Grenze International Journal of Engineering and Technology, June Issue

GeneGuard AI: Real-Time Pathogenicity Prediction of Single Nucleotide Variants using Genomic Foundation Models and Transformer Fine-Tuning

Gollapalli Abhiram[1], G. Ravi[2], B. Vamsi[3] and Krishnaprasad T.R[4]
[1,3]Dept. of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, India – 690525
Email: gollapalliabhiram30@gmail.com, vamsinarendrabandela@gmail.com
[2,4]Dept. of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, India – 690525
Email: gancheriravistar45@gmail.com, krishnaprasadtr@am.amrita.edu

***Abstract*— Single nucleotide variants (SNVs) are among the most clinically significant mutation types in humans, linked to inherited diseases, rare Mendelian conditions, and cancer predisposition. Functional evaluation of the millions of novel SNPs discovered annually is prohibitively expensive, while existing machine learning approaches depend on annotated data that may not generalize to novel variants. We present GeneGuard AI, an integrated genomics platform that combines the theoretical foundations of Evo2 (trained on 9.3 trillion base pairs) with a fine-tuned Nucleotide Transformer (500M parameters). From ~350,000 ClinVar variants, a rigorous filtering pipeline retains 345,909 high-confidence SNVs. Chromosome-stratified splitting (chromosomes 1–18 for training; 19–22, X, Y held out) prevents genomic data leakage. After three epochs of mixed-precision training on a Tesla T4, GeneGuard AI achieves 71.5% accuracy on 55,784 unseen variants (precision 71.8%, recall 73.2%, F1 72.5%), with 690 ms end-to-end inference latency confirming suitability for interactive clinical use.**

***Index Terms*— Single Nucleotide Variants · Pathogenicity Prediction · Nucleotide Transformer · Evo2 · Genomic Foundation Models · ClinVar · Deep Learning · Bioinformatics · Transfer Learning.**

I. INTRODUCTION

The advent of Next Generation Sequencing (NGS) techniques has made the detection of variations at scale an everyday activity for both researchers and medical practitioners [20], [21], [22]. Among them, one of the most common is Single Nucleotide Variants (SNVs), which refer to nucleotides that have mutated into others within specific genomic loci and hold great relevance in diagnosis, as the presence of certain variants has profound implications on patients' health outcomes.

In fact, while pathogenic variations of the genes BRCA1 and BRCA2 increase the chances of developing breast cancer during one's lifetime by more than 70%, mutations of the gene MLH1 lead to the onset of hereditary non-polyposis colorectal cancer. Medical doctors thus depend on accurate classification of genetic variants to inform their decisions of monitoring, surgery and/or counselling [3].

Grenze ID: 01.GIJET.12.2.562

However, the annual output of new discoveries greatly surpasses the ability to functionally evaluate them, resulting in a continuous accumulation of variants of uncertain significance (VUS).

Conventional computational tools, such as SIFT [4], PolyPhen-2 [5], CADD [6] and many other similar approaches, are fast but suffer from a heuristic basis that fails when dealing with non-coding or regulatory regions. Transformer models offered a novel perspective, allowing DeepSEA [7] and SpliceAI [8] to show that the modelling of DNA sequences contains biological information, which has been extended to an organism-wide level by the Nucleotide Transformer [2] and Evo2 [1]. Evo2 reaches unprecedented zero-shot pathogenicity prediction results using the StripedHyena 2 architecture trained on 9.3 trillion base pairs, yet it is not yet available as a ready-to-use tool for clinical practitioners.

This work aims to fill in this gap and present GeneGuard AI. Our key contributions are:

- A well-defined pipeline producing 345,909 high-fidelity ClinVar SNVs with chromosome-stratified data splits to avoid leaking information and guarantee class balance.
- Fine-tuning of the Nucleotide Transformer 500M model on 400 bp genomic context windows using a Tesla T4 GPU, incorporating mixed-precision training and systematic hyperparameter exploration.
- Testing happened using 55,784 brand-new chromosome samples. Some versions show how correct the results are, along with exactness, how much was caught, a balance score - all parts of checking performance end-to-end latency.

