



(AUTONOMOUS)

Affiliated to JNTUH, Approved by AICTE, Accredited by NAAC with A++ Grade, ISO 9001:2015 Certified
Kacharam, Shamshabad, Hyderabad – 501218, Telangana, India

Department of Computer Science and Engineering

III B. TECH II SEMESTER CSE-B

BATCH 2022-26	MINI PROJECT 2024-25	ACADEMIC YEAR 2024-25
--------------------------------	---------------------------------------	--

Batch Id : 22MPCS-B11

Title of the Project : DNA Classification using CNN

TEAM MEMBERS

S. No.	Roll Number	Student Name	Mail Id	Contact Number
1	22881A0570	Bugga Vinay	vinnybugga@gmail.com	6305318257
2	22881A0585	Kotari Swamy	Kotariswamy1@gmail.com	8341929472
3	22881A05A8	Paladugula Sowmya	Paladugulasowmya7@gmail.com	9014698273

GUIDE DETAILS

Name of the Guide	:	Mr.D Ganesh
Designation	:	Assistant Professor
Department	:	Computer Science and Engineering
Mail Id	:	dganesh@vardhaman.org
Contact Number	:	7207513076
Area of interest	:	High Performance Computing

*Signature of the
Projects In charge*

*Signature of the
Guide with Date*

*Signature of the
Project Coordinator*

Batch Id : 22MPCS-B11

Title of the Project : DNA Classification using CNN

Objectives

1. Accurate DNA Sequence Classification
2. Feature Extraction from DNA Sequences
3. Enhancing Classification Speed

Outcomes

1. Achieves high accuracy in classifying DNA sequences into predefined categories (e.g., species, diseases, or functional types).
2. Automatically learns relevant patterns and motifs from DNA sequences without the need for manual feature extraction.
3. Provides a fast and efficient classification model, reducing computational time compared to traditional methods.

ABSTRACT

DNA classification is crucial in genomics for species identification, genetic disorder detection, and mutation analysis. Traditional alignment-based methods are computationally expensive, prompting the need for more efficient approaches. This study proposes a Convolutional Neural Network (CNN)-based model for DNA sequence classification, where sequences are transformed into numerical representations to enable feature extraction. The CNN effectively captures spatial patterns and sequence motifs, improving classification accuracy while reducing computational complexity. Experimental results on benchmark datasets demonstrate the model's superiority over traditional machine learning methods. By leveraging deep learning, this approach enhances precision and efficiency, making it a promising tool for bioinformatics applications. Future enhancements could incorporate attention mechanisms and hybrid architectures for improved interpretability and performance. This study highlights the potential of CNNs in computational genomics, paving the way for advanced deep learning applications in DNA sequence analysis.

REFERENCES:

- [1] Trovato, M., Sartorius, R., D'Apice, L., Manco, R. and De Berardinis, P., 2020. Viral emerging diseases: challenges in developing vaccination strategies. *Frontiers in Immunology*, 11, p.2130.
- [2] Gelderblom, H.R., 1996. Structure and classification of viruses. *Medical Microbiology*. 4th edition.
- [3] Travers, A. and Muskhelishvili, G., 2015. DNA structure and function. *The FEBS journal*, 282(12), pp.2279-2295.
- [4] Randhawa, G.S., Soltysiak, M.P., El Roz, H., de Souza, C.P., Hill, K.A. and Kari, L., 2020. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. *Plos one*, 15(4), p.e0232391.
- [5] Liu, R., Qiao, M., Zheng, J. and Zhou, W., 2021. Analysis SARS-CoV-2 Genomes of G20 Areas on Phylogeny Tree, t-SNE based on Machine Learning.

Date of Submission:

*Signature of the
Guide with Date*

*Signature of the
HOD, CSE*