

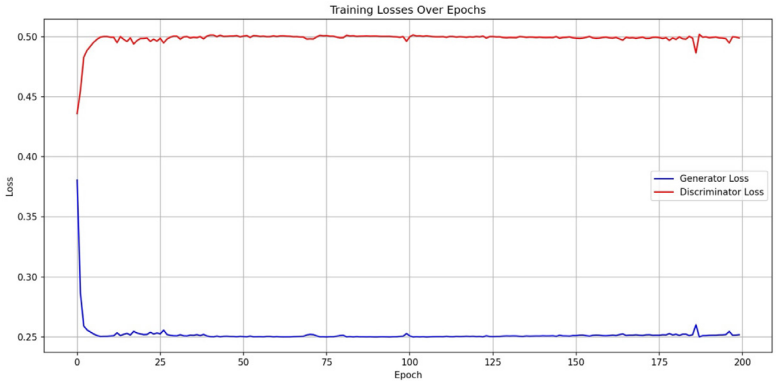

Figure 2. Training Loss Curves for Generator and Discriminator.

Error Metrics and Distributional Analysis

Mean Squared Error: The average squared difference between estimated and actual values is 7.02×10^{-5}, indicating strong model performance.

Mean Absolute Error: The linear score measuring the average difference between predicted and actual values is 0.0021, further confirming the data's resemblance.

Wasserstein Distances: With values ranging from 0.0014 to 0.0023, the distances reflect high alignment between real and generated data distributions. The average Wasserstein distance of 0.00217 underscores the model's effectiveness.

Figure 3 presents a box plot comparing real and generated data distributions across six features. The limited variability and tightly clustered data distributions demonstrate a close resemblance between real and synthetic data, with color legends provided for clarity.

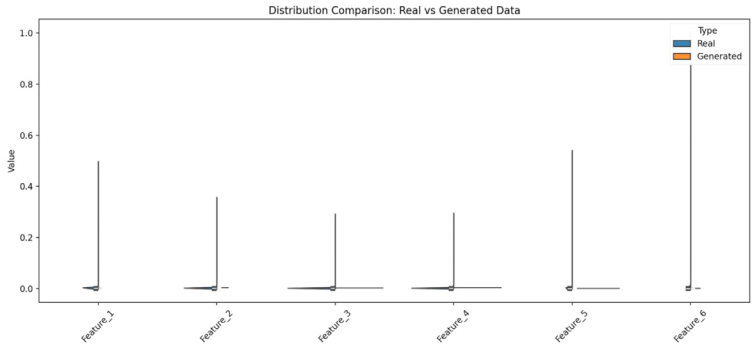

Figure 3. Box Plot Comparing Real and Generated Data Distributions.

Figure 4 contains six Q-Q plots arranged in a 2x3 grid, each corresponding to a different feature. Blue points represent observed values plotted against theoretical quantiles, while red lines denote expected values. The plots reveal tightly clustered data points, indicating low variability and limited diversity in the generated data.

Figure 5 shows six Q-Q plots arranged in a 2x3 grid, comparing real data distributions across features. The upward curves in the plots suggest heavier tails than expected under a normal distribution, highlighting inherent skewness in the real data.

Figure 6 illustrates six density plots arranged in a 2x3 grid, showing the distribution of real data features. The sharp peaks near zero indicate skewness and limited variability, reflecting the dataset's characteristic patterns.

Evaluation of LSGAN Model Performance

The LSGAN model effectively captured real data distribution, as evidenced by density plots, high cosine similarity, and low error metrics. The epoch loss curves (Figure 2) further demonstrate that both generator and discriminator losses stabilized during training, achieving a balance between producing realistic samples and distinguishing between real and synthetic data. These findings indicate the model's robustness and potential utility in generating high-quality synthetic gene expression data for genomics and bioinformatics applications.