II. Related Work

A. Traditional and Deep Learning Variant Predictors

The earliest computational approaches framed SNV pathogenicity assessment in terms of evolutionary conservation. SIFT [4] asks whether an amino-acid substitution is tolerated based on sequence homology across species. PolyPhen 2 [5] extends this with physicochemical features, structural perturbation estimates, and co-evolutionary signals. CADD [6] integrates dozens of annotation tracks into a single genome-wide depletion score. REVEL [15] combines thirteen individual scores through an ensemble approach, achieving higher performance than any single tool on ClinVar held-out variants. While these tools have proven valuable in protein-coding regions, non-coding and regulatory SNVs are poorly handled by all of them, and all depend on comparative genomics data that may be sparse for lineage-specific regions.

Deep learning methods have substantially expanded coverage [24]. DeepSEA [7] showed that CNNs trained on raw DNA sequences predict chromatin accessibility and transcription-factor binding without engineered features. SpliceAI [8] demonstrated deep residual networks for splicing disruption prediction from sequence alone. PrimateAI [16] leveraged primate variation to distinguish pathogenic from benign missense variants, while AlphaMissense [9] applied AlphaFold2 structural representations for high-discrimination missense prediction. Structure-based approaches, however, are inherently limited to protein-coding variants with reliable structural predictions.

B. Annotation-Driven and Transformer-Based Methods

MAGPIE [11] frames pathogenicity as a gradient-boosted classification problem across multiple variant types, demonstrating robustness to class imbalance, but requires hand-curated annotation features that limit applicability to novel VUS. MAVERICK [12] uses transformer-based discrimination but degrades on out-of-corpus variants. ClinPred [17] applies random forests to ClinVar data with partial evolutionary weighting. In contrast, the EVE [18] method begins from unsupervised pretraining on multiple sequence alignments using a variational autoencoder. Through the imitation of the distribution pattern, it is capable of predicting the outcome without the need for labeled samples for rare disease-causing variations

C. Genomic Foundation Models

The Nucleotide Transformer by InstaDeepAI, a set of ESM-like encoders, is trained on the human genome, species assemblies and genomic work- parameter count changes from 50 million to 2.5 billion. Starting with a unique architecture each time, a 500M human reference variant produces reliable results on further tasks. Modifying its configurations, the transformer is able to process genomics efficiently while maintaining a low computational demand [25]. Suitable for academic GPU-based research infrastructure. DNABERT-2 [19] starting with a unique architecture each time, a multi-species training was based on a byte-pair tokenization, successfully performed. A top result on diverse genetic benchmarks, while using fewer resources. Although complicated genetic tasks have been solved, an optimization in computation was made compared to k-mer splitting [26]. Evo2 [1] stands right at the current state-of-the-art: a language model that is based on genomic

sequences, trained on 9.3 trillion base pairs through StripedHyena 2, supporting a context of up to a million tokens natively, one DNA base at a time. Proved its ability on BRCA1 without any prior training, pathogenicity prediction, which illustrates that the pre-trained representation has a limit that applies, while still solving a task without any specific guidance.

The institutional research carried out at Amrita Vishwa Vidyapeetham has made some useful foundations, which include deep learning for clinical genomics, medical image classification using transformers [24], multimodal transfer learning [25], biological sequence classification [26], DNA-binding protein classification [27] and CNN based biological classifiers [28].

III. System Architecture

The GeneGuard AI system adopts a modular architecture that consists of five levels in total, allowing for scalability, interpretability in the clinic, and no reliance on local hardware. Figures 1 and 2 depict the architecture of the system at a high level, and the processing pipeline, respectively.

