A Comprehensive Survey on DNA Analysis using Machine Learning Approaches

Mr.Shivanand Maradi, Dr. Chethan L. S
Assistant Professor, Department of CSE, PESITM, Shiamoga
Associate Professor, Department of CSE, PESITM, Shiamoga.

Abstract: **Nowadays, pandemic infections causing enormous harm to the living life is most challenging and crucial part. If it is identified and medicated these kinds of infections in the early stage will help humans to survive. Today the Biomedical department is storing the umbilical cord of mother when giving birth to a child. This storing of umbilical cord help the upcoming generation to overcome unseen diseases with the help of analyzing, sequencing the DNA pattern. Human DNA extracted from saliva that was fresh or frozen for three, six and twelve months using different extraction protocols like Oragene – Commercial kit, QIAamp DNA mini kit, DNA extraction using ammonium acetate, Instagene. This Research work investigation helps the upcoming Academician, Researchers to overcome the problem. This paper focus on investigation existing literature and techniques used to study DNA by using different Machine learning algorithms.**

INTRODUCTION

Analysis of DNA has made a easy part for understanding Evolutionary Biology, Identifying genetical diseases that may affect the humans. The area of genetical, medicine and Biotechnology has revolutionized into digitized part. However analysis and prediction are majority hampered by large data which is produce by DNA sequence. The genetic material found in humans and in all species is called deoxyribonucleic acid (DNA), which is made up of two long strands that form a double helix and are made up of four different types of nucleotides: adenine (A), cytosine (C), guanine (G), and thymine (T). Biological sequence data have been applied in many fields, such as the biological sequences similarity analysis, phylogenetic relation-ships construction, gene function analysis and protein structure analysis. The role of Machine Learning (ML) in analysis DNA for Predicting, Analyzing and Pattern matching help people to identify children of a particular parent. The forensic department is doing continuous research with the help of trending techniques or algorithm available from the domain of ML. This technology or tool will help many challenges that we can fall in the area of DNA. The challenges we can see that area are:

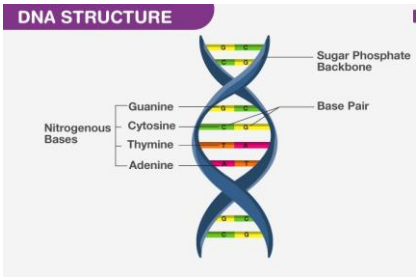
1. Searching a pattern in DNA sequence.
2. To detect the pattern in a DNA
3. To find the particular method for compressing of DNA sequence.

Every human body has unique structure of DNA pattern which contains 99.9% identical (DNA), only 0.1% of different pattern. So individual mapping of a particular cell in body is difficult. The information present in the genetic code of DNA will help biological growth and reproduce inheritance. Pattern matching focuses on finding the repetition of particular pattern in a DNA sequence. Efficiently compressing and indexing highly repetitive DNA sequences has become one of the most challenging topics in computational biology and bioinformatics.

Basics of DNA Analysis

DNA Structure and Sequencing:
The genetic material found in humans and nearly all other species is called deoxyribonucleic acid, or DNA. It is made up of two long strands that form a double helix and are made up of four different types of nucleotides: adenine (A), cytosine (C), guanine (G), and thymine (T).

The technique of precisely figuring out a DNA molecule's nucleotide sequence is known as DNA sequencing. Large-scale and quick sequencing has been made possible by technologies like Sanger sequencing and next generation sequencing (NGS).



DNA Diagram representing the DNA Structure

Role of Machine Learning in DNA Analysis The creation of algorithms for machine learning enables computers to recognize patterns in data and forecast outcomes. Machine learning (ML) can help with a number of DNA analysis issues, including predictive modeling, data integration, and pattern identification.

1. Data Preprocessing:
- Noise Reduction: By eliminating errors and artifacts, ML approaches can assist in cleaning and preprocessing raw DNA sequencing data.
- Feature extraction: the process of extracting pertinent features from unprocessed sequencing data, such as certain genetic markers or motifs.

2. Pattern Recognition:
- Sequence Alignment: Machine learning techniques can expedite and increase the precision of DNA sequence alignment, which is essential for locating homologous areas.
- Motif Discovery: Finding patterns in DNA sequences that are conserved and could point to functional or regulatory regions.

