





















- [12] J. Lázaro-Guevara and K. Garrido, "A New approximate matching compression algorithm for DNA sequences," *bioRxiv*, p. 853358, 2019.
- [13] M. Silva, D. Pratas, and A. J. Pinho, "Efficient DNA sequence compression with neural networks," *GigaScience*, vol. 9, p. g119, 2020.
- [14] S. Chowdary, S. V. Kumar, D. Nedunuri, and V. Gupta, "A novel approach to compress dna repetitive sequences in bio-informatics," in *Journal of Physics: Conference Series*, 2019, p. 012026.
- [15] K. Banerjee and V. Bali, "Design and Development of Bioinformatics Feature Based DNA Sequence Data Compression Algorithm," *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 5, p. e5, 2020.
- [16] J. Partee, R. Hazell, A. Solsi, and J. Santerre, "Compressed DNA Representation for Efficient AMR Classification," *SMU Data Science Review*, vol. 3, p. 5, 2020.
- [17] R. Wang, Y. Bai, Y.-S. Chu, Z. Wang, Y. Wang, M. Sun, *et al.*, "DeepDNA: A hybrid convolutional and recurrent neural network for compressing human mitochondrial genomes," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018, pp. 270-274.
- [18] S. Bibi, J. Iqbal, A. Iftikhar, and M. Hassan, "Analysis of Compression Techniques for DNA Sequence Data," *arXiv preprint arXiv:2006.02232*, 2020.
- [19] A. Womack, "Cigarcoil: A new algorithm for the compression of dna sequencing data," 2019.
- [20] S. E. Zahra, K. Masood, and M. Asif, "DNA Compression using an innovative Index based Coding Algorithm," in *2019 22nd International Multi-topic Conference (INMIC)*, 2019, pp. 1-6.
- [21] M. Goyal, K. Tatwadi, S. Chandak, and I. Ochoa, "Deepzip: Lossless data compression using recurrent neural networks," *arXiv preprint arXiv:1811.08162*, 2018.
- [22] S. M. Hossein, D. De, P. K. D. Mohapatra, S. P. Mondal, A. Ahmadian, F. Ghaemi, *et al.*, "DNA Sequences Compression by GP<sup>2</sup> R and Selective Encryption Using Modified RSA Technique," *IEEE Access*, vol. 8, pp. 76880-76895, 2020.
- [23] S. M. Hossein, A. Mitra, P. K. D. Mohapatra, and D. De, "A Compression & Encryption Algorithms on DNA Sequences Using R 2 P & Selective Technique."
- [24] G. P. Arya, R. Bharti, D. Prasad, and V. Garg, "An improved method for DNA sequence compression," in *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*, 2017, pp. 1-4.
- [25] A. B. Majumder and S. Gupta, "Dynamic Programming Based DNA Compression Algorithm through Substitution Method," *International Journal of Engineering and Applied Sciences*, vol. 4, p. 257356.
- [26] K. Punitha and A. Murugan, "SegmentationBased DNA Sequence Compression."
- [27] F. Kounelis and C. Makris, "Comparison between text compression algorithms in biological sequences," *Information and Computation*, vol. 270, p. 104466, 2020.
- [28] T. Paridaens, "Compression and interoperable representation of genomic information," Ghent University, 2018.
- [29] C. Yin, "Encoding and decoding DNA sequences by integer chaos game representation," *Journal of Computational Biology*, vol. 26, pp. 143-151, 2019.
- [30] K. Kryukov, M. T. Ueda, S. Nakagawa, and T. Imanishi, "Nucleotide Archival Format (NAF) enables efficient lossless reference-free compression of DNA sequences," *Bioinformatics*, vol. 35, pp. 3826-3828, 2019.
- [31] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *arXiv preprint arXiv:1409.3215*, 2014.
- [32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735-1780, 1997.
- [33] T. Islam, C. H. Kim, H. Iwata, H. Shimono, A. Kimura, H. Zaw, *et al.*, "A Deep Learning Method to Impute Missing Values and Compress Genome-wide Polymorphism Data in Rice," 2021.

## Authors Profile



**Prashanth.M.C** author is currently pursuing Ph.D. in the department of computer science, Kuvempu University, Karnataka, INDIA. He is working in the domain of DNA sequences pattern matching. This author has completed his Master degree in computer science in 2013.



**M.Ravikumar** the author has completed his B. E, M. Tech and Ph.D. degrees in the year 1996,2001 and 2016 respectively. His Ph.D. topic is Estimation of Multiple Skews in Trilingual Handwritten Document Images. His research area includes Document images analysis, pattern matching, medical image processing. He has published more than fifteen papers and journals.