International Journal of Contemporary Issues in Social Sciences
Volume 3, Issue 3, 2024 ISSN(P):2959-3808 | 2959-2461

MACHINE LEARNING TECHNIQUES BASED DNA SEQUENCE ALIGNMENT: A SYSTEMATIC LITERATURE REVIEW

Sidra Ali[1], Department of Computer Science, City University of Science and Information Technology, KPK, Email: Sidraali1289@gmail.com,

Muhammad Arif Shah[*2], Department of Computer Science, Pak-Austria Fachhochschule, Institute of Applied Sciences and Technology, Haripur Email: arif.websol@gmail.com

Hamza Shaukat[3], School of Technology, Jamk University of Applied Sciences, Finland, Email: , hamzakhan12599@gmail.com

Muhammad Zakir Khan[4], James watt school of engineering University of Glasgow Email: M.khan.6@research.gla.ac.uk

Corresponding authors*

Received: July 20, 2024 **Revised:** July 20, 2024 **Accepted:** September 20, 2024 **Published:** September 23, 2024

ABSTRACT

The massive volume of organic information, the conventional software engineering procedures and calculations neglect to take care of complex natural issues of this present reality. In any case, present day computational methodologies, for example, AI can address the restrictions of the customary strategies. AI has assumed a significant job in building up Bioinformatics as a field in its own in the course of the most recent 30 years. For solving the complex problems of biological data we use Machine Learning Techniques for Deoxyribonucleic Acid (DNA) Sequence Alignment data. We present a Systematic Literature Review Protocol (SLRP) of DNA sequence alignment using machine learning technique. This proposition played out an archived arrangement of explicit techniques which are useful to utilize the Systematic Literature Review (SLR). The anticipated outcomes of this survey recognize the DNA Sequence Alignment utilizing Machine Learning Techniques, investigate issues, order the issues, characterize the significant qualities and furthermore talk about the general attributes. The normal advantages of this investigation in future will be Systematic best in class of machine learning method in bioinformatics that will be supportive for new Researchers to represent DNA succession arrangement utilizing machine learning procedures.

Keywords: DNA sequence alignment; systematic literature review; machine learning; bioinformatics

INTRODUCTION

Bioinformatics is a multidisciplinary field that is in steady development due to technological propels in corresponded sciences (for example software engineering, science, scientific, science, and medicine) (Pevsner 2009). Genomic research is the most delegate space in bioinformatics, as it is the underlying advance of a few kinds of trials and it is likewise required in a few different bioinformatics fields. It looks at genomic highlight's DNA successions, qualities, administrative groupings, or other genomic auxiliary segments of various life forms. When all is said in done, similar genomics begins with the arrangement of genomic orthologous groupings (I.e. arrangements that share a typical heritage) for checking the degree of likeness (preservation) among groupings (or genomes) (Miller, Makova et al. 2004).

Because of the expansion of the quantity of examinations (that are likewise getting progressively perplexing) including genomic research just as DNA sequencing advances the sum and unpredictability of natural information is being

expanded. It straightforwardly influences the presentation of the computational execution of bioinformatics tests (Koboldt, Steinberg et al. 2013). So for taking care of such essential issue we utilize the DNA arrangement.

DNA is an atom that conveys the hereditary data. It comprises of genetic guidelines for building, running, keeping up a living being and giving life to the people to come. DNA is in twofold helix structure in which two individual DNA strands bend around one another in a winding. The DNA strand comprises of four nucleotide bases Adenine, Cytosine, Guanine and Thymine. Condensed as A, C, G, and T. In atomic science DNA sequencing the approach toward determining the particular entreaty of nucleotides privileged a DNA particle which is made by the reiteration of the four nucleotides: adenine, thymine, guanine and cytosine. The human genome is comprised of 3 billion of these Genetic characters. There are at any rate 26 billion base pairs (bp) (Cohen 2004).

In bioinformatics, a Sequence Alignment (SA) is considered to be a method of orchestrating groupings of proteins, RNA and DNA with a goal to discover locales of likeness which may give extra data on the practical, auxiliary, developmental and different interests between the arrangements. Attuned sequences of amino corrosive or nucleotide remains are ordinarily enunciated to as lines classified inside a framework, one on head of the other. For instance, given two arrangements, ATATAGAGGACACG and ATAGGGGACATGG, one potential arrangement. In this arrangement, the vertical lines demonstrate the match. The firmly adjusted districts in the adjusted arrangements are called comparable locales. These comparative locales are the areas generally preserved from past ages. In certain areas, exceptional characters, for example, '- ', otherwise called indels or holes are included. Focusing on the goal that vague or analogous typescripts are attuned in reformist fragments, holes are entrenched among the collections. This inclusion of an exceptional image speaks to a transformation (change) or could be seen as erasure from the other succession's viewpoint.

DNA sequence alignment have two primary sorts. The first is Multiple Sequence Alignment (MSA) is the predominant practice for reasoning organic realities from a lot of arrangements. It incorporates the arrangement of multiple groupings. A MSA can be seen as a 2-D table. In this table groupings are the lines and the segments of identical DNA Sequence are orchestrated by putting hole typescripts in realistic locations, with the end goal that the organic relationship of the successions is best described (Simossis, Kleinjung et al. 2003).

Pairwise Sequence Alignment (PSA) is an alignment between any two given sequences. Pairwise sequence alignment could be additionally named nearby and worldwide sequence alignment. Nearby sequence alignment finds the best estimated sub-sequence coordinate inside twofold specified sequences. Neighbourhood sequence alignments are planned fundamentally to scan for exceptionally comparative areas inside the two given sequences. Worldwide sequence alignment, then again, is intended to locate the best alignment of the two sequences completely. Along these lines, worldwide sequence alignment searches for worldwide planning between whole sequences (Haque, Aravind et al. 2009).

Because of immense measures of organic information and an exceptionally enormous number of potential blends and stages of different natural sequences, the traditional human knowledge-based techniques can't work viably and productively. So man-made consciousness strategies, for example, AI can assume a basic job in complex biomedical applications (Pan, Wang et al. 2014). AI (ML) is a subfield of man-made brainpower and is worried about the improvement of calculations and methods that permit PCs to learn. As of late, the measure of natural information requiring investigation has detonated and many AI techniques have been created to manage this blast of information. Consequently, AI in bioinformatics has become a significant examination territory for both PC researchers and scholars (Lacey and Xie 2014). Important AI techniques incorporate help vector machines, portion machines, highlight choice, neural systems, developmental calculation, factual learning, fluffy rationale, regulated getting the hang of, grouping, gathering learning, Bayesian systems, direct relapse, head segments investigation, concealed Markov models, entropy-based data strategies, and numerous others (Pan, Wang et al. 2014). This examination presents a SLR trying to distinguish the Machine Learning strategies in DNA sequence Alignment encouraging the SLR procedure. SLR is a settled examination strategy used to coordinate the best

accessible observational information from methodical exploration (Kitchenham 2004).

As opposed to the regular impromptu writing survey process, SLR gives dependable methodologies and set up sequences to total, assess and decipher the best accessible individual examinations to address specific exploration questions (Kitchenham 2004). It additionally permits commentators to decide the genuine impacts and wonders in zones where little, singular examinations are not effortlessly controlled or replicated (Brereton, Kitchenham et al. 2007).

I. Background

In this section background and related works about the Machine learning techniques used for DNA sequence alignment are discussed. Computational microbiology has increased noteworthy notoriety over the most recent couple of decades. The enormous volume of organic information that is put away as DNA, RNA and protein sequences requires broad processing capacity to recover and investigate sequences rapidly and precisely. Adjusting the sequences is a significant and basic advance in tackling issues, for example, foreseeing the auxiliary and tertiary structure of a protein, anticipating the tribal sequence or recognizing the normal qualities in two living beings. In any case, the multifaceted nature because of the sheer number of potential blends and searches makes sequence alignment a figure escalated issue. This intricacy increments exponentially with the size of the sequences. Both equipment and programming enhancements are generally considered as likely bearings to improve the speed and exactness (Haque, Aravind et al. 2009).

With new organic sequences being found nearly every day, the natural sequence database is developing exponentially.

This blast of information requests new calculations which are quick but then effective. The test to adjust both speed and productivity was immediately perceived by numerous scientists and in the previous hardly any decades. They utilized AI method to take care of such large issue. AI has assumed a significant job in setting up Bioinformatics as a Field in its own directly in the course of the most recent 30 years. A great deal of methods running from randomized choice trees to neural systems, bolster vector machines and concealed Markov models have been applied effectively to take care of issues in novel quality finding, developmental investigation, tranquilize and horticultural examination and protein grouping (Lacey and Xie 2014). Our concentration is the Machine learning strategies created for DNA sequence alignment.

A. Related Works

Numerous research contributions are in the region of bioinformatics for DNA sequence alignment Machine learning methods in different gatherings and Journals. As far as we could possibly know, there is no organized work or SLRP distributed for the Machine learning procedures utilizing DNA sequence alignment.

To perform orderly surveys is important to follow a pre-built up and all around characterized convention. Following the characterized orderly advances may ensure the reproducibility of the examination.

Cochrane Collaboration, a global association that produces orderly writing surveys of mediations in human services, proposing a handbook to help direct precise audits (Pentheroudakis, Greco et al. 2009). They suggest that the initial step of an efficient audit is to build up a convention that obviously characterizes the targets of the survey, the measures for consideration and avoidance of systematic review, the techniques that distinguish the investigations, and the examination strategy for the gathered examinations. The principle consequence of a Cochrane Alliance SLR is a rundown of the finest value logical investigations for a particular topic (Kanewala and Bieman 2014), (van Karnebeek and Stockler 2012). (van Karnebeek and Stockler 2012) proposed that the influence factor, quantity of references, and duration of distribution are imperative to choose and rank pertinent logical papers enveloping the audited subject.

(Goel, Singh et al. 2013) characterized that a hypothetical audit of delicate figuring strategies for quality expectation. The issue of quality expectation, alongside the issues engaged with it, is first portrayed (Goel, Singh et al. 2013). A concise depiction of delicate figuring procedures, and their application to quality expectation, a rundown of various delicate registering strategies for quality forecast lastly a few confinements of the ebb and flow examination and future exploration

bearings are introduced (Polato, Ré et al. 2014). Nikolayevskyy et al (2014) proposed a method for performing efficient surveys for programming designing analysts and make points of interest rules for new zones, for example, bioinformatics (Nikolayevskyy, Kranzer et al. 2016).

(Chowdhury and Garai 2017) characterizes a survey on numerous sequence alignment from the point of view of hereditary calculation (Chowdhury and Garai 2017). The protein alignment issue is studied for an extensive duration, unfortunately, every reachable technique yields alignment results contrastingly for a solitary alignment issue. Numerous sequence alignment is exposed as an extremely high computational intricate concern. Plentiful stochastic procedures, in this way, are reflected for improving the precision of alignment. Among them, several scientists every now and again utilize Genetic Algorithm. This examination propose several kinds of the strategy applied in alignment and the ongoing configurations in them multi unprejudiced genetic deviousness for explaining different sequence alignment. Numerous ongoing investigations have revealed noteworthy advancement in discovery of the alignment exactitude.

(Ezziane 2005) article plans investigate the utilization of AI in the areas of bioinformatics and DNA sequencing. And furthermore portrays the sort of programming programs that helps the necessities of scientists to use and support untangle the tremendous actions of evidence that are persistently being amassed in genomic research (Ezziane 2005).

Babasaheb .S. Satpute, Dr. Raghav Yadav (2007) states that because of enormous development in the measure of organic information. We need generally progressed and in fact compelling innovation and apparatuses to use for the examination of those data.so this article present the survey of Machine Learning Techniques for Bioinformatics and Computational Biology and a portion of the renowned regulated order and Clustering calculations (Khattree and Naik 2007).