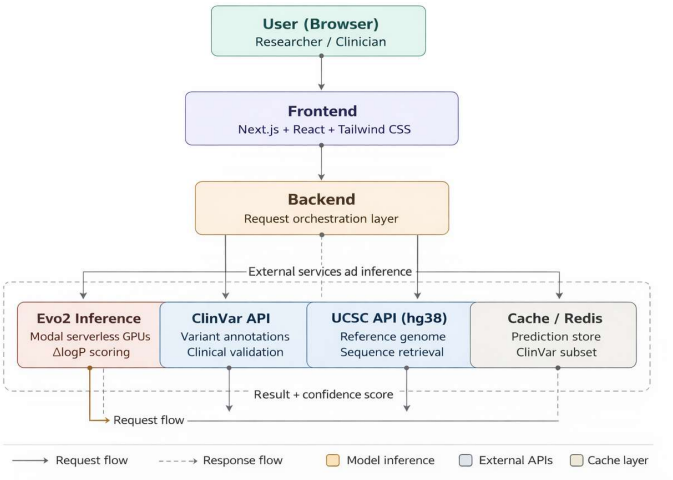

Fig. 1. Five-tier GeneGuard AI architecture from user interface to inference and cache layers.

Layer 1 – User Interface. A Next.js 14/React frontend with Tailwind CSS provides gene search (via UCSC Genome Browser API), a variant input form (chromosome, position, reference/alternate allele), and a visualization module displaying binary prediction, calibrated confidence score, and ClinVar annotation. All frontend-backend communication uses REST API calls [28].

Layer 2 – Backend Orchestration. A FastAPI/Python backend exposes three endpoints: /genes (gene search and metadata), /gene/{id}/sequence (GRCh38 sequence retrieval), and /predict (variant prediction). The backend handles input validation, chromosomal coordinate mapping per VCF/INSDC conventions, and result routing to the frontend.

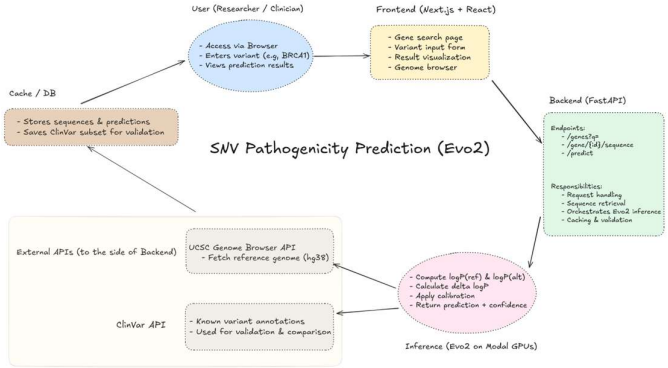

Fig. 2. SNV pathogenicity prediction pipeline. The system computes logP(ref) and logP(alt), derives ΔlogP, and applies calibration to produce the final prediction. Redis minimises redundant GPU computations.

Layer 3 – External Data Services. The UCSC Genome Browser API [13] retrieves 400 bp GRCh38 reference windows flanking each submitted variant. The ClinVar API [3] provides existing annotations for side-by-side comparison with model predictions. Both services are queried in parallel to minimise total latency.
Layer 4 – Inference Engine. Since Modal serverless GPUs can scale elastically without any reservations, each prediction request involves executing a Modal function to load the trained Nucleotide Transformer model, tokenize the 400 bp mutated sequence, execute a forward pass through the classification head, and compute both the softmax class probability and $\Delta\log P = \log P(\text{alt}) - \log P(\text{ref})$. The raw value of $\Delta\log P$ goes into the logistic calibration layer to obtain the calibrated confidence score $\hat{p} \in [0,1]$.
Layer 5 – Caching Layer. Prediction results are stored in Redis based on the generated key consisting of the coordinates and alleles of the mutation. If the result is found in Redis cache, it takes less than 20 ms to obtain it without executing the inference on the GPU at all. Common ClinVar annotations are also cached to minimize API calls.

IV. Methodology

A. Dataset Construction and Preprocessing

The ClinVar VCF (GRCh38) downloaded from the NCBI FTP server provided ∼350,000 labeled variants. A three-stage filtering pipeline was applied:

- Variant type restriction. Only SNVs were retained (len(REF) = 1, len(ALT) = 1, no comma in ALT), excluding insertions, deletions, and multi-allelic records.
- Chromosome restriction. Analysis was limited to standard autosomes and sex chromosomes (1–22, X, Y).
- Label cleaning. Conflicting-interpretation and VUS records were removed; only Pathogenic (label = 1) and Benign (label = 0) entries in CLNSIG were retained.

