

A OBTAINING UNIQUE BY ANALYZING DNA USING A NEURO-FUZZY ALGORITHM

Thaer kh. Asman¹, Hadeel M Saleh^{2*}, Alaa Hamid Mohammed³

¹College of Arts University of Anbar

^{2*}Continuing Education Center ,University of Anbar,Iraq

³Department of Medical Instruments Techniques Engineering, Al-Maarif University College ,Ramadi ,Iraq:



ARTICLE INFO

Received: 22 / 02 / 2023
Accepted: 08 / 05 / 2023
Available online: 12 / 06 / 2023

DOI: [10.37652/juaps.2023.178906](https://doi.org/10.37652/juaps.2023.178906)

Keywords:

Deoxyribonucleic Acid (DNA),
Neuro-fuzzy,
bioinformatics,
ANFIS Architecture,
biological data, G-guanine,

Copyright©Authors, 2023, College of Sciences, University of Anbar. This is an open-access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).



ABSTRACT

In this study, the problems of accurate exemplification and a trustworthy test for judging DNA base-calling were discussed. The study of this problem was done by utilizing a model to gather data and develop a suitable design according to the problem's criteria. The height, peakness, and spacing of the first most likely candidate (the base), as well as the peakness and height of the second most likely candidate, are the three input variables in our model. By gathering data from base calling, these three attributes may be computed in the three subsystems and utilized in the main system to calculate the analysis value for each base in DNA sequencing. The MATLAB program was used to create this model. It's looking into applying the neuro fuzzy algorithm to solve the problem. These methods are delivering successful outcomes with high workplace performance and easy access to very precise data.

1. Introduction:

DNA base calling is a technique used in sequence analysis to identify the four kinds of bases (G-guanine (black), T-thymine (red), A-adenine (green), and C-cytosine (blue)) from cues given by a DNA sequencing machine[1]. For high-throughput information accumulation, the DNA sequence is the sole arrangement upon which a greater range of biological events may be expected [2]. The field of genomics is responsible for collecting and analyzing data on each and every gene in a living organism [3]. It has become more important to scientists since it aids in the concentration of designs from data, which then change into biological facts and knowledge.

DNA is the basic blueprint for life [4], and it has been subjected to a variety of procedures. DNA is a two-fold helix held together by hydrogen bonds that are (A, T) and (C, G) supplementary sets [5]. DNA is made up of four nucleotides (G: guanine-black, T: thymine-red, A: adenine-green, and C: cytosine-blue) [6].

In bioinformatics, artificial intelligence may be useful for both showing biological knowledge and producing new discoveries. The application of artificial intelligence (AI) approaches in bioinformatics is gaining popularity [7]. There is a recognition that a lot of challenges in bioinformatics require a different strategy to be addressed due to the immovability of present approaches or the lack of an informed and intelligent approach to misusing biological data [8]. For identifying relevant and fitting genes from a large number of genes tested, it is necessary to integrate biological knowledge with computational methodologies.

*Corresponding author at: Continuing Education Center, University of AnbarIraq ORCID:<https://orcid.org/0000-0003-3408-3679>;
Tel:+9647832893520
E-mail address: Haddeel.mohammed@uoanbar.edu.iq

In order to improve accuracy, this study aims to give a measurement score analysis that focuses on utilizing information from DNA base calling in high-throughput sequencing.

2. Literature review

Haznedar, B., Arslan, M. T., & Kalnl, A. (2018) ANFIS is an artificial intelligence method used to classify cancer gene expression profiles, achieving the highest average classification performance of 95.56%, with the highest performance achieved at 87.65%. [9]

Using an adaptive neurofuzzy inference system for classification of microarray gene expression cancer profiles.

Bamisaye, A., Adesina, M. O., Alfred, M. O., Idowu, M. A., Adeleke, O., & Adegoke, K. A [10] shown a walnut shell-chitosan composite (WNS-CH) was created as an effective biosorbent for Pb²⁺ uptake and biofilter for *Klebsiella pneumoniae* and *Bacillus subtilis*. In order to forecast the adsorption of Pb²⁺ based on pertinent input factors, it was described using SEM and FTIR. Additionally, an intelligent and cost-effective machine learning model, an adaptive neuro-fuzzy model, grid-partitioning (GP), and fuzzy c-means (FCM) were constructed. The ideal GP-clustered ANFIS model produced the best predictions for Pb²⁺ adsorption, with correlation determination (R¹) values of 1.217, 0.563, 1.698, and 0.9966 for the root mean square error (RMSE), mean absolute deviation (MAD), and mean absolute percentage error (MAPE), respectively.