This writing talked about in this segment take a shot at the mechanization of clinical procedure particularly treatment arranging and furthermore on helps the dental specialist in discovering dental illness. It particularly empowers the dental specialist to diagnose for the disorder nature and select the appropriate cure disposition.

B. Existing Systematic Reviews of DNA Sequence Alignment

1. review on multiple sequence alignment from the perspective of Genetic algorithm (Chowdhury and Garai 2017)
2. Machine learning technology in the application of genome analysis: A systematic review (Wu and Zhao 2019).
3. Applications of artificial intelligence in bioinformatics: A review (Ezziane 2005)

There are hardly any works has been done in the precise surveys distributed in DNA sequence alignment. The principal survey introduced in the Elsevier on the multi-objective hereditary calculation for fathoming different sequence alignment in DNA sequence alignment by (Chowdhury and Garai 2017) And the second review presented in SCIENCE DIRECT in 2019 by Jie Wu and Yiqiang Zhao. and the 3rd review presented in SCIENCE DIRECT in 2006 by (Ezziane 2005).

In first review, that propose by Biswanath Chowdhury, Gautam Garai (Chowdhury and Garai 2017) the fundamental reason for this investigation center around numerous sequence alignment from the viewpoint of hereditary calculation in DNA sequence alignment with numerous parts missing in this examination, for example, information extractions, information Synthesis and cutoff determination system. Also, the investigation has numerous needs and impediments, for example, just one AI calculation is talk about, no information extraction and no information union.

The second precise audit proposed by Jie Wu, Yiqiang Zhao and the point of this investigation is to distinguish the utilization of AI in that identified with genomic examination. Also, the examination has numerous needs and confinements, for example, just talk about AI procedures, no conversation of web indexes, no information extraction and no information union.

The third article proposed by (Ezziane 2005) and the presumes to inspect the employment of AI in the precincts of bioinformatics and DNA sequencing. And furthermore portrays the sort of programming programs that helps the necessities of researcher to help interpret the huge processes of

evidence that are continually being gathered in genomic research.

The consequences of these deliberate audits are not identified with our examination. Anyway no orderly survey DNA sequence alignment utilizing AI procedures already has been distributed hence, we create methodical audit in software engineering sector to fill the slum in the momentum study territory of AI strategies in DNA sequence alignment for additional examination.

Research Method

A systematic review is an examination approach created to get, calculate, and comprehend the whole data. That exploration philosophy is worried about a particular examination question or region of concern and distinguishing holes in the ebb and flow research. Nonetheless; this examination follows the methodical audit rule proposed by Kichenham (Smith, Turner et al. 2016). This rule has been adjusted by utilizing the software engineering research issues; which are taken from clinical scientists.

A. Systematic literature review protocol

Systematic literature review protocol is utilized to indicate the strategy that plays out a recorded arrangement finish before beginning the orderly audit. Nonetheless, a decent class convention deals the effective orderly survey and most of the time refreshed during any timeframe to comprise more current distributions (Abdelmaboud, Jawawi et al. 2015).

1. Research Questions

Characterizing the Research inquiries for the deliberate audit is a significant advance. This survey tended to the accompanying examination questions:

Table 1: Research Questions

RQ#	Research Question
RQ 1	Which Machine Learning techniques are used for Sequence Alignment and what are the frequently used techniques?
RQ 2	Which dataset is used in Machine Learning Sequence Alignment papers and what are the frequently used dataset?
RQ 3	What are the Assembled techniques used in Machine Learning in DNA Sequence Alignment?
RQ 4	What are the evaluation parameters used in Machine Learning Sequence alignment papers?
RQ 5	Which Machine Learning techniques are used for Multiple Sequence Alignment?
RQ 6	Which Machine Learning techniques are used Pairwise Sequence Alignment?

2. Research Objectives

There are a few goals characterized in this deliberate survey convention the advantages of these targets to give Researcher's best in class of AI strategies in DNA sequence alignment. The Research objectives are:

Table 2: Research Objectives

ID	Objective
1	To identify Machine Learning in DNA Sequence Alignment and frequently techniques.
2	To identify datasets in Machine Learning techniques in DNA Sequence Alignment and frequently datasets.
3	To identify assembled techniques in DNA Sequence Alignment.
4	To identify evaluation parameters in Machine Learning techniques DNA Sequence Alignment.
5	To classify ML techniques used for Multiple Sequence Alignment.
6	To classify ML techniques used for Pairwise Sequence Alignment.

3. Data Sources

Different electronic databases have been utilized as essential hotspots for software engineering research distributions. All of the databases yielded various outcomes aside from Google Scholar; it restores indistinguishable outcomes from past databases Table 1 yet we utilized Google Scholar search database to give high calibre of indexed lists. Now and again a few papers are effectively imported from Google scholar instead of the other electronic databases Table 3.

Table 3: Data Sources

Database Source	URL
ACM Digital Library	https://dl.acm.org/

Science Direct	https://www.sciencedirect.com/
Google Scholar	https://scholar.google.com
IEEE Xplore	https://ieeexplore.ieee.org/
Springer Digital Library	https://link.springer.com/

4. Search Strategy and Search Strings

The search strategy conducted from September to January 2017. It is chosen to begin this efficient survey convention from year 2018 as a result of the genuine examination's in the DNA sequence alignment start during the year 2017. With centre to look through choices of sources, we guaranteed that the hunt involved diaries, magazines, meetings, book areas, conference and workshops. We tried many inquiry sequences, and the accompanying restored the quantity of correlated articles:

("DNA sequence alignment or AI method") AND ("numerous sequence alignment utilizing AI strategies") AND "pairwise sequence alignment utilizing AI procedures". All databases like Google researcher acknowledge this string.

5. Study Selection

The key investigations chosen, rely upon the study choice characterized by the consideration and avoidance models appeared in Table 4 during perusing the title and theoretical of the papers so as to guarantee that the outcomes identified with the examination territory under study. Now and again, titles and modified works are not widely inclusive and accordingly are not satisfactory. In this way we declaim the entire paper to ensure that the data delighted the incorporation and avoidance models (Inclusion/exclusion Criteria).

The methodical writing survey is recovered from the electronic databases Table 3 are 101 papers as shows in Figure 1. In the wake of understanding titles, abstracts and applying incorporation and prohibition rules the quantity of papers decreased to 83 papers and sent out to Endnote database. We evacuate copied papers 15 and perusing the full content 50 papers by considering consideration and avoidance rules we dispose of 3 papers the staying 83 papers chose as primary investigations for SLR.

Table 4: Study Selection Criteria

Inclusion Criteria	Exclusion Criteria
Papers in DNA Sequence Alignment related to computer science	Papers in DNA Sequence Alignment related to medical sciences
A scientific paper	The paper was not available in English Language.
Reviewed Paper (by electronic databases)	Published data not available;
Papers in press	DNA sequencing classification or DNA sequence compression
Papers that describe DNA Multiple Sequence Alignment or Pairwise Sequence Alignment using Machine Learning techniques.	Papers that describe only DNA Sequence Alignment.

6. Data Extraction

The query item of theoretical examinations is sent out from the database sources Table 3 with linked data (paper title, Author, paper reference type, and so on.) to Endnote database. The advantages of utilizing the Endnote database is to enrol all data identified with select examinations and simple to keep up and deal with this data. Table 5 shows the information things that were utilized in each study including portrayals and examination addresses identified with our efficient survey directed. The principal research question RQ1 shows to the DNA sequence alignment using machine learning techniques and frequently used machine learning techniques in DNA sequence alignment that require investigation for all data extraction items. The RQ2 indicates the dataset and frequently used data set, RQ3 indicates the assembled techniques and RQ4 indicates evaluation parameters in DNA sequence alignment using machine learning techniques. The fifth question RQ5 used to find out that which machine learning techniques is used for multiple sequence alignment and the sixth question RQ6 used to find out that which machine learning techniques is used for pairwise sequence alignment. Table 5 shows the Data items and Descriptions with relevant research questions.

Figure 1: Selected Papers

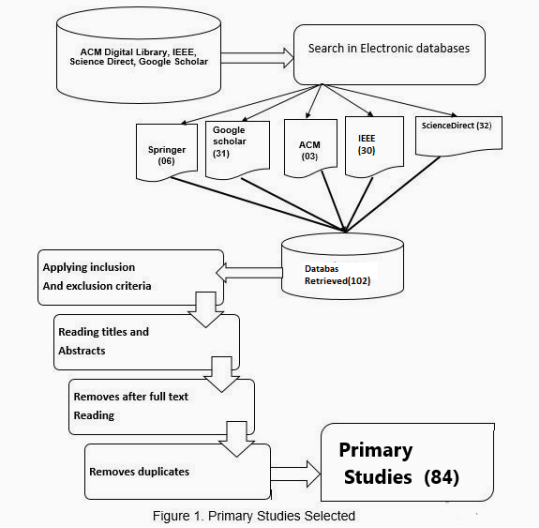


Figure 1. Primary Studies Selected

Table 5: Data Extraction Form

Data item	Description	Research Questions
Title	Title name	RQ1
Authors	Study authors name	RQ1
Reference type	Journal paper, conference paper, etc.	RQ1
ML techniques	Describe the DNA Sequence Alignment using Machine Learning techniques and mostly used techniques.	RQ1
Datasets	Describe datasets and mostly used datasets in DNA Sequence Alignment using Machine Learning techniques.	RQ2
Assembled techniques, evaluation parameters	Describe assembled techniques and evaluation parameters in DNA Sequence Alignment using ML techniques.	RQ1,RQ3,RQ4
Techniques in Multiple and Pairwise Sequence Alignment	Describe machine learning techniques used for Multiple Sequence Alignment, Pairwise Sequence Alignment.	RQ1, RQ5,RQ6
Findings	Main conclusion	RQ1

7. Data Synthesis

Data synthesis of this study aims to take care of principle questions proposed toward the start of the study. The primary fundamental inquiry is to distinguish AI procedures in DNA sequence alignment and their commitment type in bioinformatics that distributed in every year. The sub questions used to discover the AI strategies in DNA sequence alignment in bioinformatics.

The huge volume of organic information that is put away as DNA, RNA and protein sequences requires broad registering capacity to recover and dissect sequences rapidly and precisely. Adjusting the sequences is a significant and basic advance in taking care of issues, anticipating the familial sequence or recognizing the regular qualities in two creatures. They utilized AI strategy to take care of such enormous issue. AI has assumed a significant job in building up Bioinformatics from most recent 30 years. A great deal of strategies running from randomized choice trees to neural systems, bolster vector machines and concealed Markov models

have been applied effectively to take care of issues in novel quality finding, transformative examination, medicate and farming exploration and protein arrangement.

101 papers were encompassed in SLR. These papers include several machine learning techniques which is genetic algorithm, ant bee colony, artificial colony optimization, swarm particle optimization, artificial neural network , fuzzy logic etc. we group the paper based on the type on DNA sequence alignment using machine learning techniques. Five data sources are used to retrieve the relevant papers on DNA sequence alignment using machine learning techniques.

8. Threats to validity

The primary dangers to the legitimacy of our audit convention are broke down from the accompanying three perspectives are: rejection of pertinent articles, distribution inclination, and information extraction predisposition, study choice inclination.

Prohibition of significant articles: One of the significant issues we observed in this survey was discovering the important papers that tended to the research questions. To accomplish this target, we directed a hunt on databases recorded in table 3, utilizing our inquiry string on their web crawlers. Nonetheless, we perceived the likelihood that some significant studies would not be reverted by the search strings we utilized. To lessen the risk, we physically checked the reference rundown of every one of the significant investigations to search for any pertinent examinations that were missed in the robotized search.