3. Predictive Modeling:
- Disease Prediction: Based on genetic variances, machine learning models can forecast the likelihood of various diseases.
- Trait Association: Regression models and other predictive algorithms are used to find correlations between phenotypic traits and genetic markers.

4. Genomic Data Integration:
- Multi-omics analysis: combining information from transcriptomics, proteomics, genomics, and other omics to give a comprehensive picture of biological systems.
- Network Analysis: To comprehend gene-gene and protein-protein interactions, ML can assist in the construction and analysis of biological networks.

5. Personalized Medicine:
- Recommendations for Treatment: Personalized treatment regimens can be suggested by machine learning models that examine unique genetic profiles.
- Drug discovery: using genetic data to forecast the safety and effectiveness of novel medications.

Machine Learning Techniques in DNA Analysis

1. Supervised Learning:
- Classification: Determining groups or classes in genetic data, including differentiating between mutation kinds.
- Regression: Using DNA sequence data to forecast continuous outcomes, such as gene expression levels.

2. Unsupervised Learning:
- Clustering is the process of assembling genetic profiles or DNA sequences that are comparable without using labels.
- Dimensionality Reduction: Methods for bringing high-dimensional genetic data down in complexity, such as Principal Component Analysis (PCA).

3. Deep Learning:
- Neural Networks: For pattern recognition and DNA sequence analysis, convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are used.
- Autoencoders: For data compression and unsupervised feature learning.

4. Reinforcement Learning:
- Adaptive Algorithms: Developing adaptive learning algorithms that can optimize sequencing protocols or experimental designs based on feedback.

Background

Several DNA analysis techniques are crucial for understanding genetic information, diagnosing diseases, and advancing research in genomics, medicine, and biotechnology. Below are some of the most important DNA analysis techniques:

1. Polymerase Chain Reaction (PCR)

The goal is to create millions of copies of DNA from a single or limited number of copies by amplifying short regions of DNA.
Use: Infectious disease diagnosis, forensic science, gene analysis, and cloning.

2. DNA Sequencing

Goal: Establishes the precise nucleotide (A, T, C, and G) order within a DNA molecule.

- Kinds:
- Sanger Sequencing: The conventional approach, renowned for its excellent precision.
- Next-Generation Sequencing (NGS): This technique makes it possible to sequence entire genomes or particular areas at high throughput, facilitating extensive genomic research.
- Use: Personalized medicine, evolutionary study, genomics, and genetic mutation detection.

3. Restriction Fragment Length Polymorphism (RFLP)
- Goal: Uses restriction enzymes to break down DNA and examine the resultant fragments to find differences in DNA sequences.
- Use: Genetic diversity analysis, DNA fingerprinting, and genetic mapping.

4. Gel Electrophoresis
- Goal: Uses an electric field to separate DNA fragments into different sizes inside a gel matrix.
- Use: DNA analysis for genotyping, evaluating PCR product quality, and determining DNA fragment size.

5. Southern Blotting
- Goal: By transferring DNA fragments from a gel to a membrane and hybridizing with a tagged probe, this technique can identify certain DNA sequences within a complicated DNA sample.
- Use:Researching gene rearrangements, DNA fingerprinting, and gene identification.

6. DNA Microarray
- Goal: Makes it possible to identify particular DNA sequences or analyze thousands of genes' expression levels simultaneously.
- Use: Identifying mutations or polymorphisms, genotyping, and profiling gene expression.

7. CRISPR-Cas9
- Goal: A genome-editing instrument that enables accurate adjustments to the DNA sequence at particular sites.
- Use: Gene editing for scientific investigations, the creation of genetically engineered organisms, functional genomics, and medicinal development.

8. Quantitative PCR (qPCR) or Real-Time PCR
- Goal: Measures the quantity of DNA or RNA in a sample in real time, which is frequently used to assess the degree of gene expression.
- Use: Quantification of viral load, identification of genetic variants, and gene expression analysis.

9. Single Nucleotide Polymorphism (SNP) Analysis
- Goal: Identifies frequent genetic mutations in DNA, known as single nucleotide variants.
- Use: Personalized medicine, association studies, and genetic mapping.

10. Whole Genome Sequencing (WGS)
- Goal: By sequencing an organism's entire genome, this method offers a comprehensive picture of its entire genetic composition.
- Use: Determining genetic variants, researching uncommon illnesses, and understanding evolution.

