JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1130 - 1136
https://publishoa.com
ISSN: 1309-3452

Classification Techniques for Medicinal Databases Using Auto-Regression and Firefly Algorithm

M. Krishnamoorthy
[1]Research Scholar
Department of Computer Science and Engineering
Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
krishnamoorthymuniyan@gmail.com,

Dr. R. Karthikeyan
[2]Associate Professor
Department of Computer Science and Engineering
Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
rkarthikeyan1678@gmail.com,

ABSTRACT
The life form's genome contains all the hereditary data encoded by DNA sequences in medicinal database. Clarifying the genome is vital to decide the endurance, improvement and multiplication of a creature. As of late, entire genome groupings develop dramatically over the long haul. Likewise, to pack an enormous genome, we really want a substantial calculation and a legitimate calculation. In this point, the previous framework addresses a technique in light of programmed relapse demonstrating, the model not set in stone by particle swarm optimization (PSO) and the low pressure proportion for positive thinking of the PSO part. To improve, Hybrid streamlining of hospitalization-subordinate DNA pressure (HOARADNACOM) has been proposed and utilizes delicate registering strategies. This calculation works in even mode utilizing an AR-based replacement measurements strategy, and the model not entirely settled by the modified FireFly Algorithm (MFR). The essential firefly calculation fixes fireflies on the grounds that the most brilliant sparkles move haphazardly, bringing about horrible showing at specific emphases. If this splendid FireFluster can move in the improved bearing to that splendor, the exhibition of the calculations related with the best worldwide answers for all arrangements will endure. There isn't anything in this beneficial arrangement. The proposed framework expects to accomplish a bigger pressure proportion, alongside the usefulness of the succession got from pressure, which favors distance to diminish capacity, investigation, transmission upward, and air structure. Execution results show that the proposed MFF calculation can accomplish preferred pressure proportions over existing exploration techniques.

Keywords—DNA Sequence Compression, De-Compression, Compression Ratio, Nucleidoid Base, AutoRegression, Firefly Algorithm, Optimal Prediction, Soft Computing Techniques.

I INTRODUCTION

The size of the genome data set is a yearly expansion in ceaseless rate. Great many nucleotides are sequenced every day in the research facility. From 1982 to the present, the quantity of GenBank areas has nearly multiplied like clockwork. This is the quantity of 606 grouping records and the quantity of 606 succession records comparing to 1982 680338, and in April 2002, the quantity of 406 arrangement records in 2002 and the quantity of 406 succession records in 2002 and their number. Being equivalent to the quantity of bases has been noticed [1]. The Genbank arrangement set expanded likewise to 16769983 and 16769983, and the WG contained 692266338 bases and 172768 groupings. Afterward, renditions 206 and 2015 had an enormous number of quality banks and WGS bases, and various successions 181336445, with cluster quantities of 873281414087 and 205465046 and 205465046, separately. The quantity of individual genomes that vary somewhat from one another for a specific living being. Such a verbose and tremendous assortment of groupings presents clear pressure issues. Packed information limits correspondence and capacity upward. Likewise, you can utilize compacted arrangements to keep up with succession similitude. An exceptionally verbose DNA grouping has a few great elements [2] and [3] that can be utilized to pack it. The DNA succession contains ASEXON (ie, coding district or protein union) or intron (ie, non-coding locale or protein blend) A, C, G, and four nucleotide bases A, C, G, situated all things considered Ts. Furthermore, T. Notwithstanding this reality, the capacity rich instances of any base are "pressure", "GZIP", "BZIP2", "WinZip", or "WinRAR" that utilization multiple pieces in [4]. The more static and versatile Huffman coding can't be debased through the DNA succession, as these images are very liable to happen. This assignment centers around compacting that specific information. DNA is basically a twofold receptive particle, with neighboring chains associated by hydrogen connections between the bases. This hydrogen bond is available in the chain where adenine (A/A) is available in the other chain, and guanine (G/g) present in the chain present in the chain prompts the occurrence of cytosine (C/C). It is likewise a chain of the inverse. Every pressure calculation packs just a single string. These tasks are fundamental to essentially expand the size of DNA records and lead to open doors for new pressure strategies that

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1130 - 1136
https://publishoa.com
ISSN: 1309-3452

exploit this new information include. The inspiration driving this undertaking is to perceive this overt repetitiveness and track down the pressure procedure for the best similitude inside the singular request. DNA is the primary human plan since it has a succession of deoxyribonucleic corrosive (DNA).