Information extraction inclination: along with finding and choosing all the important examinations, information abstraction was the most basic errand in this study. To effectively separate information from these examinations, I read each paper freely and gathered the information introduced in Table 5 that are required to respond to the exploration addresses presented.

Publication bias: Only DNA sequence alignment using machine learning techniques studies are taken into account, the reason is that the authors may have some unfairness towards DNA sequence alignment using machine learning techniques. Therefore, there is likely a risk of miscalculating the performance of DNA sequence alignment using machine learning methods.

II. Review results

This section discusses the results associated with the Research Questions Table 2. These questions were aimed at analysing DNA sequence alignment using machine learning techniques studies from six perspectives: machine learning techniques used for DNA sequence alignment, datasets and evaluation parameters used for DNA sequence alignment using machine learning techniques, assembled techniques, machine learning techniques for multiple and pairwise sequence alignment. We talk about and decipher the outcomes identified with every one of these inquiries in the subsections beneath.

Table 6: Selected studies

S. No	Year	Authors	Title	Publisher
1	1992	(Eppstein, Galil et al. 1992)	Sparse Dynamic Programming Linear Cost Functions	ACM
2	2013	(Fukunishi, Finch et al. 2013)	A Bayesian Alignment Approach to Transliteration Mining	Acm
3	2014	(Mohanty and Tragoudas 2014)	Scalable Offline Searches in DNA Sequences	Acm
4	2016	Huazheng Zhu, Zhongshi He & Yuanyuan Jia	A Novel Approach to MSA Using Multi-objective EA Based on Decomposition	Ieee
5	2000	(Zhang, Schwartz et al. 2000)	A Greedy Algorithm for Aligning DNA Sequences (2)	GS

6	2004	(Kumar, Tamura et al. 2004)	MEGA3 Integrated software for Molecular Evolutionary Genetics Analysis and SA	GS
7	2006	(Torres and Nieto 2006)	Fuzzy Logic in Medicine and Bioinformatics	GS
8	2007	(Gondro and Kinghorn 2007)	A simple GA for MSA	GS
9	2010	(Wang, Huang et al. 2010)	BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features	GS
10	2011	(Verma, Singh et al. 2011)	DNA Sequence Assembly using PSO	GS
11	2012	(Gupta, Agarwal et al. 2012)	Genetic Algorithm Based Approach for Obtaining Alignment of Multiple Sequences	GS
12	2014	(Huang, Chen et al. 2015)	A memetic PSO algorithm for solving the DNA fragment assembly problem	GS
13	2015	(Kumar 2015)	AN ENHNCED ALGO FR MSA OF PROTEIN SEQ USNG GA	GS
14	2016	(Karaboga and Aslan 2016)	A discrete ABC algorithm for detecting transcription factor binding sites in DNA sequences	GS
15	2016	(Othman 2016)	Survey of the use of genetic algorithm for multiple sequence alignment	GS
16	2018	(Karaboga and Aslan 2019)	discovery of conserved regions in DNA sequences by (ABC) algorithm based methods	GS
17	2018	(Leslie, Eskin et al. 2004)	mismatch-string-kernels-for-svm-protein-classification	GS
18	1997	(Zhang and Wong 1997)	Toward Efficient MSA A System of Genetic and Dynamic Programming	IEEE
19	2001	(Chang and Halgamuge 2001)	fuzzy-sequence-pattern-matching-in-zinc-finger-domain-proteins	IEEE
20	2002	(Ando and Iba 2002)	ant-algorithm-for-construction-of-evolutionary-tree	IEEE
21	2002	(Nguyen, Yoshihara et al. 2002)	a-parallel-hybrid-genetic-algorithm-for-multiple-protein-sequence alignment	IEEE
22	2003	(Meksangsouy and Chaiyaratana 2003)	dna-fragment-assembly-using-an-ant-colony-system-algorithm	IEEE
23	2007	(Nasser, Vert et al. 2007)	Multiple Sequence Alignment using Fuzzy LOGIC	IEEE
24	2008	(Ramaswamy and Purdy 2008)	An Extended Library of Hardware Modules for GA with Applications to DNA Sequence Matching	IEEE
25	2008	(Zhao, Ma et al. 2008)	An Improved Ant Colony Algorithm for DNA Sequence Alignment	IEEE
26	2008	(Al Junid, Abd Majid et al. 2008)	HIGH SPEED DNA SEQUENCING ACCELERATOR USING FPGA	IEEE
27	2008	(Lei and Ruan 2008)	Particle Swarm Optimization Algorithm for Finding DNA Sequence Motifs	IEEE
28	2009	(Bir, Dongardive et al. 2009)	Building Consensus of Human Papillomavirus using Genetic Algorithm	IEEE
29	2009	(Mohamed, Othman et al. 2009)	Classification of “Gracilaria changii” Protein Sequences Using Back-Propagation Classifier	IEEE
30	2009	(Arribas-Gil, Metzler et al. 2008)	Statistical Alignment with a Sequence Evolution Model Allowing Rate Heterogeneity along the Sequence	IEEE

31	2010	(Mahmud, Hosen et al. 2010)	A Novel Two-Tier Multiple Sequence Alignment algorithms	IEEE
32	2010	(Lei, Sun et al. 2010)	Artificial Bee Colony Algorithm for Solving MSA	IEEE
33	2011	Ankit Agrawal and Xiaoqiu Huang	Pairwise Statistical Significance of LSA Using Sequence-Specific POSITION	IEEE
34	2011	(Yu 2011)	Solving Sequence Alignment Based on Chaos Particle Swarm Optimization Algorithm	IEEE
35	2012	(Nagar and Hahsler 2012)	A Novel Quasi-Alignment-Based Method for Discovering Conserved Regions in Genetic Sequences	IEEE
36	2012	(Verma 2012)	DSAPSO DNA Sequence Assembly using Continuous PSA with Smallest Position Value Rule	IEEE
37	2012	(Al Junid, Reffin et al. 2012)	implementation of GA FOR DNA seqnce alignment	IEEE
38	2012	(Othman and Abdel-Azim 2012)	MSA Based on GA with new Chromosomes representation	IEEE
39	2012	(Halgaswaththa, Atukorale et al. 2012)	Neural Network Based Phylogenetic Analysis	IEEE
40	2013	(Borovska, Gancheva et al. 2013)	Massively Parallel Algorithm for MSA	IEEE
41	2016	(Zheng, Li et al. 2016)	A Modified Multiple Alignment Fast Fourier Transform with Higher Efficiency	IEEE
42	2016	Huazheng Zhu, et al;	A Novel Approach to MSA Using Multi-objective EA Based on Decomposition	IEEE
43	2017	(Rani and Ramyachitra 2017)	Application of Genetic Algorithm by Influencing the Crossover Parameters for MSA	IEEE
44	2017	(Kaur and Sohi 2017)	Pairwise Sequence Alignment Method Using Flower pollination algo	IEEE
45	2017	(Siswanto, Hendric et al. 2017)	The Genomic Plant Warehouse Framework SLR	IEEE
46	2002	(Rajapakse and Faleel 2002)	genetic-approach-to-biosequence-alignment-gaba	IEEE
47	2005	(Ressom, Natarajan et al. 2005)	Applications of fuzzy logic in genomics	SD
48	2007	(Jangam and Chakraborti 2007)	A novel method for alignment of 2 nucleic acid sequences using ACO and GA	SD
49	2007	(Ho, Yu et al. 2007)	Design of accurate predictors for DNA-binding sites in proteins using hybrid SVM–PSSM method	SD
50	2008	(Kim, Kim et al. 2008)	A DNA sequence alignment algorithm using quality information and a fuzzy inference method	SD
51	2008	(Blum, Vallès et al. 2008)	An ant colony optimization algorithm for DNA sequencing by hybridization	SD
52	2008	(Lee, Su et al. 2008)	Genetic algorithm with ant colony optimization (GA-ACO) for multiple sequence alignment	SD
53	2008	(Qian, Yang et al. 2008)	Particle swarm optimization for SNP haplotype reconstruction problem	SD
54	2010	(Bi 2010)	Deterministic local alignment methods improved by a simple GA	SD
55	2012	(Zou, Shan et al. 2012)	A Novel Center Star Multiple Sequence Alignment Algorithm Based on Affine Gap Penalty and K-Band	SD

56	2012	(Li, Wang et al. 2012)	Improvements on a privacy-protection algorithm for DNA sequences with generalization lattices	SD
57	2013	(Hassanien, Al-Shammari et al. 2013)	Computational intelligence techniques in bioinformatics	SD
58	2014	(Kaya, Sarhan et al. 2014)	Multiple Sequence Alignment with Affine Gap by Using Multi-Objective Genetic Algorithm	SD
59	2015	(Garai and Chowdhury 2015)	A cascaded pairwise biomolecular sequence alignment	SD
60	2015	(Lee, Yeu et al. 2015)	BulkAligner A novel sequence alignment algorithm based on graph theory and Trinity	SD
61	2015	(Ji, Pu et al. 2015)	One-dimensional pairwise CNN for the global alignment of two DNA SEQUENCES	SD
62	2016	(Rajasekhar, Lynn et al. 2017)	Computing with the Collective Intelligence of Honey Bees – A Survey	SD
63	2016	(Rubio-Largo, Vega-Rodríguez et al. 2016)	Hybrid Multiobjective Artificial Bee Colony for MSA	SD
64	2016	(Rani and Ramyachitra 2016)	Multiple sequence alignment using multi-objective based bacterial foraging optimization algorithm	SD
65	2016	(Dakhli and Bellil 2016)	Wavelet Neural Networks for DNA Sequence Classification Using the Genetic Algorithms and the Least Trimmed Square	SD
66	2017	(Chowdhury and Garai 2017)	A review on multiple sequence alignment from the perspective of genetic algorithm	SD
67	2017	(Moustafa, Elhosseini et al. 2017)	Fragmented protein sequence alignment using two-layer particle swarm optimization (FTLPSO)	SD
68	2018	(Amorim, Neves et al. 2018)	An approach for COFFEE objective function to global DNA MSA	SD
69	2018	Mohamed (Issa, Hassanien et al. 2018)	ASCA-PSO Adaptive sine cosine optimization algorithm integrated with particle swarm for pairwise LSA	SD
70	2018	(Surendar, Shaik et al. 2018)	Micro Sequence Identification of DNA Data Using Pattern Mining Techniques	SD
71	2018	(Rubio-Largo, Vanneschi et al. 2018)	Swarm intelligence for optimizing the parameters of multiple sequence Aligners	SD
72	2019	(Saw, Raj et al. 2019)	Alignment-free method for DNA sequence clustering using Fuzzy integral similarity	SD
73	2019	(Wu and Zhao 2019)	Machine learning technology in the application of genome analysis A systematic review	SD
74	2004	(Horng, Wu et al. 2005)	A genetic algorithm for multiple sequence alignment	SPRINGER
75	2009	(Xu and Chen 2009)	A Method for Multiple Sequence Alignment based on PSO	SPRINGER
76	2010	(Xu and Lei 2010)	Multiple Sequence Alignment Based on ABC_SA	SPRINGER
77	2012	(Agarwal, Gupta et al. 2012)	A Genetic Algorithm for Alignment of Multiple DNA Sequences	SPRINGER
78	2018	(Majid, Khan et al. 2019)	Application of Parallel Vector Space Model for Large-Scale DNA Sequence Analysis	SPRINGER

79	2018	(Karaboga and Aslan 2019)	Discovery of conserved regions in DNA sequences by Artificial Bee Colony (ABC) algorithm based methods	SPRINGER
80	2019	(Ishaq, Khan et al. 2019)	Current Trends and Ongoing Progress in the Computational Alignment of Biological Sequences	IEEE
81	2014	(Lacey and Xie 2014)	Supervised Machine Learning Techniques in Bioinformatics: Protein Classi cation	ANN, SVM
82	2011	(Wen, Li et al. 2012)	Systematic literature review of machine learning based software development effort estimation models	SD
83	2014	(Idri, azzahra Amazal et al. 2015)	Analogy-based software development effort estimation: A systematic mapping and review	SD
84	2021	(Amr Ezz El-Din Rashed et al.2021)	Sequence Alignment using Machine Learning-Based Needleman-Wunsch Algorithm	IEEE

A. ML techniques used for DNA sequence alignment (RQ1)

We recognized 16 types of ML techniques, applied to DNA sequence alignment. They are listed as follows.