11. Whole Exome Sequencing (WES)
- Goal: The majority of mutations that cause disease are found in the exons, or protein-coding sections, which make up 1% to 2% of the whole genome. These regions are the focus of genome sequencing efforts.
- Use: Genetic research, clinical diagnostics, and locating mutations that causes disease.

12. Fluorescence In Situ Hybridization (FISH)
- Goal: Locates and detects particular DNA sequences on chromosomes using fluorescent probes.
- Use: Cancer diagnosis, gene mapping, and chromosomal abnormalities.

13. Comparative Genomic Hybridization (CGH)
- Identifies and quantifies copy number variants (CNVs) throughout the genome by contrasting the DNA of several specimens.
- Use: Investigating genomic instability, diagnosing genetic abnormalities, and cancer research.

14. Methylation Analysis
- Goal: Researches DNA methylation patterns, which are essential for controlling gene

expression and preserving the integrity of the genome.
- Use: Research on epigenetics, cancer research, and comprehending gene control.

15. Single-Cell DNA Sequencing
- Goal: Examines each cell's DNA to provide light on cellular heterogeneity and uncommon cell populations.
- Use: Investigating genetic diversity in tissues, developmental biology, and cancer research.
- These techniques form the backbone of modern DNA analysis, enabling researchers and clinicians to explore genetic information at various levels of detail and apply this knowledge across multiple fields.

LITERATURE REVIEW

Pooja Mishra [1] et.al has proposed minimum variance Huffman coding to save big data in small volume for longer period of time. An efficient algorithm to store large amount of data in small number of nucleotides has been proposed which uses minimum variance Huffman coding. In the proposed work initially transferring DNA data to into corresponding the binary data. Huffman coding with Minimum Variance is used to generate DNA trees. The encoding technique extracts DNA data using binary sequence and DNA Tree. DNA sequences are saved. DNA data that has been stored is retrieved. The decoding algorithm is used to extract binary sequences utilizing the recovered DNA data and the same DNA

Biological sequence for the usage of texture feature is proposed [2] by Weiyang Chen, et.al. The field of sequencing technology has advanced quickly in recent years. As a result, a lot of biological sequence data are produced. Biological sequences have been the subject of numerous studies due of their significance. An essential component of visual design, texture is typically employed to characterize how different image intensities are arranged in space. The field of sequencing technology has advanced quickly in recent years. As a result, a lot of biological sequence data are produced. Biological sequences have been the subject of numerous studies due of their significance. An essential component of visual design, texture is typically employed to characterize how different image intensities are arranged in space. The biological sequence's texture features and created the calculation technique by combining the biological sequence's texture features and created the calculation technique by combining the digital coding of the biological sequence with the picture texture feature calculation method. We used this method to quantify and analyze characteristics found in DNA sequences.

The problem for Efficient Compression and Indexing for Highly Repetitive DNA Sequence Collections [3] propose by Hongwei Huo et.al. says that crucial issue of indexing and searching collections of extremely repetitive DNA sequences. We suggest using dynamic programming to uncover space-efficient variants between T and the sequences in G, allowing us to create concise structures that facilitate effective search. Experimental results on the tested data show that the suggested method offers considerable improvements over the present state-of-the-art methods in terms of retrieval time and space consumption for extremely repetitive sequences. Several algorithms, such as LM, FINDDiff, and GetDif, were employed. The source code is available online.

YASMIN A. BADR [4] et. al has proposed Automatic Clustering of DNA Sequences.With Intelligent Techniques. a novel soft computing metaheuristic framework is presented in this paper for automatic clustering. A pulse coupled neural network (PCNN) is used to determine how similar or dissimilar a DNA sequence is. To solve the automatic data clustering problem, the popular genetic algorithm is hybridized with the bat algorithm. Using the extended human oral microbiome database (eHOMD), extensive computer studies are performed. The suggested hybrid algorithm outperformed the bat algorithm and the normal genetic algorithm, according to a comparison of the experimental findings. To determine the hybrid performance's superiority, it was also compared to rival algorithms found in the literature review. A Mann-Whitney-Wilcoxon rank-sum test is used to verify the resulting clusters statistically.