II RELATEDWORKS

Life has areas of strength for association and construction [5]. At the point when the 1000 Genomes Project is finished, the undertaking will produce roughly 8.2 billion bases each day, and the total grouping is assessed to surpass 6 trillion nucleotide bases. The DNA particle is made out of a mix of four nucleotides: adenine, thymine, cytosine, and guanine (A, T, C, G). Universally useful pressure calculations don't function admirably with natural groupings. In [6], we checked on the pressure calculations produced for organic groupings. Finding characteristics and contrasting genomes is a troublesome undertaking [7], [8]. According to a numerical perspective, pressure implies understanding and translation [9]. Pressure is an effective apparatus used to look at genomes and inspect different properties of the genome. DNA arrangements that broaden the existence of the code should be compressible. It is notable that DNA arrangements present in higher eukaryotes contain different couple rehashes, and fundamental qualities (like rRNA) contain various duplicates. It has likewise been shown that qualities might make their own imitations for developmental purposes. This multitude of realities shows that the DNA grouping should be compressible. Pressure of DNA successions isn't simple [10], [11], [12]. The DNA succession comprises of just four nucleotide bases {a, c, g, t}. Two pieces are sufficient to store each base. Standard pressure programming, for example, "pack", "gzip", "bzip2", and "winzip" have extended as opposed to compacting DNA genome documents. A large number of the accessible programming devices functioned admirably for English text pressure [13], yet were not reasonable for the DNA genome. Because of the expansion in genomic succession information of life forms, the size of DNA data sets has expanded a few times every year. Thusly, it is undeniably challenging to download and keep up with the information on your neighborhood framework. Different calculations explicitly intended to pack DNA groupings have not had the option to accomplish a typical pressure proportion of under 1.7 pieces/base. Civelevel [14, 15] of the calculation used to pack DNA arrangements like pressure calculations. BioCompress [16], GENCOMPRESS [17] and DNACompress [18] perform DNA succession pressure. Their typical pressure rate is around 1.74 pieces per unit. In this way, another pressure calculation was presented as "DNABIT pressure". The pressure pace of this calculation is under 1.56 pieces per base for bigger genomic (ideally 200,000 characters).

III PROPOSED METHODOLOGY

3.1 EfficientDNACompressionTechnique

The new calculation works in even mode utilizing an option factual strategy that depends on AR, and model boundaries are resolved utilizing the Modified FireFly (MFF) calculation. In the fundamental firefly calculation, the most brilliant fireflies move arbitrarily, so the splendor might diminish relying upon the course. This dials back the presentation of the calculation during that specific emphasis. MFF has been proposed to take care of this issue. On the off chance that this calculation can move this most splendid firefly in only one heading with further developed brilliance, the exhibition of the calculation won't be corrupted for the worldwide best arrangement of the relative multitude of arrangements acquired by this specific strategy rehash. The proposed framework expects to accomplish higher pressure proportions and lessens stockpiling, recovery, transmission upward, and induction construction and usefulness from pressure to bring the application from a commonsense and practical perspective [19].