Mu'azu, M. A. (2022) [11] This study aimed to predict and early detect lung cancer using an adaptive neuro fuzzy inference system (ANFIS) and artificial neural network (ANN). The accuracy of (ANN) and (ANFIS) was 98.913 and 97.087 respectively. Model evaluation showed that the (ANN) and (ANFIS) have been able to detect the absence or presence of lung cancer.[11]

Cao, Y., Pourrostan, T., Zandi, Y., Denić, N., Ćirković, B., Agdas, A. S., ... & Milic, M. (2021) [12] The purpose of this study is to use patient Pap tests to identify abnormal tissue growth in the cervix region for the diagnosis of cervical cancer. The suggested methodology uses form and a neuro-fuzzy based diagnostic model to categorize cervical cancer into

benign or malignant phases for pattern identification. The presence of cervical cancer was successfully recognized by the fuzzy expert system (FES) with 100% accuracy after a neural network was trained with 15-Pap imaging datasets. [12]

Sharafati, A., Haghbin, M., Tiwari, N. K., Bhagat, S. K., Al-Ansari, N., Chau, K. W., & Yaseen, Z. M. (2021)[13] Using input factors such sediment concentration, flow depth, velocity, size of the sediment, Froude number, extraction ratio, number of tunnels, and flow depth, hybrid data-intelligence models were created to estimate the sediment removal of the sediment ejector.[13].

3. Method

To ascertain the degree of confidence in DNA-based calling, this model was developed. The height, peakness, and spacing data, which are the three input parameters in our model, are collected using the base calling method. The concavity at the highest point of a peak is the first characteristic (P), which is specified as a measure. The second attribute (H), i.e., the peak's height, is the height. The last aspect is the base spacing (S), which describes how the region compares from one peak to the next. These variables can be calculated at any location.

By gathering data from base calling and applying it in the main system to generate the analysis value for each base in DNA sequencing, these three attributes may be determined in three subsystems.

3-The Data Collected

The sequencing machine first sorts and arranges the raw data into a succession of peaks in order to calculate the three attributes (peakness, height, and spacing). Then, it is ready for the surge of peaks, and each peak is effectively connected to a base. The base calling algorithm needed an analysis value to assess how well the framework was being executed, which was the motivation behind the development of the confidence model. The characteristics are extracted from the raw data and fed into the model that is being used. These are the crucial variables that aid in properly differentiating the bases and forecasting the confidence levels. They appear to have a role in how people naturally evaluate confidence levels. Three

characteristics are extracted from the basic calling procedure and used in the model. The first feature is the height (H), or peak height. as seen in Figure 1. This feature is calculated from the two inputs that are gathered (P1 and P2). The second is the peakness (P), which is determined from the two inputs and is a metric that is associated with the concavity at the highest point of a peak (P1 and P2). The last feature is the base spacing (ΔS), i.e., the area contrasts starting with one peak, then onto the next, and calculated from the two input data (ΔS next and ΔS previous. The base-calling depends on the expectation for the spacing between bases [14]. Peakness, height, and spacing are three qualities that must be calculated separately from one another in order to be calculated properly. The confidence level for every base in DNA base-calling is then determined using these three qualities as the three inputs to the main algorithm [15], [16].

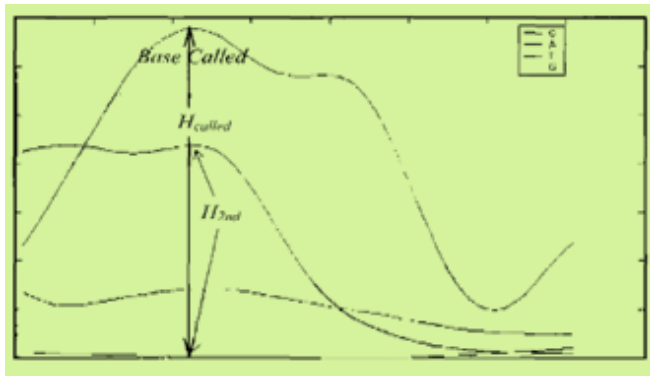


Figure (1): The clarification of the Height variables

The base call assesses the peakness, height, and slope of the signals at the local level. The base-spacing model is updated after a single base calling operation[17], and the base calling component is changed using the new spacing data. When base calling is finished, the confidence levels of the bases being called are calculated using the same parameters, height, base-spacing, and peakness.

The information variables that used in the systems for determining the features that are:

Height: The amount of each base from the baseline is used to compute height.

H called: The base's height is termed H called.

H 2nd: The second candidate's height

Peakness: it's referred to a peak's local sharpness. It is described as a full impact rather than specifically where a peak is located. Therefore, the higher the peakness, the more likely it is that there will be a peak there. The peakness numerical estimation is shown in the next sections.