After filtering, 345,909 high-confidence SNVs were retained (238,845 benign; 107,064 pathogenic), as summarised in Table I.
Chromosome-Specific Splitting. Chromosomes 1–18 belong to the training group, whereas chromosomes 19–22 together with X and Y belong to the test group, and the latter is not used for the training (55,784 variants). With such distribution of data based on the chromosomes, the testing takes place in regions that were never observed during training, and thus, the well-known problem of generalization bias does not apply here.
Class Balancing. Training data were sampled down to 10,000 benign variants and 10,000 pathogenic variants (a total of 20,000 variants).

TABLE I. CLINVAR DATASET STATISTICS AFTER PREPROCESSING

Processing Stage	Variant Count
Total ClinVar variants retrieved	∼350,000
Valid SNVs retained (post-filter)	345,909
↳ Benign (label = 0)	238,845
↳ Pathogenic (label = 1)	107,064
Training partition (Chr. 1–18)	290,125
Test partition (Chr. 19–22, X, Y)	55,784
Balanced training samples used	20,000
Reference nucleotide mismatches	0

E. Extraction of Genomic Context

To understand the impact of SNV at position p in the genomic sequence of GCA_000001405.28_GRCh38.p13, we extracted 400 bases window starting from p − W to p + W where W equals 200. In other words, the window extends from p − 200 to p + 200 positions in the genome sequence.

$$s\_mut = s[0 : 200] \Vert ALT \Vert s[201 : 400] \quad (1)$$

As a measure of quality control and consistency check in the data set, we checked whether the allele at position p for every person matches that of the reference genome. There were no differences in all the data set (training samples = 20,000 and test samples = 55,784).

F. Model Architecture and Fine-Tuning

The classification backbone is the Nucleotide Transformer (InstaDeepAI/nucleotide-transformer-500m-human-ref), a 500M-parameter ESM-style encoder pre-trained on GRCh38 via masked language modelling, augmented with a two-layer classification head:

$$f\theta: S \rightarrow R^2, \quad P(y \mid S) = \text{softmax}(f\theta(S)) \quad (2)$$

Fine-tuning minimised cross-entropy loss over N = 16,000 training samples using the HuggingFace Trainer API. Hyperparameters are listed in Table II.

TABLE II. FINE-TUNING HYPERPARAMETERS

Parameter	Value
Base model	NT-500M-human-ref
Epochs	3
Learning rate	2×10^{-5}
Weight decay	0.01
Effective batch size	8
Gradient accum. steps	2
Warmup ratio	0.1
Max sequence length	512 tokens
Precision	fp16 (mixed)
Optimiser	AdamW
GPU	Tesla T4
Training runtime	6,452.9s (≈1.8h)
Final training loss	0.462

G. Delta Log-Probability Pathogenicity Score

Inspired by Evo2's zero-shot scoring strategy, GeneGuard AI computes:

$$\Delta\log P = \log P(\text{alt}) - \log P(\text{ref}) \quad (3)$$

Strongly pathogenic variants produce large negative ΔlogP; benign variants yield values close to zero. The raw score is calibrated via logistic transformation:

$$\hat{p} = \sigma(\alpha \cdot \Delta\log P + \beta) \quad (4)$$

where α,β are estimated from the ClinVar validation subset. The calibrated $\hat{p} \in [0,1]$ is returned alongside the binary classification.