3.2 Featureselection

The put away sign was in the time area and was basically arbitrary or stochastic. Social state qualities can be time-ward and recurrence subordinate. Discrete wavelet change innovation is reasonable for this since it has ideal goal in the time and recurrence area. This sign was examined utilizing the 8-level discrete wavelet change of the eighth request Daubechies wavelet [20]. Otherwise called db8, it is utilized to break down the sign on the grounds that the state of the db8 mother wavelet looks like an EEG wave. MATLAB was utilized to foster the DWT program. Despite the fact that it tends to be gotten by the corresponding of the free sign, recreating a decayed 8-level DWT that gives 8 signs of various frequencies was not fundamental. Since the testing recurrence is 512 Hz, the 5-level detail factor is in the 1535 Hz, the 6-level detail factor is in the 715 Hz range, the 7-level estimation factor is 47 Hz, and the 8-level guess factor is in the 04 Hz recurrence range. Other itemized coefficients and a portion of the guess coefficients are not in the recurrence range, so the overall elements were disregarded and extricated.

3.3 Classification

Order precision is one of the primary entanglements of the created BCI framework. This straightforwardly influences the choices made on the BCI yield. The grouping precision of a trial is determined by separating the quantity of accurately characterized tests by the complete number of tests. In this review, we utilize k-approval cross-approval to compute order precision and assess the presentation of the proposed technique. In the k-approval methodology, the dataset is separated into k totally unrelated subsets of roughly equivalent size and the technique is rehashed k times. For each situation, one

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1130 - 1136
https://publishoa.com
ISSN: 1309-3452

of the subsets is utilized as the test set and the other k1 subsets are consolidated to frame the preparation set. Then, at that point, the typical precision of all k preliminaries is determined. In this review, we pick k = 10. This is on the grounds that it is a typical decision for k-crease cross-approval [21]. Figure 1 shows the plan of how the element vectors extricated in this study are partitioned into 10 fundamentally unrelated subsets as per the k fold cross-approval framework. As displayed in Figure 1, each subject's arrangement of element vectors is partitioned into 10 subsets, and this cycle is rehashed multiple times (folds). As displayed in Figure 1, one subset is utilized as the test set and the leftover nine subsets are utilized as the preparation put down each point in time.

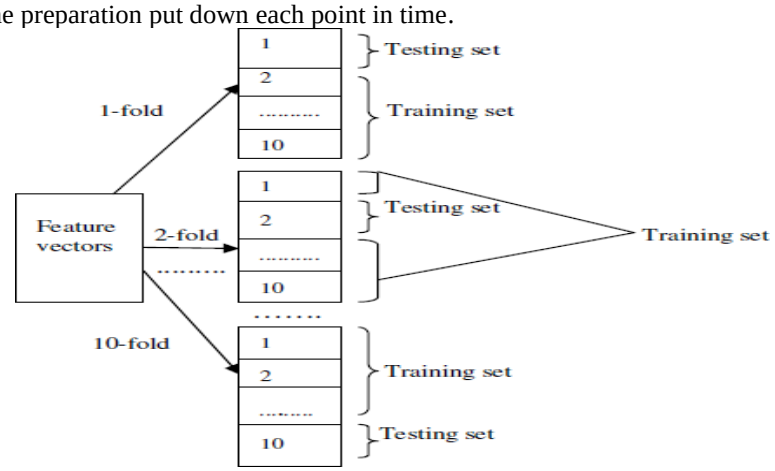

Figure 1 Partitioning design of the obtained feature vectors for the 10-foldcross-validation method

3.4 Bayesianclassification

The misleading positive (FP) case is the positive of the negative class. Misleading negatives (FNs) are cases that are anticipated to be negative whose genuine class is positive. Genuine up-sides (TPs) show cases that are anticipated to be in the positive classification, and genuine negatives (TNs) precisely demonstrate cases in the negative classification. Awareness and particularity are the most normally utilized presentation assessment boundaries [22]. Awareness is the capacity of the test to recognize individuals with the infection precisely. It is otherwise called an update. Particularity is the capacity of the test to precisely distinguish individuals who don't have the sickness. The amount of awareness and particularity overlooks the general loads of valid and bogus up-sides. The Bayesian classifier is an ideal classifier, created by the base mean mistake. All in all, this classifier attempts to find the normal least in the mistake work.