- Dynamic programming (DP)
- Artificial Neural Networks (ANN)
- Support Vector Machine (SVM)
- Genetic Algorithms (GA)
- Fuzzy logic (FL)
- Artificial bee colony(ABC)
- Ant colony optimization(ACO)
- Particle swarm optimization(PSO)
- Needle-Wunsch Algorithm

Among the above listed ML techniques, GA, PSO, ACO, FL, ABC are the five most every now and again utilized ones; they together were received by 86% of the chosen investigations, as outlined in Fig. 2. This presents just the measure of examination consideration that each variety of ML method has gotten during the previous 20 years; as a supplement to Fig. 2, Fig. 3 is plotted to additionally introduce the dispersion of exploration consideration in every distribution year. As appeared in Fig. 3, on one hand, a conspicuous distribution top shows up around year 2008.

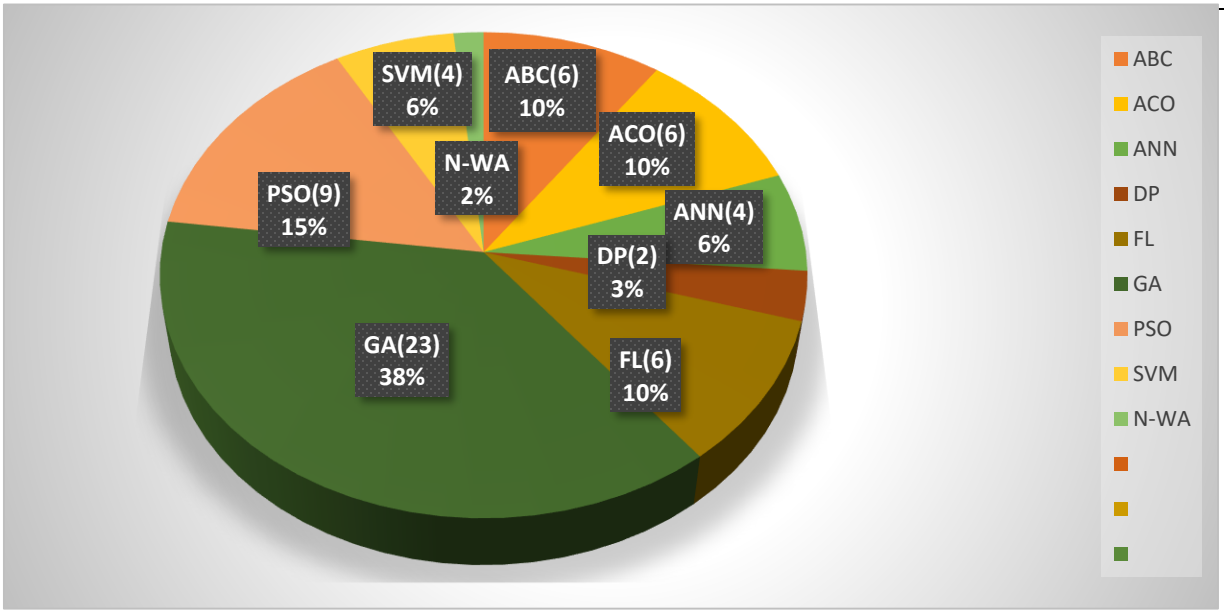


Figure 2: Distribution of the studies over type of ml technique.

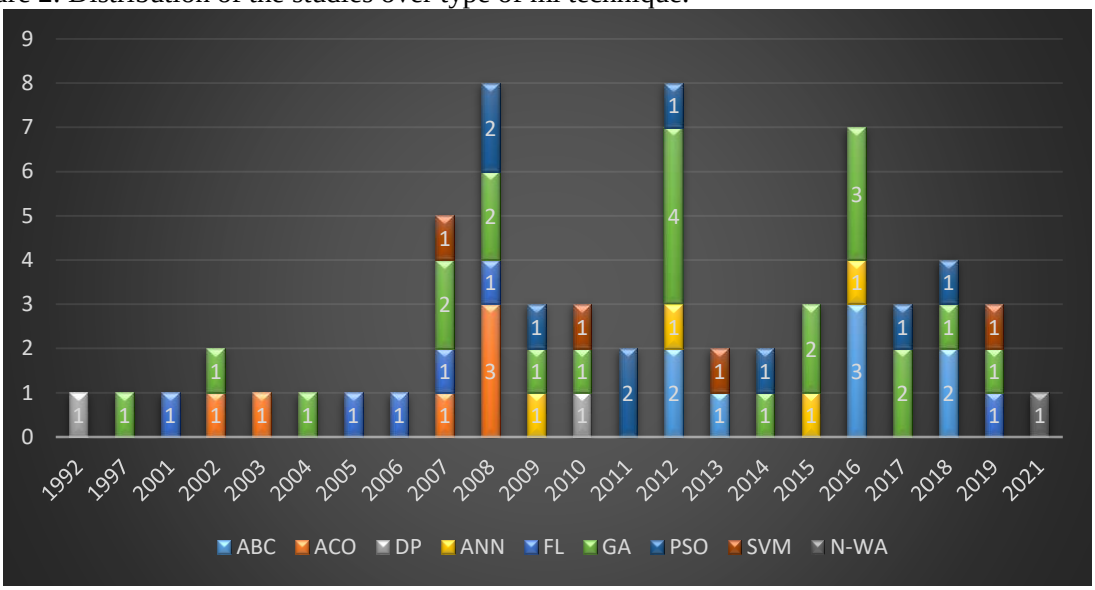


Figure 3: Distribution of the studies over publication year.

On the other hand, contrasted with other ML procedures, GA and PSO appear to have gotten prevailing exploration consideration in numerous years. Note that a few examinations encompass more than one ML approach. The recognized ML procedures were utilized for DNA sequence alignment generally in two structures: in alone or in blend. The blend structure might be gotten by consolidating at least 2 ML procedures or by joining ML strategies with non-ML methods. The run of the mill ML procedure that were frequently used to consolidate with other ML strategies are GA and fluffy rationale, individually. With respect to GA, ACO and ABC as per the chose examinations, it was seen as utilized in mix structure. The examinations detailing the utilization of GA in blend with other ML methods are, for instance, (GA with ACO), (GA with DP), (GA with ANN), and (GA with ABC).What we found about the ML strategies utilized in DNA sequence alignment space is exceptionally reliable with the discoveries of a few other significant survey works.

Table 7: SA Techniques and their relevant studies

Sr.	Techniques	Studies
1	Dynamic programming (DP)	[1], [31],
2	Artificial Neural Networks (ANN)	[29], [39], [61],
3	Support Vector Machine (SVM)	[2], [9], [17], [49]
4	Genetic Algorithms (GA)	[6], [8], [11], [13], [15], [21], [24], [28], [37], [38], [43], [46], [54], [58], [59], [66], [68], [74], [77], [80],
5	Fuzzy logic (FL)	[7], [19], [23], [47], [50], [72]
6	Artificial bee colony(ABC)	[14], [16], [32], [40], [62], [63], [76], [79],
7	Ant colony optimization(ACO)	[20], [22], [25], [51],
8	Particle swarm optimization (PSO)	[10], [12], [27], [34], [53], [67], [71], [75],
9	Hybrid	[18], [36], [48], [52], [64], [65], [69], [81], [82], [83]
10	Needleman-Wunsch Algorithm	[84]

B. Datasets used in machine learning sequence alignment papers (RQ2) We examined the related papers and their sources. Overall, 68 different datasets (public and private) were identified, used in DNA sequence alignment using machine learning techniques within the 84 papers addressing the same research questions. They are listed as follows.

- BAILBASE
- Protein sequencesTNFAIP2
- TRANSFAC
- OTHERSAmong the above listed datasets, BAILBASE are the most repeatedly used datasets; about 25% of the selected studies adopted BAILBASE, as illustrated in Fig.4. The detailed information about datasets Fig. 5 is plotted to additionally introduce the circulation of examination consideration in every distribution year.

