

Pre-Trained Language Models Based Sequence Prediction of Wnt-Sclerostin Protein Sequences in Alveolar Bone Formation

Pradeep kumar Yadalam¹, Ramya Ramadoss², Pradeep kumar R^{3,*} and Jishnu Krishna Kumar³

¹Department of Periodontics, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, Tamil Nadu, India.

²Department of Oral Biology, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, Tamil Nadu, India.

³Department of Public Health Dentistry, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, Tamil Nadu, India.

Corresponding author: Pradeep kumar R (e-mail: pradeepkumarr.sdc@saveetha.com).

©2023 the Author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)

Abstract Background and Introduction: Osteocytes, the most numerous bone cells, create sclerostin. The sclerostin protein sequence predictive model helps create novel medications and produce alveolar bone in periodontitis and other oral bone illnesses, including osteoporosis. Neural networks examine protein variants for protein engineering and predict their structure and function impacts. Proteins with improved function and stability have been engineered using LLMs and CNNs. Sequence-based models, especially protein LLMs, predict variation effects, fitness, post-translational modifications, biophysical properties, and protein structure. CNNs trained on structural data also improve enzyme function. It is unknown if these models differ or forecast similarly. This study seeks Pre-trained language models to predict Wnt-Sclerostin Protein sequences in alveolar bone formation. **Methods:** Using UniProt ID, sclerostin and related proteins (Q9BQB4, Q9BQB4-1, Q9BQB4-2, Q6X4U4, O75197) were identified and quality-checked. Deepbio analyzed FASTA sequences. Deep Bio is a one-stop web service allowing academics to build any biological deep-learning architecture. DeepBIO used deep learning to improve and visualize biological sequencing data. LLM BASED Reformer, AAPNP, TEXTRGNN, VDCNN, and RNN_CNN split sequence-based datasets into test and training. We randomly partitioned each dataset into 1000 training and 200 testing sets to change hyperparameters and measure performance. **Results:** Reformer, AAPNP, TEXTRGNN, VDCNN, RNN CNN exhibit 93, 64, 51, 91, and 64 percent accuracy. **Conclusion:** Protein sequence-based massive language models are growing, and R&D is solving complicated challenges.

Key Words LLM, Natural Language Processing, Sclerostin, Alveolar Bone Formation, Periodontitis, Dental, Reformer, AAPNP, TEXTRGNN, VDCNN, RNN_CNN

1. Introduction

In recent years, there has been a major evolution in our understanding of bone health and regulation [1], [2]. The Wnt signaling pathway is one of the major mechanisms controlling bone homeostasis; sclerostin is a critical regulator in this system. The purpose of this article is to present a thorough summary of the functions of sclerostin and Wnt signaling in bone health. The Wnt signaling system tightly regulates bone resorption and formation. It comprises the β -catenin-dependent canonical Wnt pathway and the β -catenin-independent non-canonical Wnt pathway [3], [4]. While osteoclastogenesis, bone resorption, and bone remodeling are involved in the non-canonical pathway, the canonical pathway controls osteoblast proliferation, differentiation, and survival [5].

The most prevalent cells in bone tissue [6]–[8], osteocytes, are the main producers of the protein sclerostin. Attaching itself to the receptors LRP5/6 and preventing β -catenin from being activated is a negative regulator of the Wnt signaling cascade. Sclerostin suppresses osteoblast proliferation and activity through this mechanism, which ultimately results in lower bone mass and inhibits the creation of new bone [9]. The Wnt signaling pathway is activated, favorably controlling osteoblastogenesis and encouraging new bone growth. When Wnt ligands, like Wnt1, Wnt3a, and Wnt10b, engage with LRP5/6 receptors, β -catenin [10] in the nucleus-activator is stabilized and accumulates, and target genes important in osteoblast differentiation, and function are expressed [11], [12]. On the other hand, Sclerostin binds to LRP5/6 and inhibits this process by stopping β -catenin from

activating [13], [14].