P called: The base's peakness was named as a base called.

P 2nd: The second candidate's peakness.

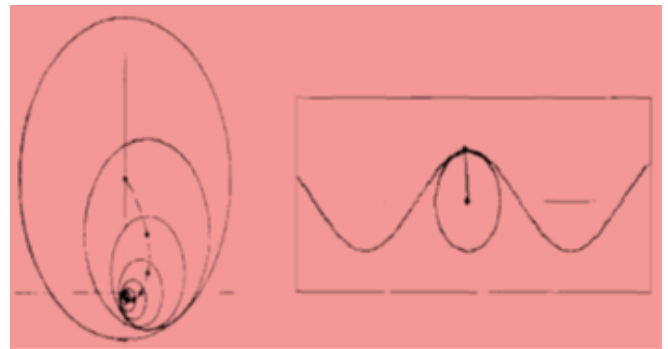


Figure (2): Surface Circle and Radius of the Curve

Spacing: Ideally, there should be an equal distance between each base. There are the raw data that are gathered through the base calling algorithm, these data are studied as the mathematical model and getting it to use as the input to the system to determine the spacing feature. There are various possible approaches to express the variation of the spacing. Phred's approach[18] is to select locations with bases that are evenly spaced apart and analyses the sequence district by area. The ΔS_{next} is processed as a divergence between the actual[14] (Calculated distance between the first called base and the second called base) as in equation (1) and the anticipated separation as explained in Figure 3.

$$\Delta S_{next} = S_{n_actual} - S_{n_predicted} \quad (1)$$

So also, for ΔS previous, the equation compares the actual separation between the called base and the previously called base with the projected separation. The model is used to compute the anticipated separation of the prior base equation (2). This serves as the input to the model that is used, which is described in the section below.

$$\Delta S_{previous} = S_{p_actual} - S_{p_predicted} \quad (2)$$

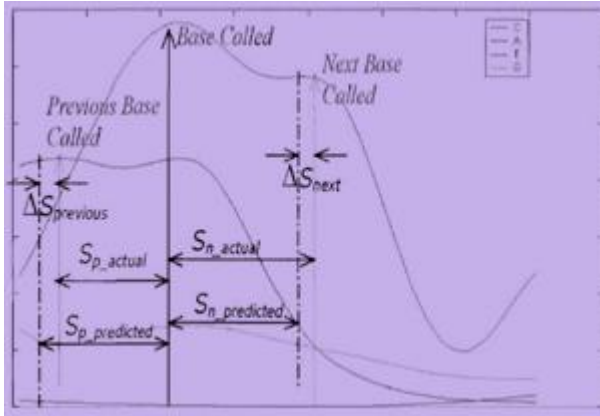


Figure (3): The clarification of the spacing variable

4- The Neuro-Fuzzy

A neuro-fuzzy (NF) theory method[19][20] is taken into consideration since it combines the benefits of fuzzy logic with neural networks. Manually setting offsets in production networks might be used to perform Fuzzy Logic[21] inferencing. Since it is widely believed that neural networks require training, this technique has been criticized. The ability to solve adjustment issues and design restrictions inherent in Fuzzy Logic is enabled by this merging.

5- ANFIS Architecture

A type of adaptive system called Adaptive Neuro-Fuzzy Inference Systems (ANFIS)[22] is essentially proportional to fuzzy induction. Both the Tsukamoto fuzzy model and the Surgeon fuzzy model represent ANFIS structures. Expect X and Y as two inputs, and Z as one output.

Rule 1: If x is A1 and y is B1, then f1 = p1x + q1y +r1

Rule 2: If x is A2 and y is B2, then f2 = p2x + q2y +r2

The ANFIS layers [23]:

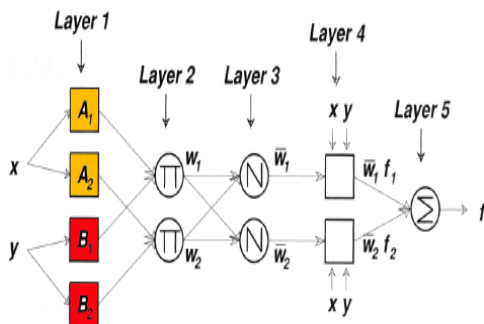


Figure (4): Layer 1

In the layer1 in Figure 4, every node i is an adaptive node with the node function

$$O_{i,1} = \mu_{Ai}[x], \text{ for } I = 1, 2 \quad (3)$$

or,

$$O_{i,1} = \mu_{Bi-2}[x], \text{ for } I = 3, 4 \quad (4)$$

Where x or y is the entrance to the node i and Ai or Bi is a linguistic label

** O 1, i is the membership grade of a fuzzy set and it indicates the extent to which the given input x or y fulfills the measures typically as in the equations (3) and (4), the membership function of a fuzzy set can be any parameterized MF, for

example, (Triangle (is specified by three parameters), Trapezoidal (is specified by four parameters), Gaussian (is specified by two parameters), and Generalized Bell (is specified by three parameters)). Variables in this layer are alluded to as antecedence Parameters.

