An Optimized Method Implemented In Analyzing of Organ System Using Intelligent Tools

Safa A. Hameed,
College of CS and IT
University of Anbar,
Al-Anbar, Iraq
la_programmer89@yahoo.com

Raed I. Hamed
College of Science and Technology,
University of Human Development,
Sulaimani, KRG, Iraq
raed.alfalahy@uhd.edu.iq

Abstract—The paper proposes an efficient approach applied in DNA base calling, which concerns efficiency and sensitivity. We utilized the Neuro-Fuzzy model in the analysis issues to determine the confidence value prediction in DNA base calling, that is solved by several attempts applied in the MATLAB tool, the model is implemented for the collected data for each base in the DNA sequencing. The model is designed by using the ANFIS tool, which contains three subsystems a main system. We obtain three features (peakness, height, and spacing) for each base from the three subsystems and in the main system use these three features as the input to predict the confidence value for each base in the DNA. This achieves a high accuracy in the obtained results with high-performance.

Index Terms— Index Terms: ANFIS, Confidence value, DNA sequencing

I. INTRODUCTION

THE Deoxyribo Nucleic Acid (DNA) molecule consists of four sorts of nitrogenous bases: A-adenine, T-thymine, G-guanine, and C-cytosine. The genetic data is encoded by substituting these four bases in the DNA strand [1]. A confidence value measure is utilized to decide how well the base calling algorithm has performed for each base call. This will obviously make it less demanding to reveal potential blunders and correct them, consequently expanding the throughput of genetic sequencing [2]. The genomic data manages the accumulation and investigation of information on every one of the genes in a life form. It has been progressively imperative to scientists, helping them to focus on designs from information, which transform into biological facts and knowledge [3]. One such tool utilized for this issue is the fuzzy logic [4], and the Neuro-fuzzy [5], that can fulfill the requirement for a DNA sequencing analysis procedure and give a deliberate and fair-minded approach to prefer this topic. More importantly that the Neuro-fuzzy approach is efficient to apply in bio system and in genomic [6], the

Neuro-fuzzy can be used as a method to

identify the changes of the statistics of selection if some "agent" is capable of predicting (and thus recognizing) [7], and in the functional analysis of gene expression data from microarray experiments [8].

II. LITERATURE REVIEW

There are several attempts to analyze the DNA base calling, some of those are used to predict the confidence value of DNA base calling. Here, we describe different procedures in analyzing the DNA base calling, such as in (Golden et al, 1993, [9]), depict a neural network model for photometric signal molding amid crude information securing with a mechanized DNA sequencer. In (Berno, 1996, [10]), another base-calling algorithm that is proposed to be autonomous of a specific sequencing innovation has been produced and appeared to be viable with information from the applied Biosystems 373 sequencing framework. In (Ewing et al, 1998, [11]), In Ewing et al, (1998, [11]), the main aspect of the work is applied in the field of confidence value work, by primarily in support of the development of Phred base calling system.

Creating a confidence value for the base calls, in this procedure of traces and quality values using consed. Ressom et al, (2005 [2]), describe the method using a fuzzy logic system to provide the confidence values of bases called in the DNA. The three variables: peakness, height and base spacing are calculated during the base calling process which are involved in the fuzzy system. Whereas David et al, (2007 [12]), they discuss the decoding Trace Peak DNA sequence base calling, which has made measures of confidence of base calls vital, by utilizing Neuro-fuzzy classifiers. In (Raed et al, 2011 [3]), a fuzzy Petri net (FPN) approach to modeling fuzzy rule-based reasoning is suggested to determine confidence values for bases called in DNA sequencing, In Safa et al,(2016 [13]), they discuss a Fuzzy logic

method implemented on the DNA base calling to obtain the confidence value for each base.