Numerous bone disorders have been linked to the deregulation of sclerostin and abnormalities in the Wnt signaling pathway. For example, osteoporosis, osteogenesis imperfecta, and juvenile idiopathic arthritis have been linked to elevated sclerostin levels and decreased Wnt signaling activity [15]–[18]. These abnormalities weaken bones and raise fracture risk by decreasing bone production and increasing bone resorption. Therapeutic strategies that target sclerostin have been investigated because of its critical function in maintaining bone homeostasis.

In periodontitis [19] and other problems involving the mouth’s bone, such as osteoporosis, the predictive model of the sclerostin protein sequences helps create healthy alveolar bone and can be useful in designing novel medications. Neural networks and other machine learning models are increasingly being utilized to investigate and forecast the impacts of protein variations on structure and function in protein engineering [18]–[20]. Convolutional neural networks (CNNs) and large language models (LLMs) [21]–[23] have effectively created proteins with improved stability and function. Protein LLMs [24], particularly sequence-based models, have successfully predicted protein structure, post-translational modifications, variation effects, and biophysical characteristics. CNNs trained on structural data have also successfully increased enzyme function activity. It’s unclear whether these models are fundamentally different or produce comparable forecasts. This work aims to predict Wnt-Sclerostin protein sequences in alveolar bone formation using pre-trained language models.

2. Methods

The following sclerostin and related proteins Q9BQB4, Q9BQB4-1, Q9BQB4-2, Q6X4U4, and O75197 were downloaded using UniProt id, and their sequences were recognized and quality-checked. The Deepbio tool was used for FASTA sequences.

Deep Bio is a one-stop shop for researchers wishing to create a deep-learning architecture for any biological subject. DeepBIO used deep learning techniques to evaluate, improve, and visualize biological sequencing data. Sequence-based datasets were divided into training and test sets by Deep Bio. We randomly divided each dataset into 1000 training and 200 testing sets to modify hyperparameters and assess performance.

Large language models and other algorithms for sequence prediction used were Reformer, AAPNP, TEXTRGNN, VD-CNN, RNN_CNN (see Table 1).

Reformer

The Reformer is a natural language processing AI model. In 2019, Google researchers published "Reformer: The Efficient Transformer," introducing it. The Reformer model uses the Transformer architecture popularised by BERT and GPT (Generative Pre-trained Transformer). Reformer efficiency in managing long-range dependencies is a significant

Cuda:	TRUE2	TRUE3
Seed:	43	43
num_workers:	4	4
num_class:	2	2
Kmer:	3	3
save_figure_type:	png	png
Mode:	train-test	train-test
Type:	prot	prot
Model:	VDCNN	RNN_CNN
datatype:	userprovide	userprovide
interval_log:	10	10
interval_valid:	1	1
interval_test:	1	1
Epoch:	50	50
optimizer:	Adam	Adam
loss_func:	CE	CE
batch_size:	32	32
LR:	0.0001	0.0001
Reg:	0.0025	0.0025
Gamma:	2	2
Alpha:	0.25	0.25
max_len:	52	52
dim_embedding:	32	32
minimode:	modelCompare	modelCompare
if_use_FL:	0	0
if_data_aug:	1	1
if_data_enh:	0	0
CDHit:	['1']	['1']

Table 1: Model parameters for hyper tuning and epoch iterations

contribution. Compared to classic Transformers. It reduces long sequence attention time and memory complexity with "Locality-Sensitive Hashing" (LSH). Reversible residual layers make Reformer training memory-efficient. This is crucial for long sequences, as regular Transformers often struggle with memory.

AAPNP

AAPNP, which stands for Approximation of Personalized Propagation of Neural Prediction, is a novel approach for semi-supervised learning on graphs. It combines the strengths of two powerful techniques: Personalized PageRank, which uses Google’s PageRank to rank network nodes by neighbor importance, and a "seed" node. This means a node’s value depends on its connections and neighbors’ connections, considering a specific focus. Neural Networks: These powerful models can learn complex relationships and patterns from data, often achieving impressive results in many tasks.