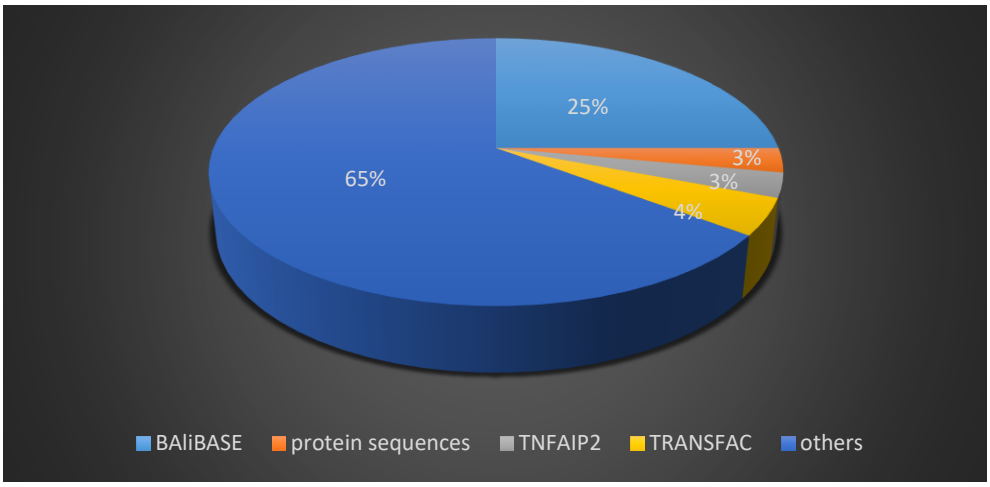


Figure 4: Distribution of the studies over type of datasets

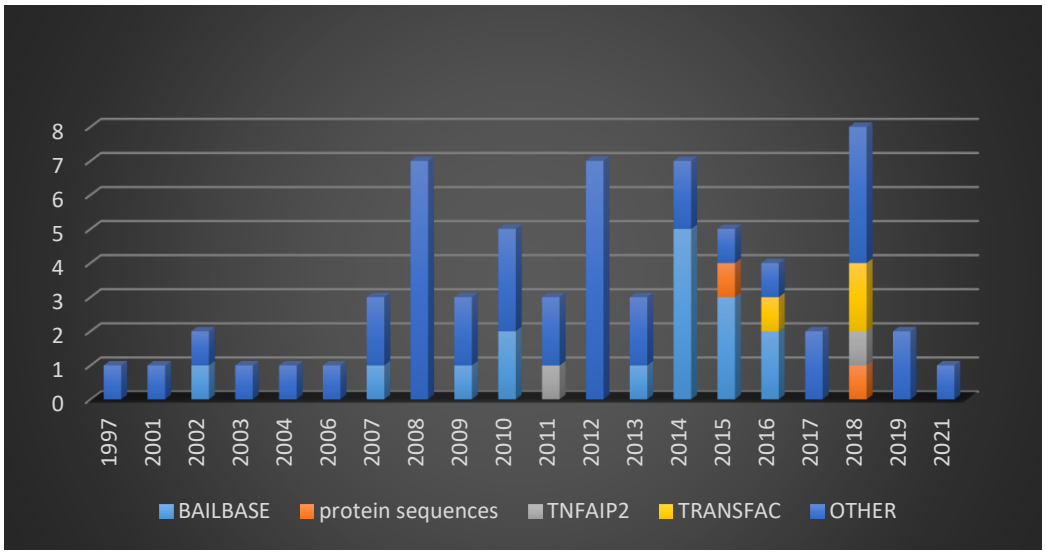


Figure 5: Utilized datasets per category

On the other hand, Benchmark Alignment Database - BaliBase (Thompson et al., 1999b; Bahr et al., 2001) appear to have gotten predominant examination consideration in numerous years. The first BaliBase (Thompson et al., 1999b) comprises of a lot of 142 reference alignments with more than 1000 sequences. Variant 2 of BaliBase (Bahr et al., 2001) improved a few alignments from the first database and stretched out it to 167 reference alignments and more than 2100 sequences incorporating sequences with rehashed areas, transmembrane sequences and round stages. BaliBase 2 is partitioned into eight classes of reference sets: 1) equidistant sequences with various degrees of preservation, 2) sequences with a profoundly disparate sequence, 3) bunches with under 25% personality, 4) sequences with N/C-terminal augmentations, 5) inner additions, 6) rehashes, 7) roundabout stages, and 8) transmembrane proteins (Bahr et al., 2001). Gathering 1 is partitioned by sequence sizes. From each gathering (and subgroups in bunch 1), sequences were arbitrarily chosen for the experiments, giving an aggregate of 32 as an agent test of the whole database. This gives a premise to assessment where an objective outcome is accessible for each arrangement of sequences. This outcome is the result of hand-curation, and doesn't speak to the ideal outcome dependent on the scoring frameworks utilized here. This implies a genuine worldwide improvement in the tests did for the most part need not to be replicated by complete alignment of the BaliBase arrangement (A straightforward hereditary calculation for various sequence alignment)

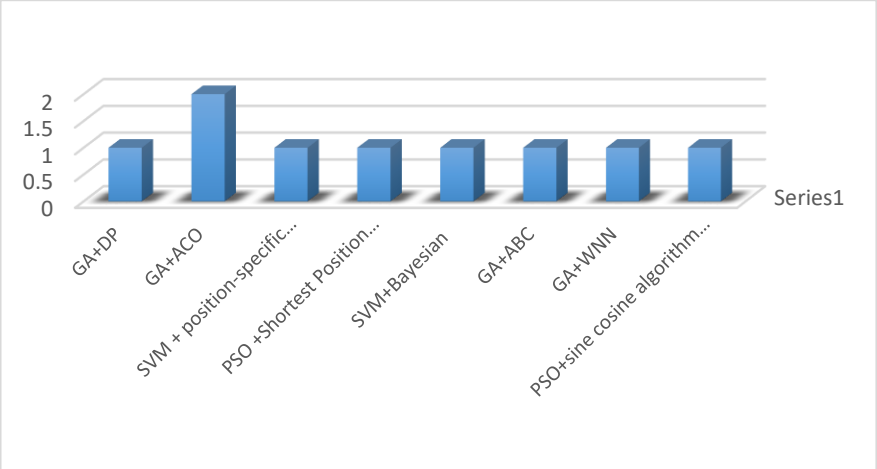
Table 8: Dataset and Relevant Articles

Sr. No.	Dataset	Studies used that dataset
1	BAILBASE	[4], [8], [13], [21], [32], [41], [42], [43], [58], [63], [64], [66], [67], [68], [71], [75], [76],
2	Protein Sequences	[59], [69], [81]
3	TNFAIP2	[78],
4	TRANSFAC	[14], [16], [79],

5	OTHERS	[2], [3], [5], [7], [9], [10], [11], [12], [17], [18], [19], [20], [22], [23], [24], [25], [26], [27], [28], [29], [31], [33], [35], [36], [39], [40], [44], [46], [49], [50], [51], [53], [54], [55], [56], [61], [65], [70], [72], [74], [77], [84]

C. Assembled techniques used in machinelearning in DNA sequence alignment

Different ML methods were utilized in blend with the AI strategies to beat a few difficulties related to DNA sequence alignment. Fig. 6 shows the assembled techniques used in DNA sequence alignment and shows that genetic algorithm and Ant colony optimization are the most regularly used practices (50%), in combination with ACO (25%), followed by SVM and PSO with 25% each.



.

Figure. 6. Distribution of techniques used in combination with MACHINE learning techniques

We investigated the use of GA in combination with other machine learning techniques in the selected studies of DNA sequence alignment. Solving a Sequence Alignment utilizing deterministic methodology is a mind-boggling streamlining task. Hereditary Algorithm (GA) is perhaps the most well-known and natural streamlining strategy. It depends on the component of characteristic development and the Darwinian hypothesis of common choice, and utilized in mix with relationship, for enhancing a mind boggling, enormous, as well as multidimensional issue and gives an ideal or a close ideal arrangement [5,10,39,40].

Table 9: Assemble SA Techniques and their Relevant Studies

Sr. No.	assembled techniques used in SA	Relevant Studies
1	Bayesian with SVM	[2],
2	(PSO) with Shortest Position Value (SPV)	[10],
3	GA and DP	[18],
4	ACO and GA	[48], [52],
5	SVM + position-specific scoring matrices	[49]
6	(GA-ABC)	[64]
7	(WNN)+GA	[65]
8	sine cosine algorithm (SCA) +(PSO)	[69]

D. Evaluation parameters used in machine learning sequence alignment papersWe identify the most regularly utilized assessment measurements in DNA sequence alignment utilizing AI strategies. We discovered 8 distinctive existing assessment measurements. These measurements were classified by discovery proficiency and computational execution. An aggregate of 74 papers talked about the assessment measurements. Altogether, 7 measurements have a place with this group:

- Accuracy

- Precision and recall:
- Time:
- Space:
- Sensitivity and specificity
- Quality
- Speed

Obviously, any individual measurement isn't a sufficient presentation meter of discovery effectiveness. For example, exactness now and again, slanted dataset, can prompt inclined outcomes in the presentation meter (Chawla, Bowyer et al. 2002) [N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321–357, 2002.]. The majority of the included papers assessed their models utilizing a few assessment measurements (see Fig. 7).

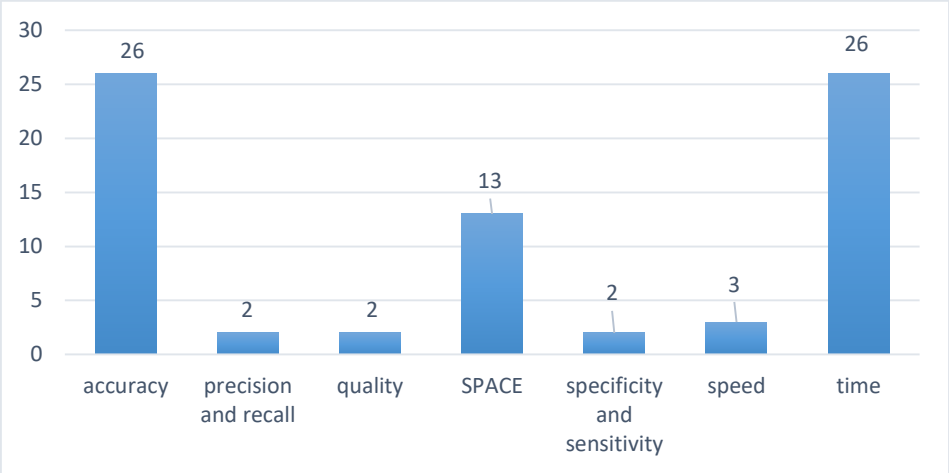


Figure 7. Detection efficiency metrics

As shown in Fig. 8, the accuracy (26), time (26), and space (13) are the most commonly used metrics to portion DNA sequence alignment using machine learning techniques detection effectiveness. To inspect the enhancement of evaluation metrics over time , we pursued the metrics to measure DNA sequence alignment using machine learning techniques. We saw the absence of thought given to certain measurements in the most recent decade including, in ascending order, speed, quality etc.

Figure 8. Popularity of efficiency metrics over time

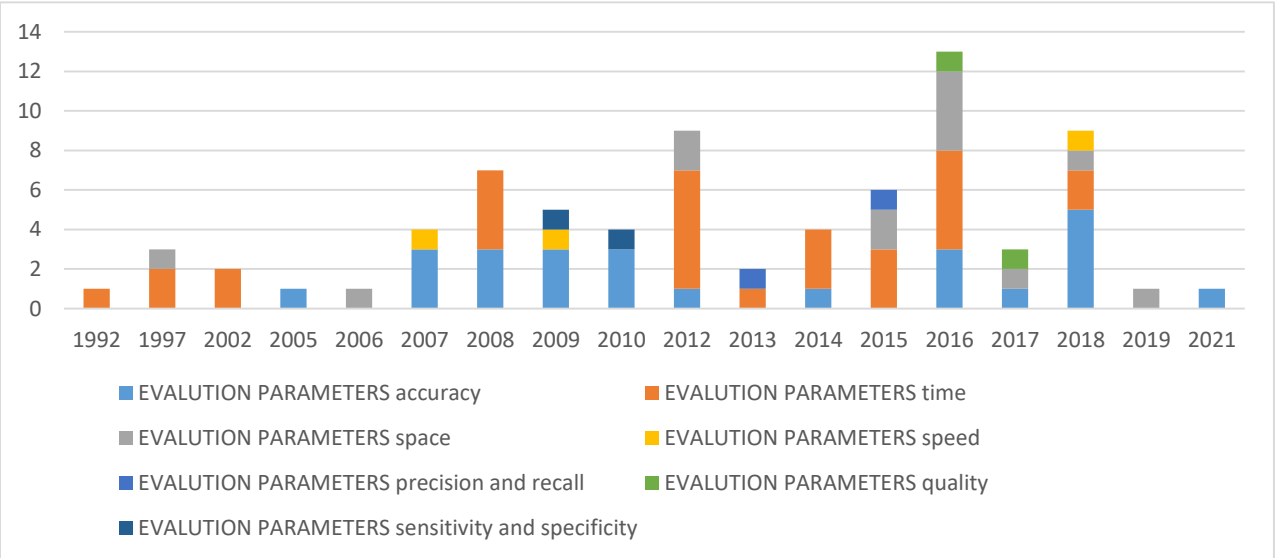


Table 10: Performance Metric and Relevant Articles

Sr. No.	Performance Metrics	Relevant Studies
1	Accuracy	[5], [11], [16], [23], [27], [28], [30], [32], [41], [47], [48], [49], [50], [53], [54], [58], [63], [64], [66], [68], [71], [76], [78], [82], [83], [84]
2	Precision and recall:	[2], [60],
3	Time	[1], [12], [20], [21], [22], [24], [25], [26], [35], [39], [40], [51], [55], [56], [61], [65], [74],
4	Space	[7], [46], [59], [80]
5	Both Time and Space	[3], [4], [15], [17], [18], [36], [37],
6	Sensitivity and specificity	[9], [29],
7	Quality	[42], [43],
8	Speed	[8], [75]
9	Size	[14], [44], [72], [79]
10	Time and Accuracy	[69]
11	Time and Speed	[70]

E. Machine learning techniques are used for multiple sequence alignment

From the chosen investigations, we recognized various kinds of ML methods that had been applied to multiple DNA sequence alignment. They are listed as follows.

- Dynamic programming (DP)
- MOMOSA
- Fast Fourier transform (FFT)
- Genetic Algorithms (GA)
- Fuzzy logic (FL)
- Artificial Bee Colony (ABC)
- Ant Colony Optimization (ACO)
- Particle Swarm Optimization (PSO)
- Needleman-Wunsch Algorithm

Among the above recorded ML methods, GA are the most often utilized one; they received by 46% of the chose examinations, as outlined in Fig. 9. Fig. 10 is plotted to additionally introduce the dispersion of exploration consideration in every distribution year. As appeared in Fig. 10, on one hand, a conspicuous distribution top shows up around year 2016.