III. OBTAINING THE CONFIDENCE VALUE USING NEURO-FUZZY

We utilize the Neuro-fuzzy method for obtaining the confidence value of DNA base calling, which is the optimal model to measure incorrect analysis and editing process in order to make it much easier and faster to obtain results. Figure 1. Illustrates the systems of the confidence Neuro-Fuzzy.
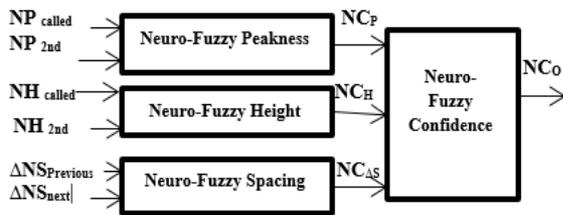

Figure 1: Model overview of the main ANFIS confidence system

This approach is proposed for the collecting of information at the base, for more details [2]. Figure 2 - sited at the end of the paper, shows the diagram for calculating the Nuero-Fuzzy confidence value.

A. The Method

This technique is applied by using the Matlab tool, the model of the ANFIS contains three subsystems and a main system, we obtain the three features (peakness, height, and spacing) from the three subsystems and in the main system use these three features as the input to predict the confidence value for each base [13].

B. Implementation

In this system, we use the collected data from the DNA base calling (for more details see [2]) as the dataset file, and divided into two parts: training dataset and testing dataset and by implementing the method, the training set we use about 80% of data from the dataset file, whereas in the testing set, we use about 20% of the dataset file.

We attempt to apply the data through several processes in the Neuro-Fuzzy model, and choose the most suitable system with less average testing error. We tried to apply several processes in ANFIS system, and by chosing different membership function for the input, and we attempted a different type of the network with different epochs, in order to achieve the most suitable system with less average testing error. Table 1 shows the input data.

TABLE I
THE INPUT DATA VALUE FOR EACH BASE

| Features	Input data	The bases of the sequences					
		A	T	C	T	C	G
Peakness	P called	0.9 98	0.9 99	0.7 94	0.9 99	0.9 30	0.9 99
	P 2nd	0.3 61	0.4 78	0.8 38	0.7 21	0.6 65	0.6 18
Height	H called	0.8 89	0.9 91	0.6 44	0.9 54	0.6 96	0.9 52
	H 2nd	0.5 60	0.4 21	0.6 04	0.6 06	0.5 31	0.4 85
Spacing	ΔS next	0.3 05	0.3 05	0.2 81	0.2 86	0.3 02	0.2 74
	ΔSprevious	0.2 98	0.3 05	0.3 05	0.2 81	0.2 86	0.3 02

IV. THE RESULTS OF THE MODEL

We used the Neuro-Fuzzy technique, which obtained the results with high performance through building the optimal designing, through training and testing the system. We performed several ANFIS models in our data for each system, in order to get the most suitable model with less average testing error and relied on it. To explain our model a section of a DNA sequence which contains six bases (ATCTCG) is used.
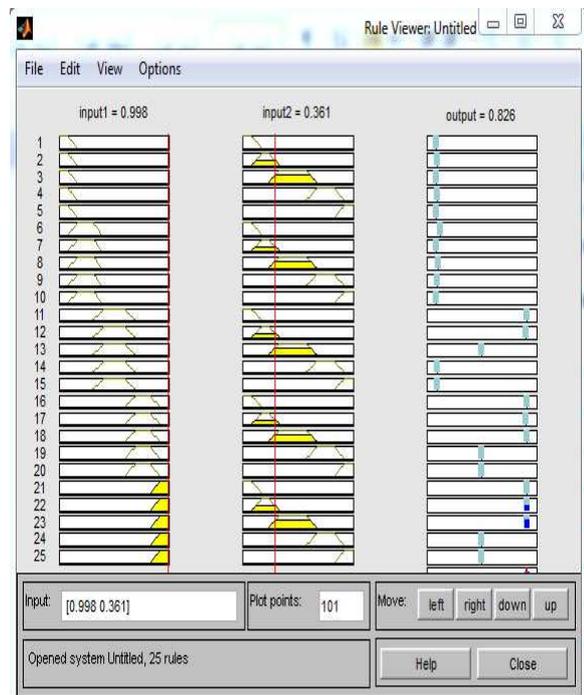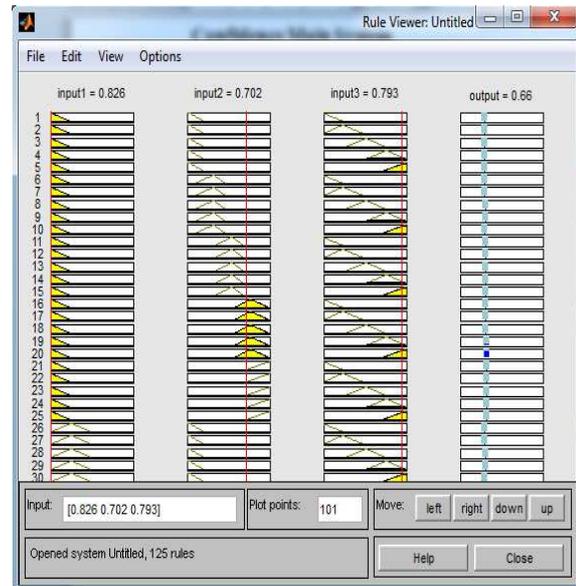