AAPNP Leverages these Two Techniques in a Two-step Process

Predict: Features are used by a neural network to forecast each node. First, guesses gather local information around each node. Propagate: Then, Personalized PageRank is ad-

justed to "spread" these predictions around the graph, taking into account nearby nodes and the focus point. This propagation stage refines initial predictions and incorporates global context by sharing information between nodes.

AAPNP's fundamental benefit is its ability to use information from a vast, configurable neighborhood around each node while preserving computational efficiency and a minimum number of parameters. This makes it useful for semi-supervised classification, where you have little labeled data but a big network of unlabeled data.

TextRGNN

A new graph neural network-based text classification architecture is Residual Graph Neural Networks (GNNs). It was introduced in a December 2021 research report and has performed well in many datasets.

Here is a breakdown of TextRGNN's key features;

- 1) **Residual Connections:** TextRGNN, unlike shallow GNN models typically using two convolutional layers, utilizes residual connections for text classification. Information flows directly from earlier to later stages due to these linkages skipping layers. TextRGNN can capture text data's short- and long-range dependencies, enhancing accuracy.
- 2) **Wider Receptive Field:** TextRGNN's residual connections enable each node to obtain information from a greater neighborhood of other nodes. This helps the model grasp sentence or document context and word-phrase relationships.
- 3) **Over-Smoothing Suppression:** GNNs may experience over-smoothing when node features become similar after multiple message-passing steps. This makes it harder for the model to discriminate text portions. TextRGNN prevents features from homogenizing, preserving their discriminative value.

TextRGNN uses a probabilistic language model (PLM) to initialize graph node embeddings to improve semantic information capture. This uses the PLM's word relationships and syntax to enrich the GNN's message-passing process.

VDCNN

VDCNN, or Very Deep Convolutional Neural Network, is an architecture designed for text classification. It uses small convolutions and pooling operations at the character level to obtain outstanding results. VDCNN breakdown:

Architectural Modularity

Choose from 9, 17, 29, and 49 layers to suit dataset sizes and complexity.

Character-level Processing

Works directly with text characters to capture fine-grained classification information.

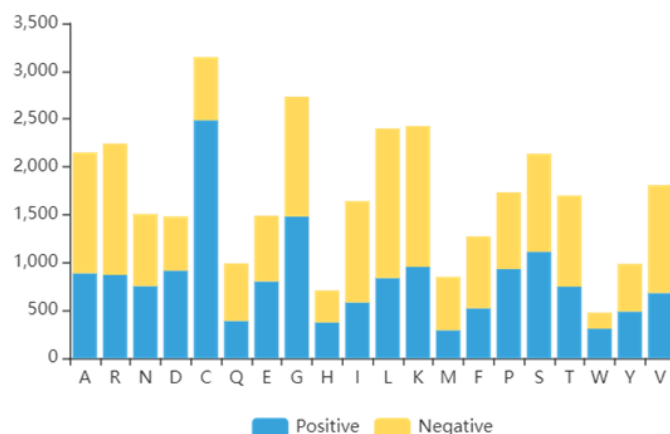


Figure 1: The data's positive (train) and negative (test)

Pooling and Small Convolutions

Uses 3 or 5 filter sizes in convolutional layers with max pooling to reduce dimensionality, reduce the computational cost, and improve noise resistance.

Multiple Nonlinear Activations

Non-linearity from ReLU activations throughout the network improves feature extraction and representation. Global average pooling aggregates information from all feature maps in the final layer to achieve efficient categorization, collecting global context.

RNN_CNN

RNN-CNN, or Recurrent Neural Network-Convolutional Neural Network, is a powerful deep learning architecture that uses its capabilities to solve complicated problems, especially in NLP and computer vision. Text or video input data is preprocessed before being fed into the network. This may entail text tokenization or video frame extraction. **CNN Extraction of Features**

The CNN uses convolutional layers to extract local information from preprocessed input. These attributes capture Edges, textures, shapes in photographs, word frequencies, and grammatical structures in text. In RNN Sequence Modeling, retrieved features are input into the RNN. The RNN successively processes features and stores their associations in its memory. Using processed features and internal memory, the RNN generates the required output. This could be a classification label, caption, or sequence prediction.