DNA Encoding-Based Nucleotide Pattern and Deep Features for Instance and Class-Based Image Retrieval [5] by Jitesh Pradhan et. al proposes that to improve the retrieval accuracy of the content-

based image retrieval (CBIR) system by using the benefits of DNA encoding. To create DNA planes from a given image, the most important bit-plane-based DNA encoding approach has been proposed first. The image's produced DNA planes effectively and compactly capture the key visual elements. The encoded DNA planes have now been applied to picture retrieval and feature extraction based on nucleotide patterns. In order to carry out classification-based picture retrieval, the translated and amplified encoded DNA planes have also been simultaneously implemented on various deep learning architectures, such as ResNet-50, VGG-16, VGG-19, and Inception V3. The suggested system's effectiveness has been assessed using the use of two corals, an object, and a dataset of medical images.

The problem on Predicting the Sequence Specificities of DNA-Binding Proteins by DNA Fine-Tuned Language Model With Decaying Learning Rates [6] proposed by Ying He et. al says that a deep learning framework for DNA- binding protein (DBP) sequence specificity prediction is presented. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid CNN-RNN models are all used in tandem. A Human Genome Language Model and a DBPs Fine-Tune Language Model are included in the system. The Classification Model combines features from both language models for prediction, and the DBPs Fine-Tune Language Model is trained by refining the Human Genome Language Model to collect DBP features. In the paper, the suggested models are contrasted with contemporary pre- trained models such as fine3 and DNABERT, as well as traditional approaches.

Inbal Preuss Inbal Preuss [7] proposed Sequencing Coverage Analysis for Combinatorial DNA-Based Storage Systems to sequencing coverage in combinatorial DNA- based storage devices is examined in this study. The paper uses binomial and multinomial distributions to model the chance of decoding combinatorial DNA sequences. It contains algorithms for figuring out how successful a sequence decoding through simulation is. Important methods include finding the computed probability based on error rates, read distribution, and sequence length, as well as the median simulated probability.

Weakly-Supervised Convolutional Neural Network Architecture for Predicting Protein- DNA Binding [8] proposed by Qinhu Zhang et. al. says that The study's focus is on a unique design known as the weakly-supervised convolutional neural network (WSCNN), which improves protein-DNA binding prediction. Convolutional neural networks (CNNs) and multiple-instance learning (MIL) are integrated to achieve this. DNA sequences are broken down into overlapping sub- sequences using the WSCNN model, which then uses fusion techniques to combine the projected scores. The results show that the model exhibits good performance on datasets that are in vitro and in vivo. The article also goes over the difficulties in implementing advanced pooling layers in the context of MIL and CNN integration, emphasizing the importance of using reverse-complement modes for precise predictions.

The problem on Needleman-Wunsch Attention: A Framework for Enhancing DNA Sequence Embedding [9] by KYELIM LEE et. al proposes the framework Needleman-Wunsch Attention, intended to improve DNA sequence embedding, is presented in this research. The method maps DNA sequences into vector representations by means of self-attention and cross-attention mechanisms. In order to rebuild input sequences into vectors while preserving alignment information, encoder and decoder networks are utilized. A variety of probability models are analyzed in the paper, along with combinatorial sequence decoding techniques.

PEYMAN NEAMATOLLAHI [10] et. Al Simple and Efficient Pattern Matching Algorithms for Biological Sequences has proposed research study is DNA sequence pattern matching algorithms. It compares and contrasts approximate versus exact pattern matching, emphasizing exact pattern matching in particular. The novel algorithms FLPM, PAPM, and LFPM are introduced along with a discussion of brute force techniques. An improvement on PAPM, LFPM is appropriate in situations requiring the search of several patterns. The focus of the paper is on word- and character-based processing methods.

CONCLUSION

In this context “A Comprehensive Survey on DNA Analysis is using Machine Learning Approaches", focus on Utilization of machine learning (ML) algorithms for DNA analysis. The key areas are Functional annotation, mutation prediction, sequence classification. The different techniques used are Supervised (SVM, neural networks) and

unsupervised learning (clustering, PCA). The finding of the survey is ML enhances efficiency and accuracy in predicting phenotypic features, identifying disease mutations, and annotating genomic regions. The implications are Potential advancements in genetics, personalized medicine, and biotechnology through improved DNA analysis methods. The study also highlights the role of deep learning frameworks and their potential to revolutionize DNA analysis, thereby advancing fields like genetics, medicine, and biotechnology.