IV RESULTS AND DISCUSSION

4.1 FireFly Algorithm

The firefly calculation is an organic entity propelled metaheuristic calculation utilized in improvement issues. This calculation is roused by the firefly's evening time squinting way of behaving. One of the three principles used to make the calculation is that all fireflies are gender neutral. This shows that every firefly is drawn to the next splendid fireflies. Second, the subsequent decide is that the brilliance of the not entirely settled by the encoded objective capacity. The last decide is that gravity increments with splendor, diminishes with distance, fireflies approach splendid ones, and move arbitrarily when there are no brilliant ones.

4.2 Modified FireflyAlgorithm

The most brilliant fireflies are the ones with the best worldwide arrangements today. On the off chance that this most brilliant firefly moves arbitrarily, as in the standard firefly calculation, its splendor can diminish in view of the bearing. This lessens the exhibition of the calculation at that specific cycle. On the off chance that this calculation can move this most splendid firefly in only one heading with further developed brilliance, the presentation of the calculation won't be debased for the worldwide best arrangement of the multitude of arrangements got by this specific technique called Rehash. The progressions introduced in this work are: To decide the course of development of the most splendid fireflies, haphazardly create a m unit vector and consider u1, u2, ..., um. Then select the course U from the haphazardly produced bearing m. Toward this path, the brilliance of the most splendid fireflies increments as the fireflies move that way. In this manner, the most splendid firefly development can be communicated as:

$$x := x + \alpha U \qquad (1)$$

Whereαrefersto arandomsteplength.

4.3 Auto-Regression and Modified Firefly Algorithm

Autoregression and a changed firefly calculation can be utilized to further develop better pressure results for DNA successions. The calculation of the recently presented strategy (HOARDNAComp) is as per the following. Peruse the DNA grouping cautiously. Update the arrangement to 5 segments. Introduce the upsides of the four substance bases (A,

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1130 - 1136
https://publishoa.com
ISSN: 1309-3452

C, G, and T) as 0.25, 0.5, 0.75, and 1, separately. H. The quantity of compound bases should not surpass 1 preceding AR is applied to each line of the transformed succession. The MFF calculation is applied to streamline the coefficients of the AR, where every firefly in every cycle addresses the coefficients of the AR model, and the coefficients addressed by every firefly are utilized to make their own model, AR. The condition applies to each column of request. Wellness is determined utilizing the accompanying equation,
*Wellness = (Number of right anticipated bases/Total bases in the sequence)*100*

Essential introduction of firefly calculation boundaries like number of essential populace NPops, greatest number of cycles, and engaging quality component Age of NPop fireflies are discussed (essential arrangement). For each sets of fireflies (arrangement), follow the means underneath If firefly I is more alluring than firefly j (or wellness work I is superior to wellness work j), firefly j moves to firefly I as per Equation 4 beneath. Change the allure in view of the distance. In the event that the circumstances for halting the calculation are not met, go to stage 3.

Analyze the consequences of the successions produced by the AR model for the anticipated nucleotides, or compound bases. On the off chance that right, the nucleotide will be eliminated from the arrangement, if not it will be set in the succession. The iFlag document yield by the calculation comprises of a progression of 1's and 0's in paired configuration to demonstrate whether every nucleotide is accurately positioned.
This segment assesses the exhibition of the recently presented strategy (HOARDNAComp) and contrasts the presentation results and the accessible arrangement and pressure methods. Java is utilized to carry out the recently presented calculations. The proposed framework is then contrasted with existing DNA pressure methods utilizing molecule swarm enhancement (DCPSO). Figure shows the construction of the Bayes classifier.