3. Results

The study reveals that prediction accuracy varies across protein structures, with LLMs yielding good accuracy. This reflects the bias/variance dilemma in machine learning, where convolution layers have an inductive bias for spatial data.

The "sensitivity" of Reformer, AAPNP, TEXTRGNN, VDCNN, and RNN CNN is 0.88, 0.48, 0.87, and 0.66 for TP



Figure 2: The accuracy of the multiple models

Model Name	ACC	Sensitivity	Specificity	AUC
Reformer	0.885	0.88	0.89	0.936
APPNP	0.56	0.48	0.64	0.614
TextRGNN	0.5	0	1	0.511
VDCNN	0.875	0.87	0.88	0.914
RNN_CNN	0.625	0.66	0.59	0.649

Table 2: The accuracy of Reformer, AAPNP, TEXTRGNN, VDCNN, and RNN_CNN which show 93%, 64%,51%, 91% AND 64 %

/(TP + FN). The model’s ability to correctly identify negative cases is known as its specificity or genuine negative rate. The results show that the specificities of Reformer, AAPNP, TEXTRGNN, VDCNN, and RNN CNN are $TN / (TN + FP)$ -0.89, 0.64,1, 0.88, and 0.59, respectively.

Roc Curve

The Receiver Operating Characteristic demonstrates how categorization thresholds affect a model’s true positive rate (sensitivity) and false positive rate (1 - specificity) (ROC). The ROC curve in the upper left corner of the plot demonstrates that Reformer’s VDCNN is accurate, whereas AAPNP, TEXTRGNN, and RNN CNN are moderate.

Precision Recall Curve

The precision-recall curve shows binary classifiers with varying probability thresholds’ recall-precision trade-off (PRC). Recall is the percentage of accurately expected positives, whereas precision is the percentage of positive predictions. This model’s unequal class performance is shown. AUC-PR is a common statistic for classifier performance. Reformer