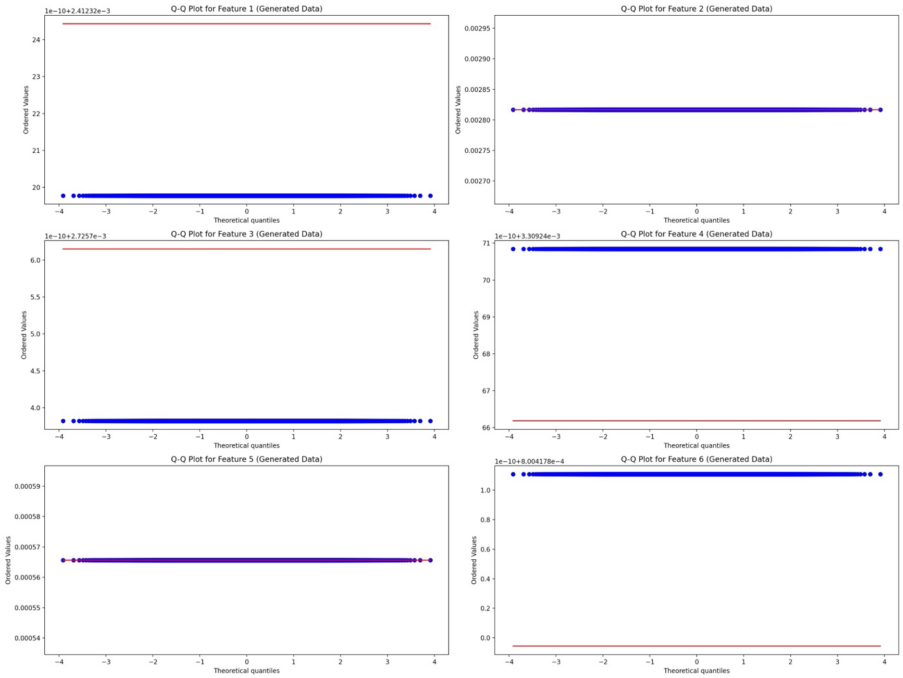

Figure 4. Q-Q Plots of Generated Data for Six Features.

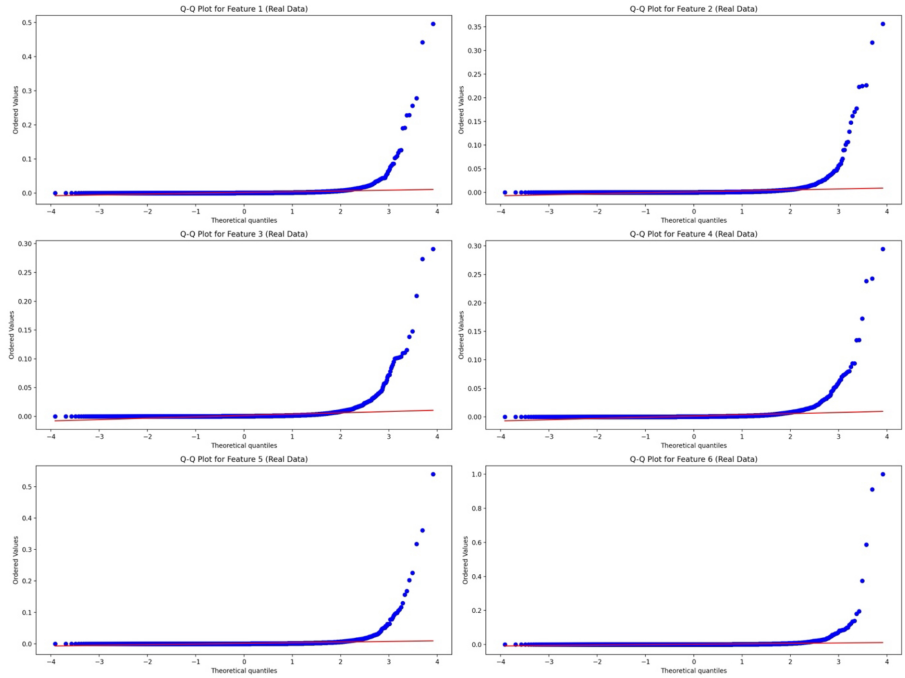

Figure 5. Q-Q Plots of Real Data for Six Features.