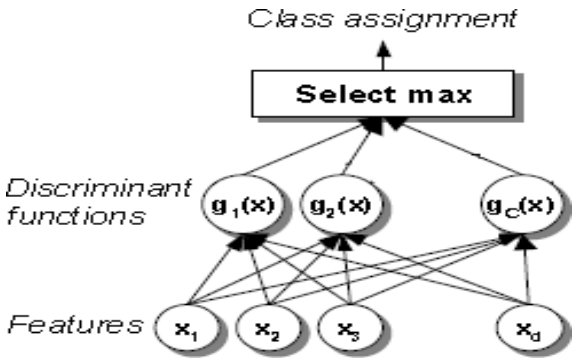

Figure 2 Structure of Bayes classifier

4.4 Naïve bayes based classification
The classifier generally plans to arrive at the best speculation H given the preparation dataset. As per Bayes' hypothesis, back probabilities (probabilities of speculations given variable qualities) can be determined in light of earlier probabilities of both found and all out information (recurrence of every speculation) as per the accompanying condition (7) [20]:

$$P(v_j|A) = \frac{P(A|v_j) \times P(v_j)}{P(A)} \qquad (2)$$

InstanceBasedK-nearestneighbor(IBK)classification
The IBK calculation is a closest neighbor classifier that utilizes a similar distance metric. The quantity of closest neighbors can be resolved naturally either unequivocally in the item proofreader or by overlooking the cross-approval center around the upper bound determined by the predefined esteem. The distance work is utilized as a boundary of the hunt strategy. The rest is equivalent to for IBL. That is the Euclidean distance. Different choices incorporate Chebyshev, Manhattan and Minkowski distances. The IBK or K Nearest Neighbor characterization groups occasions in view of similitude. This is quite possibly the most famous example acknowledgment calculation.

4.5 K-star order calculation
The Kstar calculation, otherwise called an occurrence based classifier, utilizes an entropy measure to consider the likelihood of transforming one case to one more by browsing every conceivable change. Entropy is a proportion of data that can be utilized to characterize EEG signals. This is a solid and significant way to deal with the characteristics of genuine qualities, images, and missing qualities. Ahaetal on the grounds that occasion based calculations intended for representative ascribes fall flat with the capacity of genuine qualities. There is an incorporated theory base for. (1991). A methodology is taken to deal with representative ascribes that are successful and subsequently impromptu for the qualities of genuine qualities. The treatment of missing qualities by the classifier powers a correlation question. As a rule,

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1130 - 1136
https://publishoa.com
ISSN: 1309-3452

missing characteristics are treated as various evaluations and are viewed as different as could be expected. These missing qualities are essentially disregarded and supplanted with mean qualities. These issues are best tackled by entropy-based classifiers. Information speculations assist with deciding the confinement between cases. The flightiness of moving starting with one occurrence then onto the next is really a partition between cases. Entropy as a proportion of distance: Data speculations assist with deciding the division between cases. The intricacy of transforming one case to one more is really the division between the occasions. This occurs in two phases. To begin with, it describes a restricted arrangement of changes that depict them starting with one example then onto the next. Then (a) transform one case to (b) another occurrence. It utilizes a "program" with a restricted gathering of changes that beginning and end with. Use 'r' to plan the example to itself. σ closes the whole arrangement of prefix codes P for T *. The components of set T are characterized by a one of a kind balanced change at I:

Table 1 Compression Result

Name of the sequence	Size of Sequence (In Bytes) Before Compression	Size of Sequence (InBytes) AfterCompression			
		Using Existing Algorithm (DCPSO)		Using Proposed Algorithm (HOARDNAComp)	
		Size of Sequence	Compression Ratio	Size of Sequence	Compression Ratio
CHMPXX	121024	20214	1.34	20064	1.33
CHNTXX	155943	27971	1.43	26407	1.35
HEHCMVCG	229354	41732	1.46	41407	1.44
HUMDYSTROP	38770	7190	1.48	6753	1.39
HUMGHCSA	66495	11932	1.44	11632	1.40
HUMBB	73308	13174	1.44	13161	1.44
HUMHDABCD	58864	10552	1.43	10252	1.39
HUMHPRTB	56737	10392	1.47	10255	1.45
MPOMTCG	186609	33106	1.42	32731	1.40
SCCHRIII	316613	56048	1.42	54294	1.37
VACCG	191737	31962	1.33	31708	1.32
		Average	1.42	Average	1.39