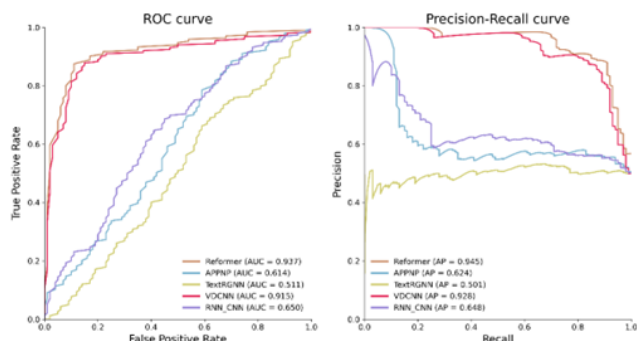


Figure 3: The ROC and precision-recall curve of the plot

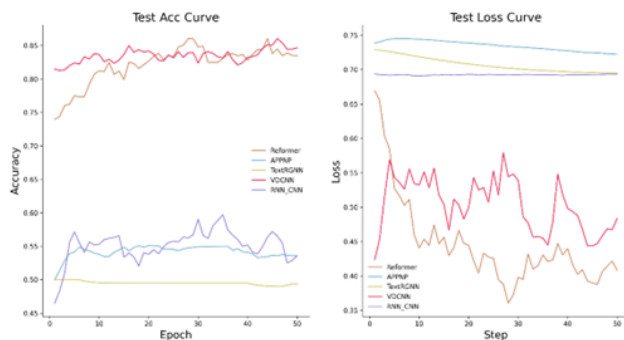


Figure 4: The epoch plot of all models

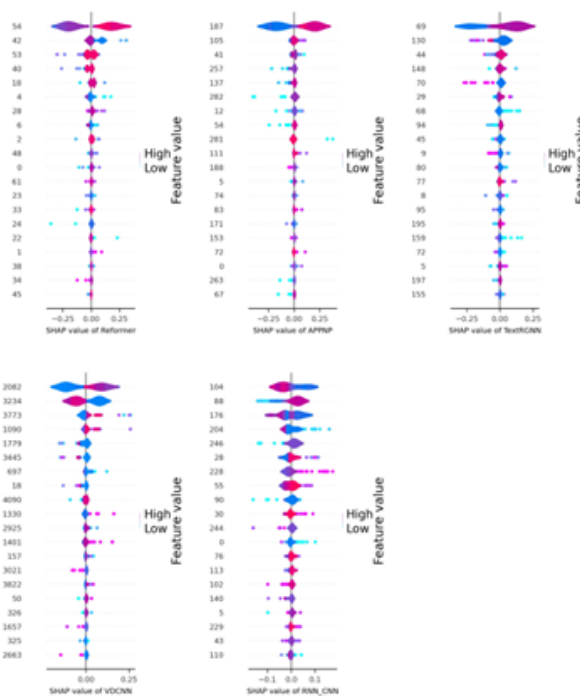


Figure 5: The shap values prediction

VDCNN model performance improves with higher AUC-PR values.

An epoch plot graphs machine learning model accuracy and loss over training. It detects overfitting and other model flaws well. Etoch plots display the number of epochs or iterations the model was trained on on the x-axis. Model accuracy or loss is represented on the y-axis. The loss shows how well the model predicts an input’s output. Accuracy measures the model’s prediction accuracy.

SHAP Values

Machine-learning models calculate each feature’s prediction value. All potential feature combinations and their relative contributions to a prediction when combined with a subset of features are analyzed to compute them. SHAP red is positive

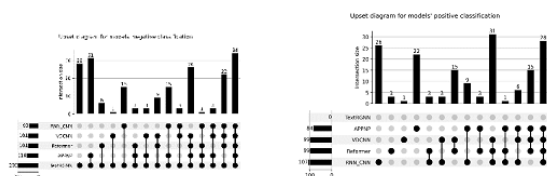


Figure 6: An upset plot of prediction

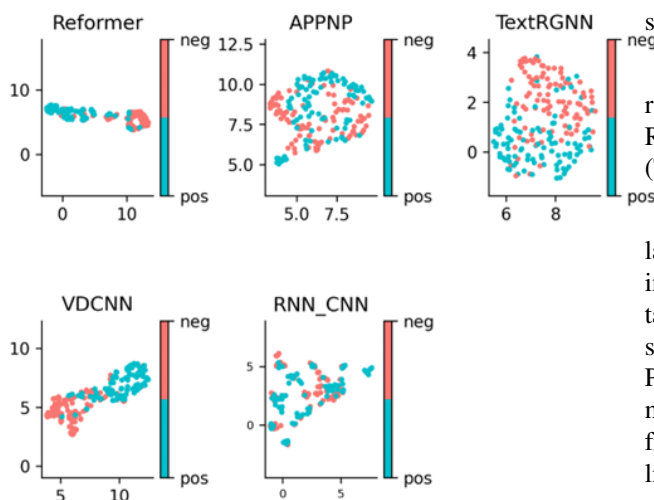


Figure 7: Umap plot of data

when a feature improves prediction. Negative SHAP blue features are less predictive.

Upset Plot

Comparison of intersection diameters shows the group frequency of common elements. Larger junctions indicate more group overlap than smaller crossings. Vertical UpSet plots show crossings as rows and sets as matrix columns. Every row has filled intersection cells showing row relationships.

UMAP shows clustering patterns in a weighted graph using high-dimensional data, with edge strength indicating "near" points. Projecting this graph reduces its size. Data demonstrate algorithm clustering. UMAP embeds high-dimensional data in low-dimensional space using nonlinear dimensionality reduction. It expects high-dimensional data points to reside near low-dimensional space.