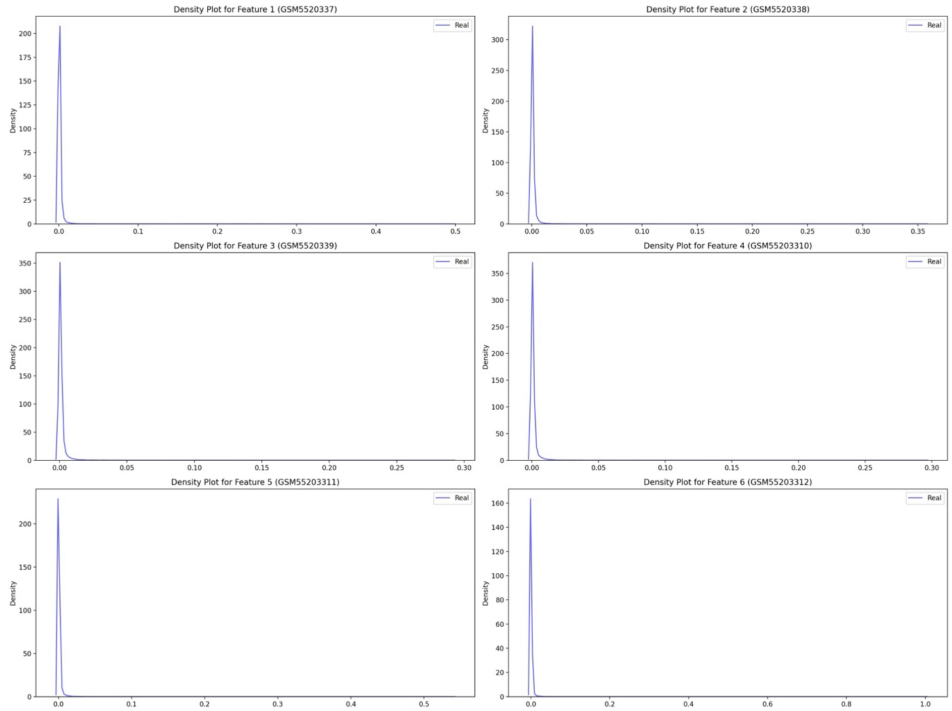

Figure 6. Density Plots of Real Data for Six Features.

DISCUSSION

Keratinization is the process by which keratin accumulates in epithelial cells, providing protective properties to various tissues in the body. In the oral cavity, keratinized tissues (1,7), such as the hard palate and gingiva, act as robust barriers against mechanical injury, microbial invasion, and dehydration. Gingival keratinization occurs within the gingival epithelium, which is subdivided into parakeratinized and orthokeratinized layers. The degree of keratinization in the gingiva varies significantly among individuals due to factors such as genetics, environmental influences, and oral hygiene practices. Keratinized gingival tissue serves multiple roles, including shielding against injury, preventing infections, and facilitating regeneration and healing. Additionally, it impacts aesthetics, periodontal health, and susceptibility to oral diseases. Understanding the processes of oral and gingival keratinization is critical for advancing dental care, with evidence suggesting that effective hygiene practices, regular check-ups, and personalized treatment plans can enhance patient outcomes (4,19).

Generative Adversarial Networks (GANs) (20) have emerged as promising tools for generating transcriptomic data related to keratinization, offering transformative potential for understanding oral and gingival health. By generating high-dimensional data, GANs address the challenges of data scarcity and facilitate model training. They can uncover regulatory mechanisms, simulate biological variability, and test hypotheses related to keratinization processes. However, GAN-generated data may not fully capture the complexity of biological systems, such as non-linear gene-gene interactions or epigenetic influences (21-23). This innovative application of AI in genomics underscores the capacity of computational tools to advance biological sciences (21,22).