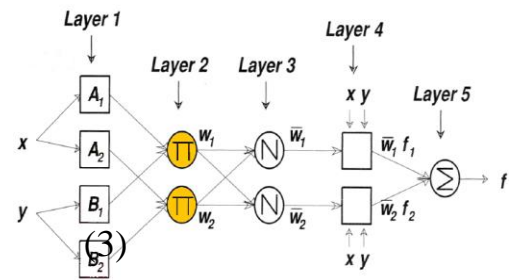


Figure (5) layer 2

Each node i in the layer2 in Figure 5, is a settled node named P, whose yield is the result of all the approaching signals as in equation (5):

$$O_{2,i} = W_i = \min\{\mu_{Ai}[X], \mu_{Bi-2}[X]\} \quad i = 1, 2 \quad (5)$$

Every node yield represents to the terminating quality of a rule.

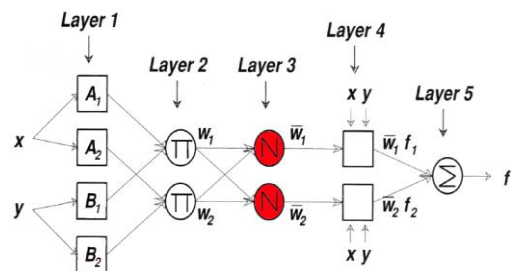


Figure (6): Layer 3

Each node in the layer 3 in Figure 6, is a settled node named N. The ith node computes the proportion of the ith rule’s terminating quality to the total of all rule’s terminating quality as in equation (2.10):

$$O_{3,i} = W_i = \frac{W_i}{[W_1+W_2]}, I = 1, 2 \quad (6)$$

[normalized terminating quality]

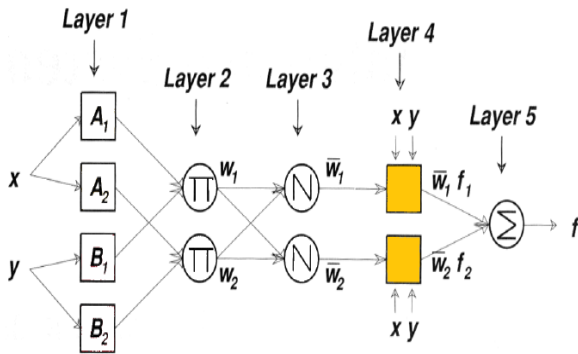


Figure (7): Layer 4

According to the layer4 in Figure 7, every node i , is on an adaptation node with the node function as in equation (7).

$$O_{i,4} = W_i F_i = W_i [P_i X + q_i Y + r_i] \quad (7)$$

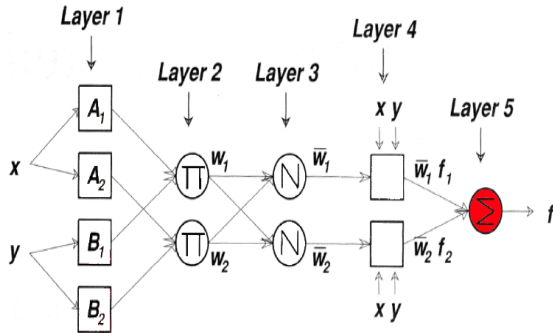


Figure (8): Layer 5

The single node in the layer5 in figure 2.10, is a settled node named S, which calculates the overall yield as the summation of every approaching signal as in

$$O_{5,i} = \sum_i W_i F_i \quad (8)$$

6-Approach Method

The DNA base calling analysis was carried out using the model in conjunction with the acquired data, and an appropriate design was created in accordance with the issue criteria.

6-1 Obtain The analysis score using Neuro-Fuzzy

When it comes to DNA sequencing, our Neuro-fuzzy model has a proven track record of delivering findings with a high accuracy ratio and efficiency.

For establishing the analysis value of DNA base calling, we utilize the Neuro-fuzzy methodology, which

is the best way for measuring inaccurate analysis and editing processes, making it faster and easier with the findings and providing a believable and accurate representation of the values. Figure 9 depicts the systems of the Neuro-Fuzzy Confidence.