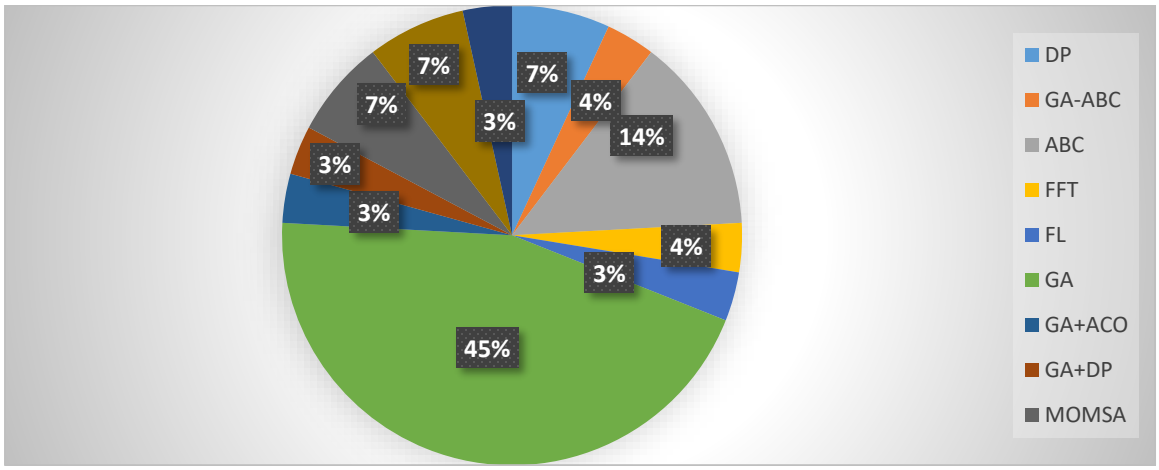


Figure 9: Distribution of the studies multiple SA using ML techniques

On the other hand, contrasted with other ML strategies, GA and ABC appear to have gotten predominant examination consideration in numerous years. Note that a few examinations comprise more than one ML method. The recognized ML methods were utilized for

different DNA sequence alignment ordinarily in two structures: in alone or in mix. The mix structure might be gotten by joining at least two ML strategies or by consolidating ML methods with other ML procedures. The run of the mill ML procedure that were regularly used to consolidate with other ML strategies are GA and DP, individually. With respect to GA, ACO and ABC as indicated by the chose investigations, it was seen as utilized in blend structure. The investigations detailing the utilization of GA in mix with other ML procedures are, for instance, (GA with ACO), (GA with DP) and (GA with ABC).

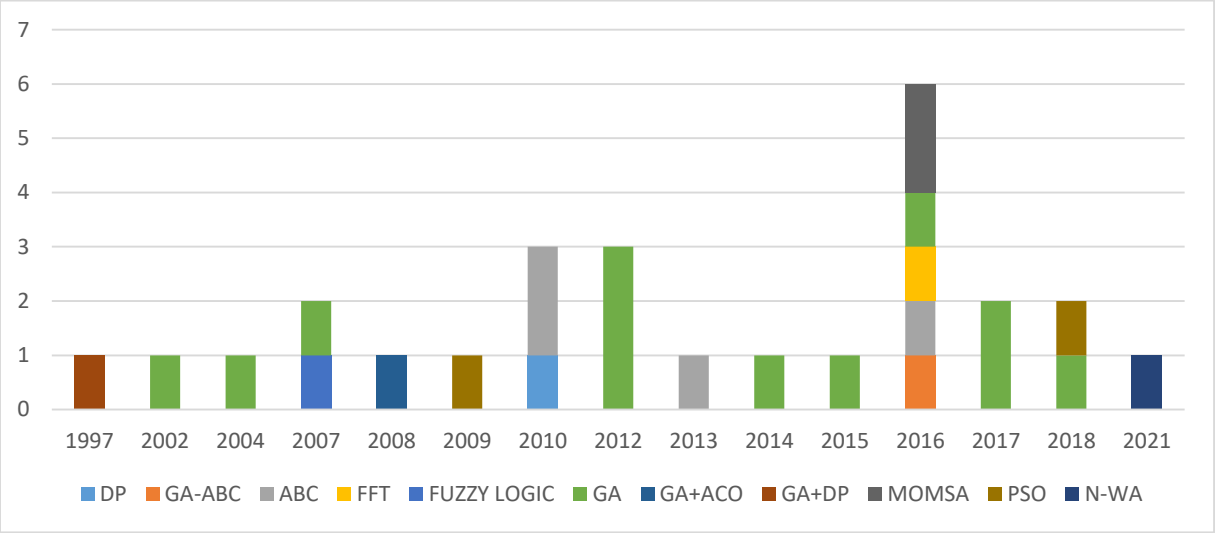


Figure 10: Distribution of the studies over publication year

F. Machine learning techniques are used for pairwise sequence alignment

From the chosen examinations, we recognized various kinds of ML procedures that had been applied to Pairwise DNA sequence alignment. They are listed as follows.

- Cellular Neural Network (CNN)
- Genetic Algorithms (GA)
- Flower Pollination Algorithm
- Artificial Bee Colony (ABC)
- Ant Colony Optimization (ACO)
- Particle Swarm Optimization (PSO)

Among the above listed ML techniques, GA are the most frequently used one and combine used with other machine learning techniques (GA with ACO); about 36% of the selected studies adopted the techniques, as illustrated in Fig. 11. Fig. 12 is plotted to additionally introduce the dissemination of examination consideration in every distribution year. As appeared in Fig. 12, on one hand, a conspicuous equivalent distribution top shows up around every year.

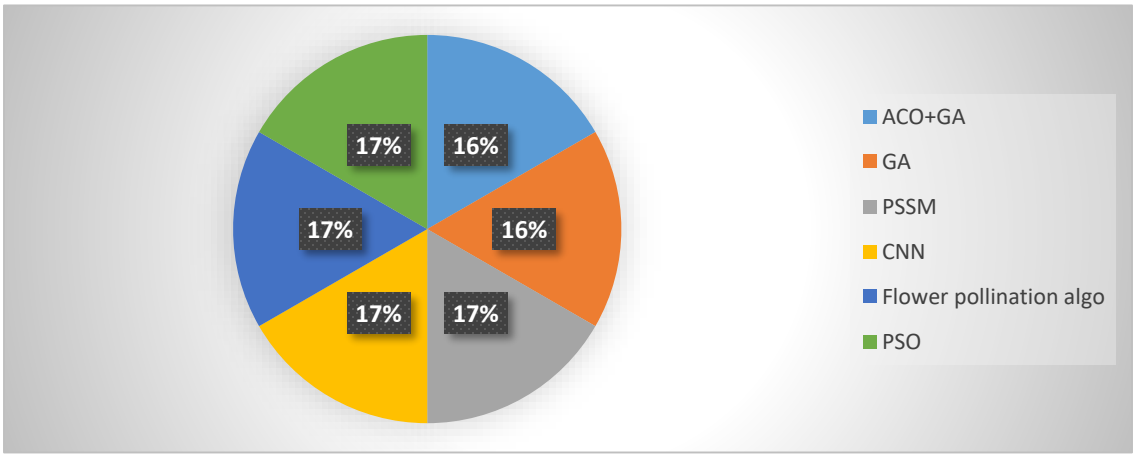


Figure 11: Distribution of the studies Pairwise SA using ML techniques

The identified ML techniques were used for pairwise DNA sequence alignment usually in two structures: in alone or in mix. The mix structure might be acquired by joining at least two ML strategies or by consolidating ML procedures with other ML methods. Concerning GA, ACO as per the chose examinations, it was seen as utilized in mix structure.

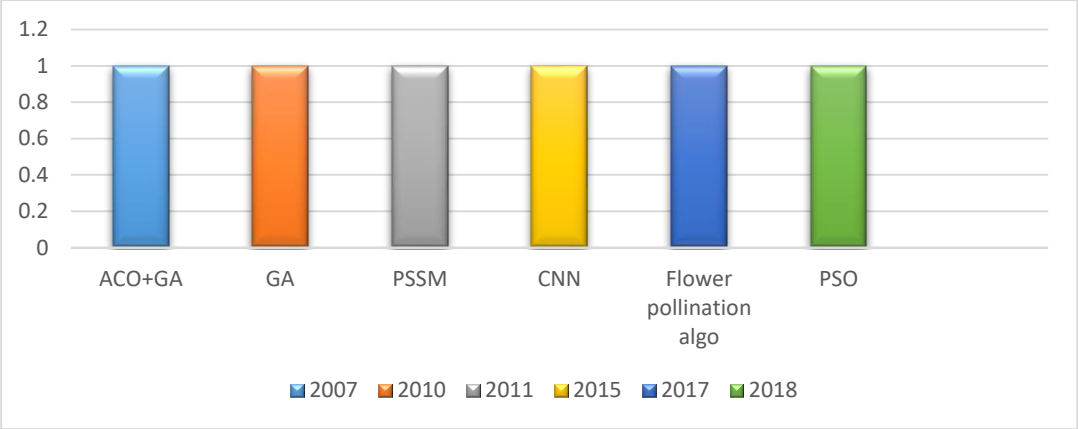


Figure 12: Distribution of the studies over publication year.

Repercussions for research and preparation:

This audit has discovered that the observational examinations on the utilization of SVM, ACO, GA, DP, PSO and ABC strategies. In this manner, scientists are urged to lead increasingly observational investigations on these ML methods to additionally fortify the experimental proof about their exhibition. In addition, analysts are additionally urged to investigate the conceivable outcomes of utilizing the ML strategies to evaluate programming improvement exertion. So as to search for the ML procedures and to utilize them all the more effectively, specialists would be advised to monitor the related teaches, for example, AI, information mining, measurements, and man-made reasoning, since these controls may give significant thoughts and techniques to address DNA sequence alignment issues. In spite of the fact that this audit has discovered that ML models are typically progressively precise and the ML model performs essentially and reliably superior to the current model.

III. Conclusion

The exponential development of the measure of organic information emerges issues: proficient data stockpiling and the board and the extraction of valuable data from this information. To take care of such issue AI strategies are utilized DNA sequence alignment. The DNA sequence alignment utilizing AI strategies in bioinformatics is turning out to be basic and significant issue for both scholastic exploration and in clinical field.

The best of our insight no precise writing survey beforehand has been distributed in the field of DNA sequence alignment utilizing AI methods in the bioinformatics research region. Hence, this SLR give valuable data to the AI strategies and general and DNA sequence alignment in bioinformatics as explicit in the field of SE.

In our paper, we talk about SLR of DNA sequence alignment utilizing machine learning procedures in bioinformatics to give composed arrangement to increasing high caliber of SLR to lead a fruitful precise writing survey before beginning it. The normal consequences of this audit will distinguish AI strategies utilized for DNA sequence alignment in bioinformatics. The normal advantages of this audit will give the analysts cutting edge of machine learning procedures utilized for DNA sequence alignment in an orderly manner.

Reference

[1] S. N. Atluri and S. Shen, "Global weak forms, weighted residuals, finite elements, boundary elements & local weak forms," in The Meshless Local Petrov-Galerkin (MLPG) Method, 1st ed., vol. 1. Henderson, NV, USA: Tech Science Press, 2004, pp. 15–64.

[2] S. N. Atluri, "The Meshless Method (MLPG) for Domain & BIE Discretization". Henderson, NV, USA: Tech Science Press, 2004. [Online]. Available: https://www.techscience.com/books/mlpg_atluri.html

[3] A. M. Farhan, "Effect of rotation on the propagation of waves in hollow poroelastic circular cylinder with magnetic field," Computers, Materials & Continua, vol. 53, no. 2, pp. 129–156, 2017.