(a) Peakness
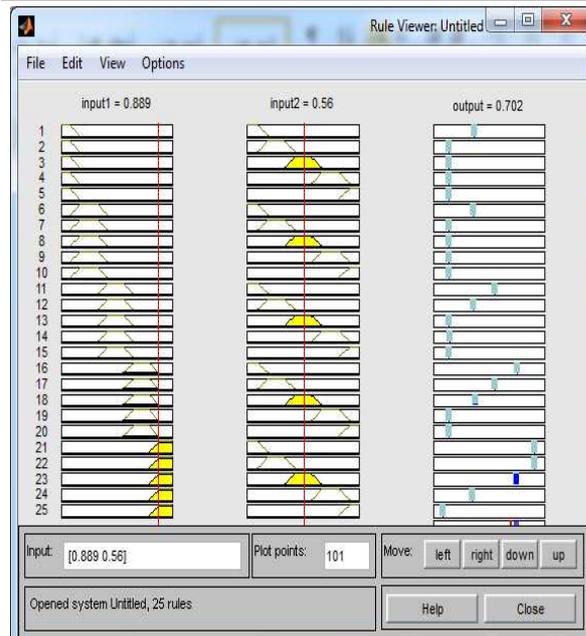
(b) Height
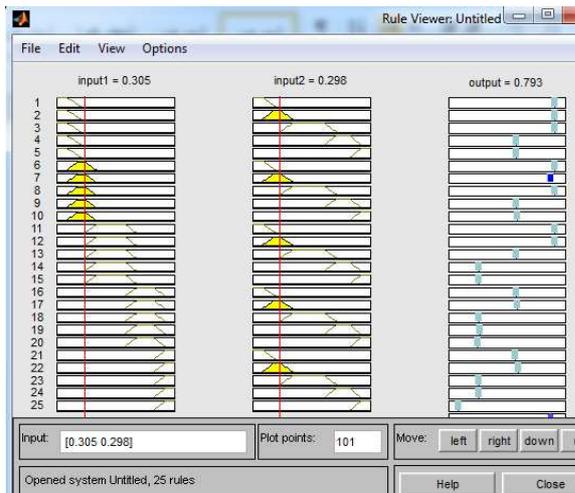
(d) Confidence value

Figure 2: The view rules (a), (b), (c) of the three subsystems and (d) the view rules of the main system

In figure 3.a,b, and c. We use the collected data that are (P called and P 2nd), (H called and H 2nd) and (ΔS next and ΔSprevious) as the two inputs in each of the three Neuro- Fuzzy subsystems to obtain the three features that are peakness, height and spacing and use it the three inputs in the main system to obtain the confidence value as shown in figure 3.d.

TABLE II
THE OBTAINED RESULT FROM THE MAIN NEURO- FUZZY CONFIDENCE

Bases Input data	A	T	C	T	C	G
Peakness	0.826	0.825	0.501	0.499	0.614	0.774
Height	0.702	0.868	0.229	0.749	0.359	0.772
Spacing	0.793	0.794	0.794	0.791	0.791	0.795
Confidence	0.66	0.831	0.187	0.758	0.414	0.73