4.6 Multi Layer Perception (MLP) classification

MLP is a fake brain network method that is enlivened by the action of neurons in the cerebrum. Utilizing this method, a model is worked from the prepared preparation information. The MLP versatile brain network comprises of three layers. Each layer contains different hubs, copying neurons. The principal layer is an info layer containing hubs addressing the quantity of highlights in the preparation dataset. The subsequent layer can incorporate different secret layers. The quantity of secret layers and holder units is inconsistent. The third layer is the result layer. For instance, the fake brain network portrayed in Figure 12 comprises of an info layer, a secret layer, and a result layer. The information layer contains four buttons. Every hub is associated with four hubs in the secret layer, which are likewise associated with the hub in the result layer. Associating lines convey loads that address how much electrical motivations in the neuron.

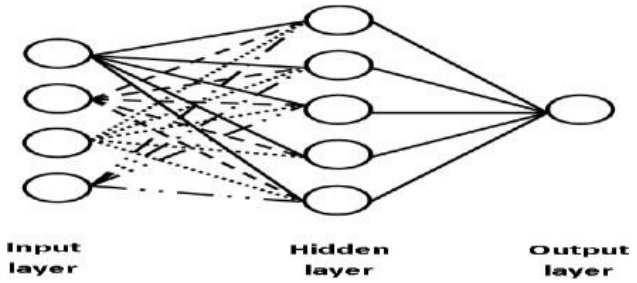

Figure 3 MLP process

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1130 - 1136
https://publishoa.com
ISSN: 1309-3452

As should be visible from Table 1 underneath, the presentation file is considered in contrast to the pressure proportion utilizing the current [19] and the proposed strategy. The outcomes show that the proposed procedure (HOARDNAComp) gives a superior pressure pace of 1.39 bpb than existing DNA pressure calculations.

4.7 Straightforward CART Method

In 1984, Leo Breiman, Jerome Friedman, Richard Olsen and Charles Stone together evolved Classification and Regression Trees (CART) and gave an overall technique to creating factual models from single element information. Truck is strong in light of the fact that it handles fragmented information, information with info and expectation attributes. The tree created via CART will contain intelligible standards. The calculation will consider the arrangement of tests, the subject of the qualities of the information will prompt scaling back the information and will go on until specific halting models are reached18. Truck can deal with both unmitigated and numeric factors. The Gini file estimates how well a given characteristic isolates learning designs into fixation classes. Here the twofold division of properties happens. This is the most generally utilized factual strategy. It gives a univariate paired choice order.

4.8 The Algorithm

Step1: The first is the means by which the parting trait is chosen.
Step2: The second is settling on what halting principles should be set up.
Step3: The latter is the way hubs are relegated to classes