[4] X. Chen and J. H. Jiang, "A method of virtual machine placement for fault-tolerant cloud applications," Intelligent Automation & Soft Computing, vol. 22, no. 4, pp. 587–597, 2016.

[5] X. F. Li, Y. B. Zhuang and S. X. Yang, "Cloud computing for big data processing," Intelligent Automation & Soft Computing, vol. 23, no. 4, pp. 545–546, 2017.

[6] L. Ali, R. Sidek, I. Aris and M. A. M. Ali, "Design of a testchip for low cost IC testing," Intelligent Automation & Soft Computing, vol. 15, no. 1, pp. 63–72, 2009.

[7] J. Cheng, R. M. Xu, X. Y. Tang, V. S. Sheng and C. T. Cai, "An abnormal network flow feature sequence prediction approach for DDoS attacks detection in big data environment," Computers, Materials & Continua, vol. 55, no. 1, pp. 95–119, 2018.

[8] W. J. Yang, P. P. Dong, W. S. Tang, X. P. Lou, H. J. Zhou et al., "A MPTCP scheduler for web transfer," Computers, Materials & Continua, vol. 57, no. 2, pp. 205–222, 2018.

[9] A. Abdelmaboud, D. N.Jawawi, I. Ghani, A. Elsafi and B. Kitchenham, "Quality of service approaches in cloud computing: A systematic mapping study," Journal of systems and software, vol. 101, pp. 159-179, 2015.

[10] D. Ebehard and E. Voges, "Digital single sideband detection for interferometric sensors," presented at the 2nd Int. Conf. Optical Fiber Sensors, Stuttgart, Germany, Jan. 2-5, 1984.

[11] P. Agarwal, R. Gupta, T. Maheswari, P. Agarwal, S. Yadav et al., "A Genetic Algorithm for Alignment of Multiple DNA Sequences," presented at the Int. Conf. Advances in Communication, Network, and Computing, Springer. 2012.

[12] S.A.M. Al Junid, Z Abd Majid and A.K. Halim, "High speed DNA sequencing accelerator using FPGA," presented in Int. Conf. on Electronic Design, IEEE, 2008.

[13] S.A.M. Al Junid, M.S. Reffin, Z. Abd Majid, N.M. Tahir and M.A. Haron, "Implementation of genetic algorithm for optimizing DNA sequence alignment," presented in IEEE Business, Engineering & Industrial Applications Colloquium (BEIAC), IEEE, 2012.

[14] A.R. Amorim, L.A. Neves, C.R. Valêncio, G.F. Roberto and G.F.D Zafalon, "An approach for COFFEE objective function to global DNA multiple sequence alignment," Computational biology and chemistry, vol. 75, pp. 39-44, 2018.

[15] S. Ando and H. Iba, "Ant algorithm for construction of evolutionary tree, Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600), IEEE, 2002.

[16] A. Arribas-Gil, D. Metzler and J.L Plouhinec, "Statistical alignment with a sequence evolution model allowing rate heterogeneity along the sequence," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 6, no. 2, pp. 281-295, 2008.

[17] C. Bi, "Deterministic local alignment methods improved by a simple genetic algorithm," Neurocomputing, vol 73, no. 13-15, pp. 2394-2406, 2010.

[18] A. Bir, J. Dongardive, S. Jamkhedkar and S. Abraham, "Building Consensus of Human Papillomavirus using Genetic Algorithm," 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), IEEE, 2009.

[19] C. Blum, M. Y. Vallès and M.J. Blesa, "An ant colony optimization algorithm for DNA sequencing by hybridization," Computers & Operations Research, vol. 35, no. 11, pp. 3620-3635, 2008.

[20] P. Borovska, V. Gancheva and N. Landzhev, "Massively parallel algorithm for multiple biological sequences alignment," presented at 36th International Conference on Telecommunications and Signal Processing (TSP), IEEE, 2013.

[21] P. Brereton, B.A. Kitchenham, D. Budgen, M. Turner and M Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," Journal of systems and software, vol. 80, no. 4, pp.571-583, 2007.

[22] Chang, B. C. and S. K. Halgamuge (2001). Fuzzy sequence pattern matching in zinc finger domain proteins. Proceedings Joint 9th IFSA World Congress and 20th NAFIPS International Conference (Cat. No. 01TH8569), IEEE.

[23] N.V. Chawla, K.W. Bowyer, L.O. Hall and W.P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of artificial intelligence research, vol. 16, pp. 321-357, 2002.

[24] B. Chowdhury and G. Garai, "A review on multiple sequence alignment from the perspective of genetic algorithm," Genomics, vol. 109, no. 5-6, pp. 419-431, 2017.

[25] J. Cohen, "Bioinformatics—an introduction for computer scientists," ACM Computing Surveys (CSUR), vol. 36, no. 2, pp. 122-158, 2004.

[26] A. Dakhli and W. Bellil, "Wavelet neural networks for DNA sequence classification using the genetic algorithms and the least trimmed square," Procedia Computer Science, vol. 96, pp. 418-427, 2016.

[27] D. Eppstein, Z. Galil, R. Giancarlo and G. F. Italiano, "Sparse dynamic programming I: linear cost functions," Journal of the ACM (JACM), vol. 39, no. 3, pp. 519-545, 1992.

[28] Z. Ezziane, "DNA computing: applications and challenges." Nanotechnology, vol. 17, no. 2, pp. R27, 2005.

[29] T. Fukunishi, A. Finch, S. Yamamoto and E. Sumita, "A Bayesian alignment approach to transliteration mining," ACM Transactions on Asian Language Information Processing (TALIP), vol. 12, no 3, pp. 1-22, 2013.

[30] G. Garai and B. Chowdhury, "A cascaded pairwise biomolecular sequence alignment technique using evolutionary algorithm," Information Sciences, vol. 297, pp. 118-139, 2015.

[31] Goel, Neelam, Shailendra Singh, and Trilok Chand Aseri. "A review of soft computing techniques for gene prediction." International Scholarly Research Notices 2013 (2013).

[32] C. Gondro and B. P. Kinghorn, "A simple genetic algorithm for multiple sequence alignment," Genetics and Molecular Research, vol. 6, no. 4, pp. 964-982, 2007.

[33] R. Gupta, P. Agarwal and A.K. Soni, "Genetic algorithm based approach for obtaining alignment of multiple sequences, " International Journal of Advanced Computer Science & Applications (IJACSA), vol. 3, no. 12, 2012.

[34] T. Halgaswaththa, A.S. Atukorale, M. Jayawardena and J. Weerasena, "Neural network based phylogenetic analysis," in proc. 2012 International Conference on Biomedical Engineering (ICoBE), IEEE, 2012.

[35] T. Halgaswaththa, A.S. Atukorale, M. Jayawardena and J. Weerasena, "Pairwise sequence alignment algorithms: a survey," in Proc. of the 2009 conference on Information Science, Technology and Applications, 2009.

[36] T. Halgaswaththa, A.S. Atukorale, M. Jayawardena and J. Weerasena, "Computational intelligence techniques in bioinformatics," Computational biology and chemistry, vol. 47, no. pp. 37-47, 2013.

[37] S.Y. Ho, F.C. Yu, C.Y. Chang and H.L. Huang, "Design of accurate predictors for DNA-binding sites in proteins using hybrid SVM–PSSM method," Biosystems, vol. 90, no. 1, pp. 234-241, 2007.

[38] J.T. Horng, L.C. Wu, C.M. Lin and B.H. Yang, "A genetic algorithm for multiple sequence alignment," Soft Computing, vol. 9, no. 6, pp. 407-420, 2005.

[39] K.W. Huang, J.L. Chen, C.S. Yang and C.W. Tsai, "A memetic particle swarm optimization algorithm for solving the DNA fragment assembly problem," Neural Computing and Applications, vol. 26, no. 3, pp. 495-506, 2015.

[40] A. Idri, A. F. azzahra and A. Abran, "Analogy-based software development effort estimation: A systematic mapping and review," Information and Software Technology, vol. 58, pp. 206-230, 2015.

[41] M. Ishaq, A. Khan, M. Khan and M. Imran, "Current Trends and Ongoing Progress in the Computational Alignment of Biological Sequences," IEEE Access, vol. 7, pp. 68380-68391, 2019.

[42] M. Issa, A.E. Hassanien, D. Oliva, A. Helmi, I. Ziedan and A. Alzohairy, "ASCA-PSO: Adaptive sine cosine optimization algorithm integrated with particle swarm for pairwise local sequence alignment." Expert Systems with Applications, vol. 99, pp. 56-70, 2018.

[43] S. R Jangam. and N. Chakraborti, "A novel method for alignment of two nucleic acid sequences using ant colony optimization and genetic algorithms," Applied Soft Computing, vol. 7, no. 3, pp. 1121-1130, 2007.

[44] L. Ji, X. Pu, H. Qu and G. Liu, "One-dimensional pairwise CNN for the global alignment of two DNA sequences," Neurocomputing, vol. 149, pp. 505-514, 2015.

[45] U. Kanewala and J. M. Bieman, "Testing scientific software: A systematic literature review," Information and Software Technology, vol 56, no. 10, pp. 1219-1232, 2014.

[46] D. Karaboga and S. Aslan, "A discrete artificial bee colony algorithm for detecting transcription factor binding sites in DNA sequences," Genet Mol Res, vol. 15, no. 2, pp. 1-11, 2016.

[47] D. Karaboga and S. Aslan "Discovery of conserved regions in DNA sequences by

Artificial Bee Colony (ABC) algorithm based methods," Natural Computing, vol. 18, no. 2, pp. 333-350, 2019.

[48] Y. Kaur and N. Sohi, "Pairwise sequence alignment method using flower pollination algorithm," in proc. 2017 4th International Conference on Signal Processing, Computing and Control (ISPCC), IEEE, 2017.

[49] M. Kaya, A. Sarhan and R. Alhajj, "Multiple sequence alignment with affine gap by using multi-objective genetic algorithm," Computer methods and programs in biomedicine, vol. 114, no. 1, pp. 38-49, 2014.

[50] R. Khattree and D. Naik "Machine learning techniques for bioinformatics" Computational Methods in Biomedical Research, pp. 57-88. Chapman and Hall/CRC, 2007.

[51] K. Kim, M. Kim and Y. Woo, "A DNA sequence alignment algorithm using quality information and a fuzzy inference method." Progress in Natural Science, vol. 18, no. 5, pp. 595-602, 2008.

[52] B. Kitchenham, (). "Procedures for performing systematic reviews, " Keele, UK, Keele University, vol. 33, pp. 1-26, 2004.

[53] D.C. Koboldt, K.M. Steinberg, D.E. Larson, R.K. Wilson and E.R. Mardis, "The next-generation sequencing revolution and its impact on genomics, " Cell, vol. 155, no. 1, pp. 27-38, 2013.

[54] M. Kumar, "An enhanced algorithm for multiple sequence alignment of protein sequences using genetic algorithm," EXCLI journal, vol 14, pp. 1232, 2015.

[55] S. Kumar, K. Tamura and M. Nei, "MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment," Briefings in bioinformatics, vol. 5, no. 2, pp. 150-163, 2004.

[56] A. Lacey and X. Xie, "Supervised Machine Learning Techniques in Bioinformatics: Protein Classification." (2014).

[57] J. Lee, Y. Yeu, H. Roh, Y. Yoon and S. Park, "BulkAligner: A novel sequence alignment algorithm based on graph theory and Trinity," Information Sciences, vol 303, pp. 120-133, 2015.