The Least Squares Generative Adversarial Network (LSGAN) (9,12,15) is a specialized GAN method designed to generate realistic data by optimizing the labeling system for

discriminators. Its primary goal is to produce synthetic samples that the discriminator incorrectly classifies as real, achieving a balance at the decision boundary between real and fake data. LSGAN employs a least-squares loss function, which ensures smoother gradient transitions and incorporates gradient penalties to generate high-quality data while mitigating issues like vanishing gradients. This study employed LSGAN to generate synthetic omics data with enhanced realism. Despite these advantages, LSGANs may struggle with rare gene expression patterns, potentially biasing downstream analyses (24,25). Similarly, the Wasserstein GAN with Penalty Loss (MDWGAN-GP) model (23-25), leveraging multiple discriminators and linear graph convolutional networks, has demonstrated superior performance in generating high-quality gene expression data. Data augmentation strategies powered by GANs have been shown to improve cancer phenotype classification accuracy, with binary classification reaching 94 % accuracy and tissue classification achieving 70 %. Complex GAN architectures are associated with higher-quality outputs and better augmentation results.

The synthetic gene expression data generated in this study, analyzed using LSGAN, exhibited low mean squared error (MSE) and mean absolute error (MAE) values (7.020×10^5), closely resembling real data distributions. Wasserstein distances (Earth Mover's Distances) revealed comparable distributions between the synthetic and real datasets, further validating the quality of the generated data. However, these metrics do not assess whether synthetic data preserves biologically meaningful gene co-expression networks (24). While LSGAN demonstrates promise in generating data related to gingival keratinization, future directions should include enhancing model architecture and variability, incorporating feature-specific training, exploring multimodal data generation, and validating biological relevance. For instance, integrating single-cell RNA-seq data could improve cellular heterogeneity modeling (25). Longitudinal data simulation and improved evaluation metrics are also critical. Although evaluation metrics such as MSE and MAE offer valuable insights, they do not comprehensively assess data quality. Addressing GAN-specific challenges like mode collapse and training instability, along with improving interpretability and ethical considerations, remains essential for future research.

CONCLUSION

The Least Squares Generative Adversarial Network effectively generates synthetic gene expression data that aligns closely with real data distributions, demonstrating significant potential for applications requiring realistic synthetic data. Synthetic datasets related to gingival keratinization provide valuable insights into biological processes and variability. Advanced modeling techniques enable the creation of datasets that accurately reflect patterns observed in clinical studies, supported by robust statistical metrics such as low MSE and high correlation coefficients.

This study highlights the potential of synthetic data to simulate variability in keratinization levels, offering a nuanced understanding of factors influencing gingival health. Such datasets can be instrumental in hypothesis testing, machine learning model training, and simulating rare or underrepresented conditions. The findings underscore the value of innovative computational approaches in oral health research and suggest future research avenues, including the incorporation of longitudinal data, biological variability, and enhanced modeling frameworks, to expand the utility of synthetic data in dentistry and genomics further.

Conflict of interest

No potential conflict of interest relevant to this article was reported

Acknowledgements

None

REFERENCES

1. Thoma DS, Gil A, Hämmerle CHF, Jung RE. Management and prevention of soft tissue complications in implant dentistry. Periodontol 2000. 2022;88:116-129.