REFERENCES

[1] Pooja Mishra , Chiranjeev Bhaya , Arup Kumar Pal, and Abhay Kumar Singh "Compressed DNA Coding Using Minimum Variance Huffman Tree" IEEE COMMUNICATIONS LETTERS, VOL. 24, NO. 8, AUGUST 2020.

[2] Weiyang Chen, and Weiwei Li "Definition and usage of texture feature for biological sequence" IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS

[3] Hongwei Huo, *Senior Member, IEEE,* Xiaoyang Chen, Xu Guo, Jeffrey Scott Vitter, *Fellow, IEEE "Efficient Compression and Indexing for Highly Repetitive DNA Sequence Collections" JOURNAL OF LATEX CLASS FILES, VOL. 14, NO. 8, AUGUST 2019*

[4] YASMIN A. BADR 1 , KHALED T. WASSIF 2 , AND MAHMOUD OTHMAN 1 " Automatic Clustering of DNA Sequences With Intelligent Techniques" Received September 19, 2021, accepted October 8, 2021, date of publication October 13, 2021, date of current version October 21, 2021. Digital Object Identifier 10.1109/ACCESS.2021.3119560.

[5] Jitesh Pradhan , Arup Kumar Pal , SK Hafizul Islam , Senior Member, IEEE, and Chiranjeev Bhaya " DNA Encoding-Based Nucleotide Pattern and Deep Features for Instance and Class-Based Image Retrieval" IEEE TRANSACTIONS ON NANOBIOSCIENCE, VOL. 23, NO. 1, JANUARY 2024.

[6] Ying He , Qinhu Zhang , Siguo Wang , Zhanheng Chen , Zhen Cui , Zhen-Hao Guo, and De-Shuang Huang "Predicting the Sequence Specificities of DNA-Binding Proteins by DNA Fine-Tuned Language Model With Decaying Learning Rates" IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, VOL. 20, NO.1, JANUARY/FEBRUARY 2023.

[7] Inbal Preuss, Student Member, IEEE, Ben Galili , Zohar Yakhini , Member, IEEE, and Leon Anavy " Sequencing Coverage Analysis for Combinatorial DNA-Based Storage Systems " IEEE TRANSACTIONS ON MOLECULAR, BIOLOGICAL, AND MULTI-SCALE COMMUNICATIONS, VOL. 10, NO. 2, JUNE 2024.

[8] Qinhu Zhang, Lin Zhu , Wenzheng Bao , and De- Shuang Huang " Weakly-Supervised Convolutional Neural Network Architecture for Predicting Protein- DNA Binding" IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, VOL. 17, NO.2, MARCH/APRIL 2020.

[9] KYELIM LEE 1AND ALBERT NO 2 , (Member, IEEE) " Needleman-Wunsch Attention: A Framework for Enhancing DNA Sequence Embedding" Received 5 March 2024, accepted 12 May 2024, date of publication 15 May 2024, date of current version 22 May 2024. Digital Object Identifier 10.1109/ACCESS.2024.3401464.

[10] PEYMAN NEAMATOLLAHI , MONTASSIR HADI , AND MAHMOUD NAGHIBZADEH " Simple and Efficient Pattern Matching Algorithms for Biological Sequences" Received November 14, 2019, accepted January 13, 2020, date of publication January 23, 2020, date of current version February 6, 2020. Digital Object Identifier 10.1109/ACCESS.2020.29.

Author	Year	Data Sets	Techniques	Results
Pooja Mishra , Chiranjeev Bhaya , Arup Kumar Pal, and Abhay Kumar Singh	2020	490 ASCII characters and two images of size 256 × 256 and 512 × 512,	Huffman algorithm, minimum variance binary Huffman tree has been used to create DNA	The algorithm gives a high code-rate of the DNA sequences, and therefore reduces synthesizing cost.