4. Discussion

Understanding the mechanisms and regulation of sclerostin protein sequences can provide insights into developing therapies for various bone-related disorders, including those affecting the alveolar bone. Protein sequence prediction using Large Language Models (LLMs) [24], [25] is a rapidly emerging field with exciting protein engineering and drug discovery possibilities. LLMs trained on massive protein databases can learn the underlying patterns and rules of

protein sequences. This allows them to generate new sequences with desired properties, like increased stability, specific binding affinities, or even new functionalities. LLMs can statistically predict the most likely missing amino acids when faced with incomplete protein sequences based on the surrounding context. This can be crucial for structural modeling and understanding protein function [26], [27]. Analyzing the relationships between different protein sequences is key to understanding their evolution and function. LLMs can help uncover these relationships by learning the subtle changes in sequences that translate to functional differences.

Sclerostin protein sequences prediction shows the accuracy of Reformer, AAPNP, TEXTRGNN, VDCNN, and RNN_CNN, which show 93%, 64%, 51%, 91% AND 64% (Table 2, Figures 1-7).

Various Models Like ProteinBERT [28], [29] is a deep language model specifically designed for proteins, combining language modeling with Gene Ontology (GO) annotation prediction. It offers efficient and flexible biological sequence performance with local and global representations. ProteinBERT [30], [31] achieves near-state-of-the-art performance on various protein properties, making it an efficient framework for rapidly training protein predictors, even with limited labeled data. Transformer-based architectures have revolutionized protein design, enabling the creation of personalized proteins for various applications. ProtGPT2 [11], [32], a language model trained on protein space, generates de novo protein sequences based on natural principles, displaying natural amino acid propensities and distantly related to natural sequences, thereby exploring unexplored regions of protein space. Sclerostin protein sequence prediction is useful for designing novel drugs and increasing alveolar bone formation [14], [16], [17], [33]–[35]. Antisclerostin monoclonal antibodies have shown significant osteoanabolic effects in animal studies, including increased bone mineral density in mice and reversing bone loss in ovariectomized rats. Antisclerostin therapy improved nonhuman primates' fracture healing, alveolar bone repair, and callus density. Sclerostin is a protein that plays a crucial role in alveolar bone formation. Alveolar bone [36] refers to the bone surrounding the teeth and helps provide support and stability. Sclerostin is primarily produced and secreted by osteocytes, mature bone cells within the bone tissue. In summary, sclerostin is vital in alveolar bone formation by inhibiting excessive alveolar bone formation and maintaining a balanced bone remodeling process. This AI model will help predict difficult sequence information and aid in novel protein drug designs targeting sclerostin.

5. Conclusion

This predictive AI model will solve complex sclerostin protein sequences and help design novel drugs to target sclerostin for alveolar bone formation.

Conflict of Interest

The authors declare no conflict of interests. All authors read and approved final version of the paper.

Authors Contribution

All authors contributed equally in this paper.