2. Barootchi S, Tavelli L, Zucchelli G, Giannobile W V, Wang HL. Gingival phenotype modification therapies on natural teeth: A network meta-analysis. J Periodontol. 2020;91:1386-1399.
3. Burra Anand D, Ramamurthy J, Kannan B, Jayaseelan VP, Arumugam P. N6-methyladenosine-mediated overexpression of TREM-1 is associated with periodontal disease. Odontology. 2025;113(2):834-843.
4. Thoma DS, Naenni N, Figuero E, Hämmerle CHF, Schwarz F, Jung RE, et al. Effects of soft tissue augmentation procedures on peri-implant health or disease: A systematic review and meta-analysis. Clin Oral Implants Res. 2018;29(Suppl 1):32-49.
5. Ramamurthy J, Mg V. Comparison of effect of hiora mouthwash versus chlorhexidine mouthwash in gingivitis patients: A clinical trial. Asian J Pharmaceut Clin Res. 2018;11:84-88.
6. Deepika BA. Comparative clinical data for gingivitis treatment using gels from Ocimum sanctum (Tulsi) and chlorhexidine (CHX). Bioinformation. 2021;17:1091-1098.
7. Malpartida-Carrillo V, Tinedo-Lopez PL, Guerrero ME, Amaya-Pajares SP, Özcan M, Rösing CK. Periodontal phenotype: A review of historical and current classifications evaluating different methods and characteristics. J Esthet Restor Dent. 2021;33:432-445.
8. Barak O, Lovelace T, Piekos S, Chu T, Cao Z, Sadovsky E, et al. Integrated unbiased multiomics defines disease-independent placental clusters in common obstetrical syndromes. BMC Med. 2023;21:349.
9. Touati R, Kadoury S. A least square generative network based on invariant contrastive feature pair learning for multimodal MR image synthesis. Int J Comput Assist Radiol Surg. 2023;18:971-979.
10. Dong J, Fu T, Lin Y, Deng Q, Fan J, Song H, et al. Hole-filling based on content loss indexed 3D partial convolution network for freehand ultrasound reconstruction. Comput Methods Programs Biomed. 2021;211:106421.
11. Veiner J, Alajaji F, Gharesifard B. A Unifying Generator Loss Function for Generative Adversarial Networks. Entropy (Basel). 2024;26:290.
12. Aguirre N, Cymberknop LJ, Grall-Maës E, Ipar E, Armentano RL. Central Arterial Dynamic Evaluation from Peripheral Blood Pressure Waveforms Using CycleGAN: An In Silico Approach. Sensors (Basel). 2023;23:1559.
13. Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP. On the Effectiveness of Least Squares Generative Adversarial Networks. IEEE Trans Pattern Anal Mach Intell. 2019;41:2947-2960.
14. Qu N, Chen D, Ma B, Zhang L, Wang Q, Wang Y, et al. Integrated proteogenomic and metabolomic characterization of papillary thyroid cancer with different recurrence risks. Nat Commun. 2024;15:3175.
15. Bhatia H, Paul W, Alajaji F, Gharesifard B, Burlina P. Least kth-Order and Rényi Generative Adversarial Networks. Neural Comput. 2021;33:2473-2510.
16. Clough E, Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. Nucleic Acids Res. 2024;52:D138-44.
17. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: Archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41(Database issue):D991-5.
18. Davis AP, Wiegers TC, Johnson RJ, Sciaky D, Wiegers J, Mattingly CJ. Comparative Toxicogenomics Database (CTD): update 2023. Nucleic Acids Res. 2023;51:D1257-262.
19. Ren S, Li J, Dorado J, Sierra A, González-Díaz H, Duardo A, et al. From molecular mechanisms of prostate cancer to translational applications: Based on multi-omics fusion analysis and intelligent medicine. Health Inf Sci Syst. 2024;12:6.
20. Mohsen F, Ali H, El Hajj N, Shah Z. Artificial intelligence-based methods for fusion of electronic health records and imaging data. Sci Rep. 2022;12:17981.
21. Lacan A, Sebag M, Hanczar B. GAN-based data augmentation for transcriptomics: Survey and comparative assessment. Bioinformatics. 2023;39(39 Suppl 1):i111-i120.
22. Li R, Wu J, Li G, Liu J, Xuan J, Zhu Q. Mdwgan-gp: Data augmentation for gene expression data based on multiple discriminator WGAN-GP. BMC Bioinformatics. 2023;24:427.
23. Wang TH, Lee CY, Lee TY, Huang HD, Hsu JBK, Chang TH. Biomarker Identification through Multiomics data analysis of prostate cancer prognostication using a deep learning model and similarity network fusion. Cancers (Basel). 2021;13:2528.
24. Waters MR, Inkman M, Jayachandran K, Kowalchuk RM, Robinson C, Schwarz JK, et al. GAiN: An integrative tool utilizing generative adversarial neural networks for augmented gene expression analysis. Patterns (N Y). 2024;5:100910.
25. Ding DY, Li S, Narasimhan B, Tibshirani R. Cooperative learning for multiview analysis. Proc Natl Acad Sci U S A. 2022;119:e2202113119.