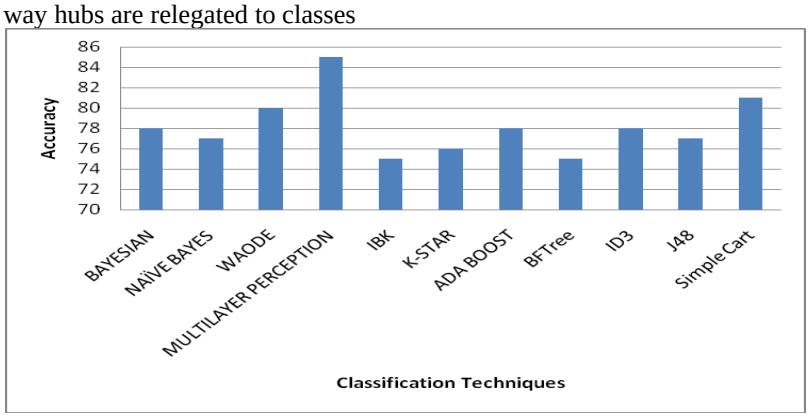

Figure 4 Accuracy performance comparison for all classification schemes

The presentation of the proposed MLP plan will be assessed alongside existing grouping plans like Bayesian, Naive Bayes, Ward, IBK, kstar, adaboost, BF Tree, ID3, J48, Simple Cart. Figure 14 shows a graphical portrayal of the exactness execution of all grouping plans. This shows that the proposed MLP accomplished superior execution contrasted with existing plans. Because of the viable preparation process, the introduced MLP accomplished improved results. The ROC execution of the proposed MLP plan will be assessed alongside existing order plans like Bayesian, Naive Bayes, Ward, IBK, kstar, adaboost, BF Tree, ID3, J48 and Simple Cart. Figure 15.8 shows a diagram of collector working trademark (ROC) execution for all order plans. This shows that the proposed MLP accomplishes elite execution contrasted with existing plans. Because of the powerful preparation process, the introduced MLP accomplished improved results.

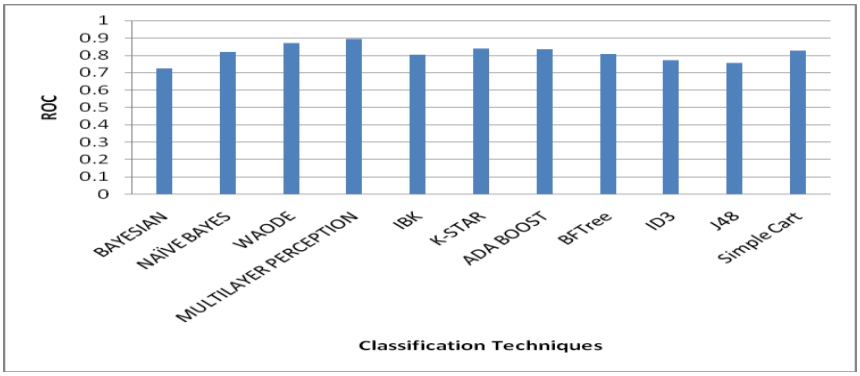

Figure 5 ROC performance comparison for all classification schemes

It shows that the accuracy and ROC performance of proposed MLP attained high compared than otheralgorithms.The accuracy of proposed scheme attained 84% and ROC attained 0.9 high values compared than other schemes.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 1130 - 1136
https://publishoa.com
ISSN: 1309-3452

V CONCLUSION

Delicate processing strategies are utilized to track down improved answers for complex genuine issues. The proposed framework utilizes normal delicate figuring methods. Autoregressive and firefly calculations to further develop the pressure aftereffects of DNA arrangements. The new calculation works in level mode utilizing an option factual technique that depends on AR, and model boundaries are set utilizing the changed FireFly calculation. In the fundamental firefly calculation, the most splendid fireflies move arbitrarily, so the splendor might diminish relying upon the heading. This decreases the presentation of the calculation at that specific cycle. A changed firefly has been proposed to take care of this issue. In the event that this calculation can move this most brilliant firefly in only one course with further developed splendor, the exhibition of the calculation won't be debased for the worldwide best arrangement of the relative multitude of arrangements acquired by this specific technique. The recently presented framework utilizes MFF. The proposed HOARDNA Comp technique intends to accomplish more noteworthy pressure proportions and is useful and practical to lessen the upward of stockpiling, recovery, transmission, and end structures, as well as the usefulness of the subsequent successions. It gives the application a benefit according to a perspective.