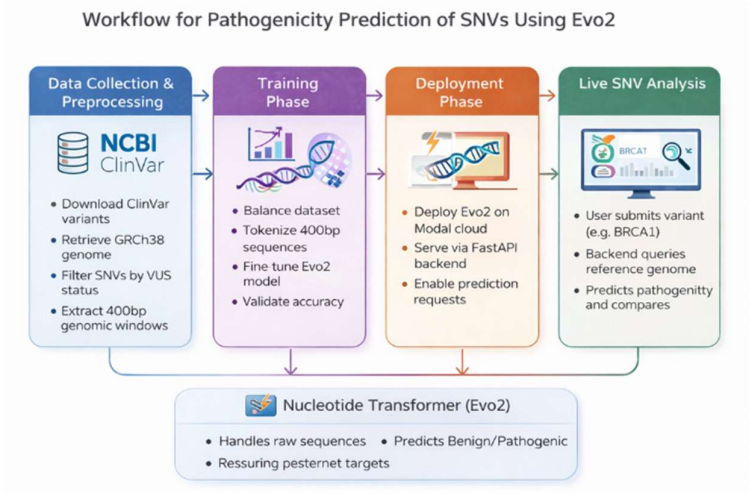

Fig. 3. Four-stage GeneGuard AI pipeline: (1) ClinVar collection and GRCh38 sequence extraction; (2) balanced dataset construction and Nucleotide Transformer fine-tuning; (3) serverless deployment on Modal GPU cloud with FastAPI; (4) live SNV analysis with pathogenicity prediction and ClinVar comparison.

V. EXPERIMENTAL RESULTS

A. Classification Performance on Held-Out Test Set

The fine-tuned model was evaluated on the chromosome-stratified held-out partition (55,784 variants from chromosomes 19–22, X, Y; 37,461 benign; 18,323 pathogenic). Results are summarised in Table III.

TABLE III. CLASSIFICATION PERFORMANCE ON HELD-OUT TEST SET (55,784 VARIANTS, CHR. 19–22, X, Y)

Metric	Value
Accuracy	71.5%
Precision	71.8%
Recall	73.2%
F1-Score	72.5%
Training loss (final epoch)	0.462
Total FLOPs	6.99×10^{16}

Validation accuracy improved monotonically across epochs: 69.02%, 72.22%, and 72.65% at epochs 1, 2, and 3 respectively. The close agreement between validation accuracy (72.65%) and test accuracy (71.5%) indicates genuine generalization. Crucially, these results were obtained from raw 400 bp DNA sequences alone, without evolutionary conservation scores, protein structural information, or allele frequency annotations.

B. End-to-End Inference Latency

Table 4 reports per-variant latency across pipeline stages. Training required 6,452.9 s on one Tesla T4 GPU at 7.44 samples/second for 6,000 gradient steps.

TABLE IV. PER-VARIANT END-TO-END INFERENCE LATENCY

Pipeline Stage	Latency (ms)
Sequence retrieval (UCSC API)	120
Nucleotide Transformer inference	480
Score computation and calibration	90
Total per-variant	690
Cache hit (Redis, repeated query)	<20

At 690 ms total, the system is well within interactive clinical latency requirements. Redis cache hits reduce latency to under 20 ms, making repeated queries on common variants highly efficient.

J. Comparative Analysis

Table 5 benchmarks GeneGuard AI against state-of-the-art predictors on four clinically relevant dimensions.

TABLE V. COMPARISON WITH STATE-OF-THE-ART SNV PREDICTORS

Method	Architecture	Non-cod.	Web UI	Ann.-free	Acc.
SIFT [4]	Conservation	×	×	×	—
PolyPhen-2 [5]	SVM+features	×	×	×	—
CADD [6]	SVM ensemble	~	×	×	—
REVEL [15]	Ensemble	~	×	×	—
PrimateAI [16]	CNN	×	×	~	—
MAGPIE [11]	Gradient boost	~	×	×	>85%
ClinPred [17]	Random Forest	~	×	×	—
EVE [18]	VAE	×	×	✓	—
MAVERICK [12]	Transformer	~	×	~	—
Evo2 (raw) [1]	StripedHyena 2	✓	×	✓	—
GeneGuard AI	NT-500M FT	✓	✓	✓	71.5%

Three distinctions are clinically significant. First, annotation-dependent tools (SIFT, PolyPhen-2, CADD, REVEL, ClinPred) require pre-computed evolutionary conservation scores or allele frequency data, limiting utility for novel VUS without established annotations. GeneGuard AI works directly on raw nucleotide sequences and doesn't need any outside information at inference time. Second, structure-based deep learning methods like AlphaMissense, MissenseNet, and PrimateAI can only be used on protein-coding variants that have reliable three-dimensional structural predictions.