References

- [1] Galli, C., Passeri, G., & Macaluso, G. M. (2010). Osteocytes and WNT: the mechanical control of bone formation. *Journal of Dental Research*, 89(4), 331-343.
- [2] Chen, G., Deng, C., & Li, Y. P. (2012). TGF- β and BMP signaling in osteoblast differentiation and bone formation. *International Journal Of Biological Sciences*, 8(2), 272-288.
- [3] Tan, Z., Ding, N., Lu, H., Kessler, J. A., & Kan, L. (2019). Wnt signaling in physiological and pathological bone formation. *Histology and Histopathology*, 34(4), 303-312.
- [4] Clevers, H., & Nusse, R. (2012). Wnt/ β -catenin signaling and disease. *Cell*, 149(6), 1192-1205.
- [5] Rossini, M., Gatti, D., & Adami, S. (2013). Involvement of WNT/ β -catenin signaling in the treatment of osteoporosis. *Calcified Tissue International*, 93, 121-132.
- [6] Devi, S., & Duraisamy, R. (2020). Crestal Bone Loss in Implants Post-loading and Its Association with Age, Gender, and Implant Site: A Retrospective Study. *Journal of Long-Term Effects of Medical Implants*, 30(3), 205-211.
- [7] Menon, A., Kareem, N., & Vadivel, J. K. (2021). Evaluation Of Periodontal Flap Procedures Done Using Guided Tissue Regeneration (Gtr) Versus Guided Tissue Regeneration (Gtr) With Bone Graft. *International Journal of Dentistry and Oral Science*, 8(8), 4065-4069.
- [8] Shah, P., & Thangavelu, L. (2021). Knowledge Of Osteoporosis Among Students Of Private Dental College In Chennai. *Int J Dentistry Oral Sci*, 8(03), 2045-2047.
- [9] Boudin, E., Fijalkowski, I., Piters, E., & Van Hul, W. (2013, October). The role of extracellular modulators of canonical Wnt signaling in bone metabolism and diseases. In *Seminars in Arthritis and Rheumatism* (Vol. 43, No. 2, pp. 220-240). WB Saunders.
- [10] JJoiner, D. M., Ke, J., Zhong, Z., Xu, H. E., & Williams, B. O. (2013). LRP5 and LRP6 in development and disease. *Trends in Endocrinology & Metabolism*, 24(1), 31-39.
- [11] Ferruz, N., Schmidt, S., & HÅ¶cker, B. (2022). ProtGPT2 is a deep unsupervised language model for protein design. *Nature Communications*, 13(1), 4348.
- [12] Arvind, T. P., Jain, R. K., Nagi, R., & Tiwari, A. (2022). Evaluation of alveolar bone microstructure around impacted maxillary canines using fractal analysis in dravidian population: a retrospective CBCT study. *The Journal of Contemporary Dental Practice*, 23(6), 593-600.
- [13] Maeda, K., Kobayashi, Y., Koide, M., Uehara, S., Okamoto, M., Ishihara, A., ... & Marumo, K. (2019). The regulation of bone metabolism and disorders by Wnt signaling. *International Journal of Molecular Sciences*, 20(22), 5525.
- [14] Marini, F., Giusti, F., Palmi, G., & Brandi, M. L. (2023). Role of Wnt signaling and sclerostin in bone and as therapeutic targets in skeletal disorders. *Osteoporosis International*, 34(2), 213-238.
- [15] Matsumoto, T. (2015). Pharmacology of bone anabolic agents. *Nihon rinsho. Japanese Journal of Clinical Medicine*, 73(10), 1639-1644.
- [16] Ahn, V. E., Chu, M. L. H., Choi, H. J., Tran, D., Abo, A., & Weis, W. I. (2011). Structural basis of Wnt signaling inhibition by Dickkopf binding to LRP5/6. *Developmental cell*, 21(5), 862-873.
- [17] Ring, L., Neth, P., Weber, C., Steffens, S., & Faussner, A. (2014). β -Catenin-dependent pathway activation by both promiscuous "canonical" WNT3a-, and specific "noncanonical" WNT4- and WNT5a-FZD receptor combinations with strong differences in LRP5 and LRP6 dependency. *Cellular Signalling*, 26(2), 260-267.
- [18] Bao, J., Zheng, J. J., & Wu, D. (2012). The structural basis of DKK-mediated inhibition of Wnt/LRP signaling. *Science Signaling*, 5(224), pe22-pe22.
- [19] Rajasekar, A., & Varghese, S. S. (2022). Microbiological Profile in Periodontitis and Peri-Implantitis: A Systematic Review. *Journal of Long-Term Effects of Medical Implants*, 32(4), 83-94.