			sequences of nucleotides which preserve the constraints.	
Weiyang Weiwei Chen, and Li	2020	DNA, biological data	Clustering analysis to classify both twelve primate species and nineteen features.	In our method, we can not only calculate various co-occurrence frequencies, but also define various texture features and corresponding calculation methods. Third, this method can be extended to analyze any biological sequence, and all biological sequences can be digitally coded and texture features can be calculated by this method.
Hongwei Huo, *Senior Member, IEEE,* Xiaoyang Chen, Xu Guo, Jeffrey Scott Vitter, *Fellow, IEEE*	2019	Datasets contains 100 sequences; each sequence is of size 1M. The alphabet set of each of the four datasets is *{A, T, G, C, N}*	propose a new reference sequence-based compressed indexing structure, called FMQ, for highly repetitive DNA sequence collections *G* that provides fast search in *G*.	FMQ is 42.18, 7.63, 2.09, and 1.72 times faster than the best query time of the other compared indexing methods on *DNA*, *E. coli*, *para*, and *cere*, respectively.
Yasmin A. Badr 1 , Khaled T. Wassif 2 , And Mahmoud Othman 1	2021	Human Oral Microbiome Database (eHOMD)	HYBRID GENETIC BAT ALGORITHM (GABAT)	The simulation results have shown that the hybrid GABAT outperformed the two state-of-the-art clustering algorithms and other metaheuristic algorithms, namely, firefly, particle swarm optimization and FAPSO.
Jitesh Pradhan , Arup Kumar Pal , SK Hafizul Islam , Senior Member, IEEE, and Chiranjeev Bhaya	2024	coral and an object dataset. These datasets are Corel-1000, GHIM-10K, and COIL-100	DNA Encoding, CBIR	The proposed methods use the properties of DNA encoding on images for content based image retrieval. It extracts the intensity values from every pixel of the image and quantifies it to a string of DNA nucleotides. The comparative study validate that DNA features work significantly better than direct image features in CBIR
Ying He , Qinhu Zhang , Siguo Wang , Zhanheng Chen , Zhen Cui Zhen-Hao Guo, and De-Shuang Huang	2023	DFLM mainly used two types of data sets, the unlabeled large-scale Genome Reference Consortium Human Build 38(GRCh38/hg38) [1] and about 500 labeled ChIP-seq data set [2].	DNA Fine-tuned Language Model (DFLM).	The proposed method have visualized and analyzed three coding methods one-hot, pre-training methods and fine tuning to understand the model in a more intuitive way. In conclusion, the experiments illustrate that our model performs better across all models combined, with fine3 performing the best.

Inbal Preuss , *Student Member, IEEE,* Ben Galili , Zohar Yakhini , *Member, IEEE,* and Leon Anavy	2024	standard DNA-based data storage	Combinatorial DNA encoding. Decoding Probability of a Single Combinatorial Letter, Combinatorial Sequence, Complete Combinatorial Message	One of the key aspects of the combinatorial approach is the strategic selection of Ω that consists of easily distinguishable k- mers. This, together with the use of a threshold $t > 1$ in the reconstruction algorithm (See Algorithm 1), effectively mitigates k-mer mix-up errors, as was demonstrated in [15]. We therefore chose to ignore k-mer mix-up errors in the model used for the reconstruction probability.
Qinhu Zhang , Lin Zhu , Wenzheng Bao , and De-Shuang Huang	2020	48 public transcription factor ChIPseq datasets from the ENCODE project, where each ChIPseq dataset has an average of 45000 positive sequences,and each sequence consists of 199 bps.	Weakly-Supervised Convolutional Neural Network (WSCNN))	The experimental results on in vivo and in vitro datasets show that WSCNN significantly outperforms the competing methods, suggesting the effectiveness of WSCNN. In addition, we give a quantitative analysis of the importance of the reverse-complement mode in predicting in vivo protein-DNA binding, and demonstrate that WSCNN has a good generalization ability if provided with enough data, and compare the performance of WSCNN when the four fusion methods are separately used, and explain why not use advanced pooling layers to combine MIL with CNN.
KYELIM LEE 1 AND ALBERT NO 2,	2024	Qiita [42] and DNA-Fountain [43] dataset	Needleman-Wunsch, sequence embedding	We conducted a comprehensive evaluation of various encoders, applying the Needleman-Wunsch Attention (NWA) framework in some instances, to demonstrate the improvements in embedding performance achieved through our proposed methodology.
Peyman Eamatollahi , Montassir Hadi , And Mahmoud Naghibzadeh	2020	The HRG dataset	First-Last Pattern Matching Algorithm, Processor-Aware Pattern Matching Algorithm, Least Frequency Pattern Matching Algorithm.	The present work's experimental results reveal that the proposed algorithms, especially LFPM, surpass the other simulated algorithms in terms of time cost. This improvement is mainly due to having decreased the number of found windows.

Table 1: Summary of overall Literature Review