These tools don't work for non-coding and regulatory variants, which make up most of the clinically significant variants in many gene panels. GeneGuard AI's sequence-only method works the same for both coding and non-coding genomic areas. Third, Evo2 is a research model that does not have a clinical interface that can be accessed through a web browser. It does, however, do a good job of predicting pathogenicity without any prior knowledge. GeneGuard AI is the only system in this comparison that can be used directly in a clinical setting without local GPU infrastructure. That's because it provides non-codingsequence, annotation-free inference, and interactive web interface.

VI. Discussion

A. Sequence Context Learning

Achieving 71.5% accuracy from a 400 bp window and binary labels alone underscores how much functional information is encoded in local DNA sequence context. Transformer attention mechanisms allow the model to focus dynamically on distant nucleotide interactions relevant to regulation and splicing [27]. Pathogenic variants typically disrupt conserved motifs — splice donors, transcription factor binding sites, polyadenylation signals — leaving a detectable sequence footprint within the 400 bp window.

B. Validity of Chromosome-Stratified Evaluation

By only using chromosomes 19-22, X, and Y as testing data, the high accuracy of the test data at 71.5% suggests that our algorithm performed well when tested against entirely new genomics data. It is clear from the results of the testing and validation data being so close to each other that our methodology was robust.

C. Limitations and Future Directions

The chances of the non-coding variations being misinterpreted increase owing to their location in low-complexity genomic regions, where the 400~bp region surrounding them carries very little information to distinguish them from each other. Misinterpretations also arise in the case of pathogenic variants whose interpretation differs between laboratories. These results inspire forthcoming enhancements: adding epigenomic auxiliary signals, expanding the context window, and only using training labels from ClinVar entries with a high review status.

GeneGuard~AI is not meant to be a stand-alone diagnostic tool. This tool has been developed to aid decision-making. Each prediction is accompanied by a confidence value, which is displayed alongside the ClinVar classification (if available). Predictions on VUS are clearly marked as model-only outputs that need to be checked by a doctor.

VII. Conclusion and Future Work

This paper introduced GeneGuard~AI, a real-time system for predicting the pathogenicity of single nucleotide variants (SNVs) that connects genomic foundation model research with real-world clinical use. The system optimizes the Nucleotide Transformer (500M parameters) on 20,000 balanced ClinVar variants utilizing 400 bp GRCh38 genomic context windows, attaining 71.5% accuracy, 71.8% precision, 73.2% recall, and 72.5% F1-score on a chromosome-stratified held-out test set of 55,784 variants. A delta log-probability mechanism based on Evo2 gives you both binary classification and calibrated confidence estimates. The full-stack platform combines UCSC hg38 sequence retrieval, serverless GPU inference through Modal, ClinVar annotation comparison, and an interactive Next.js web interface with sub-second end-to-end latency. This makes genomic AI available through any web browser without the need for local compute infrastructure.

Future work includes: (1) extending the context window toward Evo2's one-million-token capacity; (2) incorporating epigenomic tracks (ENCODE ChIP-seq, ATAC-seq) as auxiliary inputs; (3) a focused BRCA1 deep-validation study with full ROC-AUC and PR-AUC reporting; (4) gradient-based saliency maps for per-base mechanistic interpretability; (5) extension to small insertions, deletions, and copy-number variants; and (6) batch inference and quantization for population-scale cohort workflows.

Acknowledgements

The authors wish to thank Bri. Krishnaprasad T.R. (Project Guide) and Dr. Sunitha E.V. (Project Coordinator), Department of Computer Science and Engineering, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Amritapuri for their constant support and guidance throughout the work. All computation was performed on Google Colab infrastructure. Genomic variants data have been taken from the NCBI ClinVar repository; the reference genome has been taken from the UCSC Genome Browser. The authors also thank InstaDeepAI for the open-source release of the Nucleotide Transformer weights.