REFERENCES

1. Kircher M and Kelso J, 2010, High-throughput DNA sequencing – concepts and limitations, Bioessays, Wiley Online Library, 32, 6, 524–536.
2. Paula WCP, 2009, An Approach to Multiple DNA Sequences Compression-A thesis submittedin partial fulfillment of the requirements for the Degree of Master of Philosophy, The HongKong Polytechnic University, HongKong.
3. Shiu HJ, Ng KL, Fang JF, Lee RCT and Huang CH, 2010, Data hiding methods based uponDNAsequences,Information Sciences, Elsevier,180,2196–2208.
4. Mridula TV and Samuel P, 2011 , Lossless segment based DNA compression, Proceedings ofthe 3rd International Conference on Electronics Computer Technology, IEEE Xplore Press,298-302.
5. Ariey, F., Witkowski, B., Amaratunga, C., Beghain, J., Langlois, A. C., Khim, N & Lim, P.(2014). A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature, 505(7481), 50.
6. Kuscu, C., Arslan, S., Singh, R., Thorpe, J., &Adli, M. (2014). Genome-wide analysis revealscharacteristics of off-target sites bound by the Cas9 endonuclease. Nature biotechnology, 32(7),677-683.
7. De Bourcy, C. F., De Vlaminck, I., Kanbar, J. N., Wang, J., Gawad, C., & Quake, S. R. (2014).A quantitative comparison of single-cell whole genome amplification methods. PloS one, 9(8),e105585.
8. Maguire,P.,Moser,P.,&Maguire,R.(2016).UnderstandingConsciousnessasDataCompression.Journal of Cognitive Science,17(1),63-94.
9. Alves, F., Cogo, V., Wandelt, S., Leser, U., &Bessani, A. (2015). On-demand indexing forreferentialcompression ofDNAsequences.PloSone,10(7), e0132460.
10. AlOkaily,A.,Almarri,B.,AlYami,S.,&Huang,C.H.(2017).TowardaBetterCompressionforDNASequencesUsingHuffmanEncoding. JournalofComputationalBiology,24(4), 280-288.
11. X Chen et al. A compression algorithm for DNA sequences and its applications in Genomecomparison. In Proceedings of the Fourth Annual International Conference on ComputationalMolecular Biology, Tokyo, Japan, April8-11, 2000. [PMID: 11072342].
12. Hossain, M. M., Habib, A., &Rahman, M. S. (2016). Performance Improvement Of Bengali Text Compression Using Transliteration And Huffman Principle,International Journal of Engineerng Research and Applications, 6(9), 88-97.
13. Kärkkäinen,J.,Kempa,D.,&Puglisi,S.J.(2013,June).Lightweightlempel-zivparsing.In InternationalSymposiumonExperimentalAlgorithms (pp.139-150).Springer,Berlin,Heidelberg.
14. Roy,S.,Mondal,S.,Khatua,S.,&Biswas,M.(2015).AnEfficientCompressionAlgorithmfor Forthcoming New Species. International Journal of Hybrid Information Technology, 8(11),323-332.
15. X Chen et al. In Proceedings of the Fourth Annual International Conference on ComputationalMolecularBiology, Tokyo, Japan, April 8-11, 2000.
16. XChenetal.Bioinformatics18:1696(2002)[PMID:12490460]
17. Govind Prasad Arya, RajendraBharti, DNA Compression Using Particle Swarm Optimization(DCPSO), "Journal of Advanced Research in Dynamical & Control Systems, 05-Special Issue,July2017",ISSN:1943-023X, page: 295-302.
18. Govind Prasad Arya, RajendraBharti,A Modified Compression Algorithm for Nucleotide DataBased on Differential Direct Coding and Variable Length Look up Table (LUT), "Advances inComputational Sciences and Technology", ISSN 0973-6107 Volume 10, Number 6 (2017) pp.1571-1576.
19. Govind Prasad Arya, Rajendra Bharti, An Improvement over Direct Coding Technique to Compress Repeated & Non Repeated Nucleotide Data,"IEEE Explore Digital Library", International Conference on Computing, Communication and Automation, ICCCA-2016.
20. Govind Prasad Arya, Rajendra Bharti, A Compression Algorithm for Nucleotide Data Based onDifferential Direct Coding and Variable Length Look up Table (LUT), "(IJCSIT) InternationalJournalofComputerScience andInformationTechnologies",Vol. 3(3) , 2012, 4411-4416