[58] Z.J. Lee, S.F. Su, C.C. Chuang and K.H. Liu, "Genetic algorithm with ant colony optimization (GA-ACO) for multiple sequence alignment," Applied Soft Computing, vol. 8, no. 1, pp. 55-78, 2008.

[59] Lei, C. and J. Ruan, "A particle swarm optimization algorithm for finding DNA sequence motifs," in proc. 2008 IEEE International Conference on Bioinformatics and Biomeidcine Workshops, IEEE, 2008.

[60] X. Lei, J. Sun, X. Xu and L. Guo, "Artificial bee colony algorithm for solving multiple sequence alignment," in proc. 2010 IEEE fifth international conference on bio-inspired computing: theories and applications (BIC-TA), IEEE, 2010.

[61] C.S. Leslie, E. Eskin, A. Cohen, J. Weston and W.S. Noble, "Mismatch string kernels for discriminative protein classification," Bioinformatics, vol. 20, no. 4, pp. 467-476, 2004.

[62] G. Li, Y. Wang and X. Su, "Improvements on a privacy-protection algorithm for DNA sequences with generalization lattices." Computer methods and programs in biomedicine, vol. 108, no. 1, pp. 1-9, 2012.

[63] M.S.I. Mahmud, M.A. Hosen, M. Saroer-E-Azam, M.A. Mottalib and H.A Al-Mamun, "A novel two-tier multiple sequence alignment algorithm," in proc. 2010 13th International Conference on Computer and Information Technology (ICCIT), IEEE, 2010.

[64] A. Majid, M. Khan, N. Iqbal, M.A. Jan and M. Khan, "Application of parallel vector space model for large-scale dna sequence analysis," Journal of Grid Computing, vol. 17, no. 2, pp. 313-324, 2019.

[65] P. Meksangsouy and N. Chaiyaratana, "DNA fragment assembly using an ant colony system algorithm," in porc. The 2003 Congress on Evolutionary Computation, 2003. CEC'03., IEEE, 2003.

[66] A.L. Caicedo and M.D. Purugganan, "Comparative genomics," Annu. Rev. Genomics Hum. Genet, vol. 5, pp. 15-56, 2005.

[67] N.S. Mohamed, Z.A. Othman and A.A. Bakar, "A classification of "Gracilaria changii" protein sequences using back-propagation classifier," in proc. 2009 2nd Conference on Data Mining and Optimization, IEEE, 2009.

[68] P. Mohanty and S. Tragoudas, "Scalable Offline Searches in DNA Sequences," ACM Journal on Emerging Technologies in Computing Systems (JETC), vol. 11, no. 2, pp. 1-25, 2014.

[69] N. Moustafa, M. Elhosseini, T.H. Taha and M. Salem, "Fragmented protein sequence alignment using two-layer particle swarm optimization (FTLPSO)," Journal of King Saud University-Science, vol. 29, no. 2, pp.191-205, 2017.

[70] A. Nagar and M. Hahsler, "A novel quasi-alignment-based method for discovering conserved regions in genetic sequences," in proc. 2012 IEEE International Conference on

Bioinformatics and Biomedicine Workshops, IEEE, 2012.

[71] Nasser, S., et al. (2007). Multiple sequence alignment using fuzzy logic. 2007 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology, IEEE.

[72] Nguyen, H. D., et al. (2002). A parallel hybrid genetic algorithm for multiple protein sequence alignment. Proceedings of the 2002 Congress on Evolutionary Computation. CEC'02 (Cat. No. 02TH8600), IEEE.

[73] Nikolayevskyy, Vlad, Katharina Kranzer, Stefan Niemann, and Francis Drobniewski. "Whole genome sequencing of Mycobacterium tuberculosis for detection of recent transmission and tracing outbreaks: A systematic review." Tuberculosis 98 (2016): 77-85.

[74] Othman, Mohamed Tahar Ben. "Survey of the use of genetic algorithm for multiple sequence alignment." Journal of Advanced Computer Science & Technology 5, no. 2 (2016): 28..

[75] Othman, Mohamed Tahar Ben, and Gamil Abdel-Azim. "Multiple sequence alignment based on genetic algorithms with new chromosomes representation." In 2012 16th IEEE Mediterranean Electrotechnical Conference, pp. 1030-1033. IEEE, 2012.

[76] Pan, Yi, Jianxin Wang, and Min Li. "Wiley Series on Bioinformatics: Computational Techniques and Engineering." (2014): 513-514.

[77] Pentheroudakis, George, F. A. Greco, and Nicholas Pavlidis. "Molecular assignment of tissue of origin in cancer of unknown primary may not predict response to therapy or outcome: a systematic literature review." Cancer treatment reviews 35, no. 3 (2009): 221-227.

[78] Pevsner, J. "Completed Genomes: Bacteria and Archaea." Bioinformatics and Functional Genomics, Second Edition, John Wiley & Sons, Inc., Hoboken, NJ, USA. doi 10 (2009): 9780470451496.

[79] Polato, Ivanilton, Reginaldo Ré, Alfredo Goldman, and Fabio Kon. "A comprehensive view of Hadoop research—A systematic literature review." Journal of Network and Computer Applications 46 (2014): 1-25.

[80] Qian, Weiyi, Yingjie Yang, Ningning Yang, and Chun Li. "Particle swarm optimization for SNP haplotype reconstruction problem." Applied mathematics and Computation 196, no. 1 (2008): 266-272.

[81] Rajapakse, J. C., and I. Faleel. "Genetic approach to biosequence alignment (GABA)." In Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02., vol. 2, pp. 611-615. IEEE, 2002..

[82] Rajasekhar, Anguluri, Nandar Lynn, Swagatam Das, and Ponnuthurai N. Suganthan. "Computing with the collective intelligence of honey bees–a survey." Swarm and Evolutionary Computation 32 (2017): 25-48.

[83] Ramaswamy, Harish, and Carla Purdy. "An extended library of hardware modules for genetic algorithms, with applications to DNA sequence matching." In 2008 51st Midwest Symposium on Circuits and Systems, pp. 209-212. IEEE, 2008.

[84] Rani, R. Ranjani, and D. Ramyachitra. "Multiple sequence alignment using multi-objective based bacterial foraging optimization algorithm." Biosystems 150 (2016): 177-189.

[85] Rani, R. Ranjani, and D. Ramyachitra. "Application of genetic algorithm by influencing the crossover parameters for multiple sequence alignment." In 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON), pp. 33-38. IEEE, 2017.

[86] Ressom, H., Padma Natarajan, Rency S. Varghese, and Mohamad T. Musavi. "Applications of fuzzy logic in genomics." Fuzzy sets and systems 152, no. 1 (2005): 125-138.

[87] Rubio-Largo, Alvaro, Leonardo Vanneschi, Mauro Castelli, and Miguel A. Vega-Rodríguez. "Swarm intelligence for optimizing the parameters of multiple sequence aligners." Swarm and Evolutionary Computation 42 (2018): 16-28.

[88] Rubio-Largo, Álvaro, Miguel A. Vega-Rodríguez, and David L. González-Álvarez. "Hybrid multiobjective artificial bee colony for multiple sequence alignment." Applied Soft Computing 41 (2016): 157-168.

[89] Saw, Ajay Kumar, Garima Raj, Manashi Das, Narayan Chandra Talukdar, Binod Chandra Tripathy, and Soumyadeep Nandi. "Alignment-free method for DNA sequence clustering using Fuzzy integral similarity." Scientific reports 9, no. 1 (2019): 1-18.

[90] Simossis, Victor, Jens Kleinjung, and Jaap Heringa. "An overview of multiple sequence alignment." Current protocols in bioinformatics 3, no. 1 (2003): 3-7.

[91] Siswanto, Teddy, Spits Warnars Harco Leslie Hendric, Harjanto Prabowo, Nesti Fronika Sianipar, Bahtiar Saleh Abbas, and Achmad Nizar Hidayanto. "The genomic plant warehouse framework: A systematic literature review." In 2017 International Conference on

Information Management and Technology (ICIMTech), pp. 244-248. IEEE, 2017.
[92] Smith, Anna Jo, Elizabeth L. Turner, and Sanjay Kinra. "Universal cholesterol screening in childhood: a systematic review." Academic pediatrics 16, no. 8 (2016): 716-725.
[93] Surendar, A., Sadulla Shaik, and N. Usha Rani Rani. "Micro Sequence Identification of DNA Data Using Pattern Mining Techniques." Materials Today: Proceedings 5, no. 1 (2018): 578-587.
[94] Torres, Angela, and Juan J. Nieto. "Fuzzy logic in medicine and bioinformatics." journal of Biomedicine and Biotechnology 2006 (2006).
[95] van Karnebeek, Clara DM, and Sylvia Stockler. "Treatable inborn errors of metabolism causing intellectual disability: a systematic literature review." Molecular genetics and metabolism 105, no. 3 (2012): 368-381.
[96] Verma, Ravi Shankar. "DSAPSO: DNA sequence assembly using continuous particle swarm optimization with smallest position value rule." In 2012 1st international conference on recent advances in information technology (RAIT), pp. 410-415. IEEE, 2012.
[97] Verma, Ravi Shankar, Vikas Singh, and Sanjay Kumar. "Dna sequence assembly using particle swarm optimization." International Journal of Computer Applications 28, no. 10 (2011): 33-38.
[98] Wang, Liangjiang, Caiyan Huang, Mary Qu Yang, and Jack Y. Yang. "BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features." BMC Systems Biology 4, no. 1 (2010): 1-9.
[99] Wen, Jianfeng, Shixian Li, Zhiyong Lin, Yong Hu, and Changqin Huang. "Systematic literature review of machine learning based software development effort estimation models." Information and Software Technology 54, no. 1 (2012): 41-59.
[100] Wu, Jie, and Yiqiang Zhao. "Machine learning technology in the application of genome analysis: a systematic review." Gene 705 (2019): 149-156.
[101] Xu, Fasheng, and Yuehui Chen. "A method for multiple sequence alignment based on particle swarm optimization." In International Conference on Intelligent Computing, pp. 965-973. Springer, Berlin, Heidelberg, 2009..
[102] Xu, Xiaojun, and Xiujuan Lei. "Multiple sequence alignment based on abc_sa." In International Conference on Artificial Intelligence and Computational Intelligence, pp. 98-105. Springer, Berlin, Heidelberg, 2010.
[103] Yu, J. (2011). Solving sequence alignment based on chaos particle swarm optimization algorithm. 2011 International Conference on Computer Science and Service System (CSSS), IEEE.
[104] Zhang, Ching, and Andrew KC Wong. "Toward efficient multiple molecular sequence alignment: a system of genetic algorithm and dynamic programming." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 27, no. 6 (1997): 918-932.
[105] Zhang, Zheng, Scott Schwartz, Lukas Wagner, and Webb Miller. "A greedy algorithm for aligning DNA sequences." Journal of Computational biology 7, no. 1-2 (2000): 203-214.
[106] Zhao, Y., et al. (2008). An improved ant colony algorithm for DNA sequence alignment. 2008 International Symposium on Information Science and Engineering, IEEE.
[107] Zheng, Weihua, Kenli Li, Keqin Li, and Hing Cheung So. "A modified multiple alignment fast Fourier transform with higher efficiency." IEEE/ACM transactions on computational biology and bioinformatics 14, no. 3 (2016): 634-645..
[108] Zou, Quan, Xiao Shan, and Yi Jiang. "A novel center star multiple sequence alignment algorithm based on affine gap penalty and k-band." Physics Procedia 33 (2012): 322-327.

Appendix A. Example of appendix

Authors that need to include an Appendix section should place it after the References section. Multiple appendices should all have headings in the style used for above. They should be ordered as such: A, B, and C etc.

[109] Amr Ezz El-Din Rashed1, Hanan Abdelfatah1, Mervat El-Seddek2, and Hossam El~Din Mou