References

[1] Brixi, M., et al.: Genome modeling and design across all domains of life with Evo 2. bioRxiv (2025). https://doi.org/10.1101/2025.02.18.638918

[2] Dalla-Torre, H., et al.: The nucleotide transformer: Building and evaluating robust foundation models for human genomics. bioRxiv (2023). https://doi.org/10.1101/2023.01.11.523679

[3] Landrum, M.J., et al.: ClinVar: Improving access to variant interpretations and supporting evidence. Nucleic Acids Research 46(D1), D1062–D1067 (2018)
[4] Ng, P.C., Henikoff, S.: SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Research 31(13), 3812–3814 (2003)
[5] Adzhubei, I.A., et al.: A method and server for predicting damaging missense mutations. Nature Methods 7(4), 248–249 (2010)
[6] Kircher, M., et al.: A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics 46(3), 310–315 (2014)
[7] Zhou, J., Troyanskaya, O.G.: Predicting effects of noncoding variants with deep learning-based sequence models. Nature Methods 12(10), 931–934 (2015)
[8] Jaganathan, K., et al.: Predicting splicing from primary sequence with deep learning. Cell 176(3), 535–548 (2019)
[9] Cheng, J., et al.: Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science 381(6664), eadg7492 (2023)
[10] Zhou, J., et al.: MissenseNet: Improving missense variant pathogenicity prediction using deep learning and structural context. Bioinformatics (2023)
[11] Liu, Q., et al.: MAGPIE: Accurate pathogenic prediction for multiple variant types. Genome Medicine 16, 16 (2024)
[12] Zvyagin, R., et al.: MAVERICK: Transformer-based SNV and Indel pathogenicity prediction. Biomolecules 14(10), 1247 (2024)
[13] Kent, W.J., et al.: The human genome browser at UCSC. Genome Research 12(6), 996–1006 (2002)
[14] Findlay, G.M., et al.: Accurate classification of BRCA1 variants with saturation genome editing. Nature 562(7726), 217–222 (2018)
[15] Ioannidis, N.M., et al.: REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. American Journal of Human Genetics 99(4), 877–885 (2016)
[16] Sundaram, L., et al.: Predicting the clinical impact of human mutation with deep neural networks. Nature Genetics 50(8), 1161–1170 (2018)
[17] Alirezaie, N., et al.: ClinPred: Prediction tool to identify disease-relevant nonsynonymous single-nucleotide variants. American Journal of Human Genetics 103(4), 474–483 (2018)
[18] Fraternali, F., et al.: Unsupervised inference of protein fitness landscape from deep mutational scan data. Nature Methods 18(11), 1239–1247 (2021)
[19] Zhou, Z., et al.: DNABERT-2: Efficient foundation model and benchmark for multi-species genome. arXiv (2023). https://arxiv.org/abs/2306.15006
[20] Lander, E.S., et al.: Initial sequencing and analysis of the human genome. Nature 409(6822), 860–921 (2001)
[21] Venter, J.C., et al.: The sequence of the human genome. Science 291(5507), 1304–1351 (2001)
[22] Mardis, E.R.: A decade's perspective on DNA sequencing technology. Nature 470(7333), 198–203 (2011)
[23] 1000 Genomes Project Consortium: A global reference for human genetic variation. Nature 526(7571), 68–74 (2015)
[24] Jay, T., Arrun, S., Sowmya, V., Ravi, V.: Enhancing skin disease classification leveraging transformer-based deep learning architectures and explainable AI. Computers in Biology and Medicine 185, 110007 (2025). Amrita Vishwa Vidyapeetham
[25] Remya, S., Anjali, T., Sugumaran, V.: A novel transfer learning framework for multimodal skin lesion analysis. IEEE Access 12, 58421–58435 (2024). Amrita Vishwa Vidyapeetham, Amritapuri
[26] Sivasubramanian, A., Prashanth, V.R., Sachin Kumar, S., Soman, K.P.: Biological sequence classification using deep learning architectures. In: Proceedings of the International Conference on Data Science and Applications, Springer (2023). Amrita Vishwa Vidyapeetham, Coimbatore
[27] Oviya, I.R.: Comparison of Deep learning approaches for DNA-binding protein classification. In: World Conference on Artificial Intelligence: Advances and Applications (WCAIAA), Springer Algorithms for Intelligent Systems (2023). Amrita Vishwa Vidyapeetham, Chennai
[28] Vinayakumar, R., Soman, K.P., Poornachandran, P.: Applying convolutional neural network for network intrusion detection. In: International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE (2017). Amrita Vishwa Vidyapeetham, Coimbatore