V. THE ACCURACY OF THE NEURO-FUZZY CONFIDENCE VALUE

In the Neuro-Fuzzy method, we can analyze the DNA base calling using the data gathered for each base and measured the confidence value and can improve the accuracy of the DNA base calling, Figure 4. Shows the accuracy of the Neuro-Fuzzy confidence model for each subsystem and the main system.
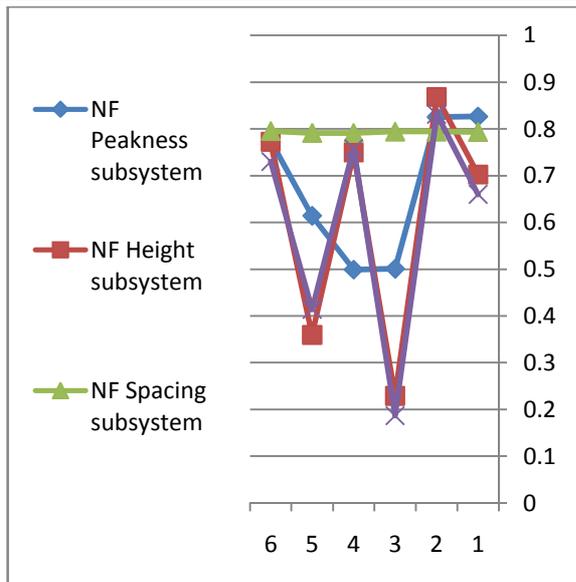

(c) Spacing

Figure 3: The accuracy for the Overall Neuro-Fuzzy Confidence value

VI. CONCLUSION

In this method, we utilize the Neuro-Fuzzy technique to obtain the confidence value for each base in DNA, the efficient properties of Neuro-Fuzzy technique which was an optimized way to implement our model, and give the results in high accuracy, we implement the several ANFIS for each in the three subsystems to obtain the three features, and several ANFIS in the main system to obtain the confidence value for each base, our model developed the method in previous and enhance the implementation with high performance.

REFERENCES

Basic format for books:

[1] T. A. Hiwarkar1, R. S. Iyer2, "New Applications of Soft Computing, Artificial Intelligence, Fuzzy Logic & Genetic Algorithm in Bioinformatics", International Journal of Computer Science and Mobile Computing, 2013.

[2] H. Ressom, P. Natarjan, R.S. Varghese, M. T. Musavi, "Applications of fuzzy logic in genomics", Journal of Fuzzy Sets and Systems 152, 125–138, 2005.

[3] R. I.Hamed, S.I. Ahson, "Confidence value prediction of DNA sequencing with Petri net model", Journal of King Saud University – Computer and Information Sciences , 23, 79–89, 2011.

[4] L.A. Zadeh, "Fuzzy logic, neural networks and soft computing", Comm. ACM 37 (3) 77–84, 1994.

[5] J. L. Pálsdóttir, " NEURO-FUZZY CONTROL ", Soft Computing, 2005.

[6] X. Ji, W. Wang, "A Neural Fuzzy System for Vibration Control in Flexible Structures", Intelligent Control and Automation, 2, 258-266, 2011.

[7] H. N. Teodorescu, "Genetics, Gene Prediction, and Neuro-Fuzzy Systems", the Context and A Program Proposal, F.S.A.I., Vol. 9, Nos. 1–3, pp. 15–22.

[8] D. Neagu, V. Palade, "A neuro-fuzzy approach for functional genomics data interpretation and analysis Neural Comput & Applic", 12: 153–159, 2003.

[9] J. B. Golden, D. Torgersen, and C., " Pattern recognition for automated DNA sequencing: I. On-line signal conditioning and feature extraction for basecalling". In Proceedings of the First International Conference on Intelligent Systems for Molecular Biology (ed. L. Hunter, D. Searls, and J. Shavlick), pp. 136–144. AAAI Press, Menlo, Park, CA, 1993.

[10] A.J. Berno, "A graph theoretic approach to the analysis of DNA sequencing data", Genome Res. 6: 80–91, 1996.

[11] B. Ewing and P. Green, "Base-calling of automated sequencer traces using phred: II. Error probabilities", Genome Research, 8, 186-194, 1998.

[12] D. Thornley, S. Petridis, " Decoding Trace Peak Behaviour A Neuro-Fuzzy Approach ", 2007.

[13] S. A. Hameed, R. I. Hamed, " Expert System of Fuzzy Logic Reasoning Based Implementation for DNA base calling ", International journal of business and ict, Vol.2, No.3-4, 2016.