- [20] Hua, Y., Yang, Y., Li, Q., He, X., Zhu, W., Wang, J., & Gan, X. (2018). Oligomerization of Frizzled and LRP5/6 protein initiates intracellular signaling for the canonical WNT/ β -catenin pathway. *Journal of Biological Chemistry*, 293(51), 19710-19724.
- [21] Schapke, J., Tavares, A., & Recamonde-Mendoza, M. (2021). Epgat: gene essentiality prediction with graph attention networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(3), 1615-1626.
- [22] Danishuddin, Kumar, V., Lee, G., Yoo, J., Ro, H. S., & Lee, K. W. (2022). An attention mechanism-based LSTM network for cancer kinase activity prediction. SAR and QSAR in Environmental Research, 33(8), 631-647.
- [23] Huang, G., Luo, W., Zhang, G., Zheng, P., Yao, Y., Lyu, J., ... & Wei, D. Q. (2022). Enhancer-LSTMAtt: a Bi-LSTM and attention-based deep learning method for enhancer recognition. *Biomolecules*, 12(7), 995.
- [24] Perez, R., Li, X., Giannakoulis, S., & Petersson, E. J. (2023). AggBERT: Best in Class Prediction of Hexapeptide Amyloidogenesis with a Semi-Supervised ProtBERT Model. *Journal of Chemical Information and Modeling*, 63(18), 5727-5733.
- [25] Guntuboina, C., Das, A., Mollaei, P., Kim, S., & Barati Farimani, A. (2023). Peptidebert: A language model based on transformers for peptide property prediction. *The Journal of Physical Chemistry Letters*, 14, 10427-10434.
- [26] Ghazikhani, H., & Butler, G. (2023). Enhanced identification of membrane transport proteins: a hybrid approach combining ProtBERT-BFD and convolutional neural networks. *Journal of Integrative Bioinformatics*, 20(2), 20220055.
- [27] Geffen, Y., Ofra, Y., & Unger, R. (2022). DistilProtBert: a distilled protein language model used to distinguish between real proteins and their randomly shuffled counterparts. *Bioinformatics*, 38(Supplement_2), ii95-ii98.
- [28] Yadalam, P. K., Trivedi, S. S., Krishnamurthi, I., Aneundi, R. V., Mathew, A., Al Shayeb, M., ... & Rajkumar, R. (2022). Machine Learning Predicts Patient Tangible Outcomes After Dental Implant Surgery. *IEEE Access*, 10, 131481-131488.
- [29] Kumar, V. S., Kumar, P. R., Yadalam, P. K., Aneundi, R. V., Shrivastava, D., Alfurhud, A. A., ... & Srivastava, K. C. (2023). Machine Learning in the Detection of Dental Cyst, Tumor, and Abscess Lesions. *BMC Oral Health*, 23(1), 833.
- [30] Marks, D. S., Hopf, T. A., & Sander, C. (2012). Protein structure prediction from sequence variation. *Nature biotechnology*, 30(11), 1072-1080.
- [31] Lupo, U., Sgarbossa, D., & Bitbol, A. F. (2022). Protein language models trained on multiple sequence alignments learn phylogenetic relationships. *Nature Communications*, 13(1), 6298.
- [32] Brandes, N., Ofer, D., Peleg, Y., Rappoport, N., & Linial, M. (2022). ProteinBERT: a universal deep-learning model of protein sequence and function. *Bioinformatics*, 38(8), 2102-2110.
- [33] Hua, Y., Yang, Y., Li, Q., He, X., Zhu, W., Wang, J., & Gan, X. (2018). Oligomerization of Frizzled and LRP5/6 protein initiates intracellular signaling for the canonical WNT/ β -catenin pathway. *Journal of Biological Chemistry*, 293(51), 19710-19724.
- [34] Bagherian, M., Sabeti, E., Wang, K., Sartor, M. A., Nikolovska-Coleska, Z., & Najarian, K. (2021). Machine learning approaches and databases for prediction of drug-target interaction: a survey paper. *Briefings in Bioinformatics*, 22(1), 247-269.
- [35] Liu, M., Cai, R., Hu, Y., Matheny, M. E., Sun, J., Hu, J., & Xu, H. (2014). Determining molecular predictors of adverse drug reactions with causality analysis based on structure learning. *Journal of the American Medical Informatics Association*, 21(2), 245-251.
- [36] Cheng, F., & Zhao, Z. (2014). Machine learning-based prediction of drug interactions by integrating drug phenotypic, therapeutic, chemical, and genomic properties. *Journal of the American Medical Informatics Association*, 21(e2), e278-e286.