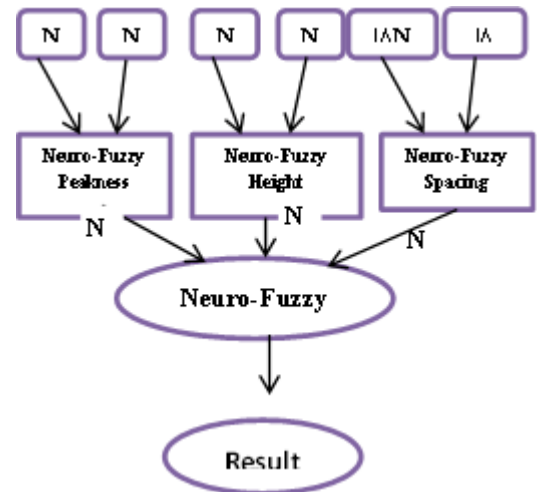


Figure (9): Model overview of the main ANFIS Analysis score system

The procedures for using the MATLAB tool to calculate the Neuro-Fuzzy analysis value were shown in the process. Three subsystems and a primary system are included in the simulation model for creating the ANFIS. Confidence amount for every bases in the major system may be estimated if we use three characteristics (peakness, height, and spacing) from the three subsystems.

6-2 The method for calculating the Neuro-Fuzzy analysis score is as follows:

- Step 1:** Gathered data from the base call algorithm (NP called, NP 2nd, NH called, NH 2nd, and NS next, NS prior) and saved it as dataset files.
- Step 2:** Split the dataset file into two files, one for training and one for testing. Take 80% of the data set for training and 20% for testing.
- Step 3:** Create numerous ANFIS models with various MF and use the dataset file for training and testing to determine the average testing error and select the best one.
- Step 4:** the three Neuro Fuzzy subsystems are loaded with these dataset files to extract the properties (Peakness, Height, and Spacing).

Step 5: Input the Main Neuro Fuzzy Confidence system with the dataset files comprising the results from the three Neuro Fuzzy subsystems.

Step 6: Create numerous ANFIS models with various MF and use the results dataset file for training and testing to determine the average testing error and select the best one.

Step 7: Utilizing the chosen ANFIS model, set the Confidence amount for every base of the basic Neuro-fuzzy system.

6-3 Implementation

Through implementing the method, the dataset file, which is split into training and testing datasets, is created using the data gathered from DNA base calling. We utilise around 80% of the dataset file for the training set and 20% of the dataset file for the testing set.

We try to apply the data through several processes in the Neuro-Fuzzy model and choose the optimal system with the lowest average testing error. In training step, we load the file training set of data samples for DNA base calling into each of the three subsystems. The Neuro-Fuzzy inference system is then generated, with the main system generating three inputs and one output to get each base's confidence level and the three subsystems generating two inputs and one output to get the three features. According to the system's membership function, we employ the AND logical operation in our rules to construct the most appropriate membership function, as mentioned, and have the ANFIS structure based on the number of membership functions and the input and output.

7- The mechanism method of the sub-systems and the main system

In the three sub-systems and in the main system, we trained each on the data with a percentage of 80%, and then tested the system on the data with a percentage of 20%, and we adopted the appropriate system with the lowest possible error rate, by choosing the appropriate member function for inputs and output, number of epochs and type of network, we changed it more than once and adopted the most appropriate system for each of the three systems as shown in next sections. The

system consists of two inputs (Pcalled, P2nd) to obtain a Peakness value as shown in Figure10.

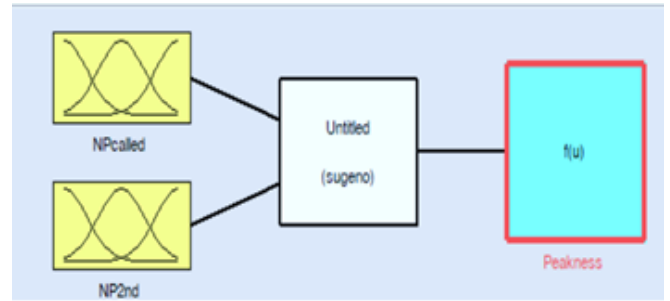


Figure (10): The Peakness Subsystem's Input/output Variables

We have applied several options and the most appropriate system was adopted after the training and testing process on the data to obtain the lowest possible error rate as shown in the Table (1) and in Figure (11)

Table (1): the parameters of peakness subsystem

Type of MF of inputs	Triangular
Type of MF of outputs	Constant
Type of network	back propagation
The number of epochs	500

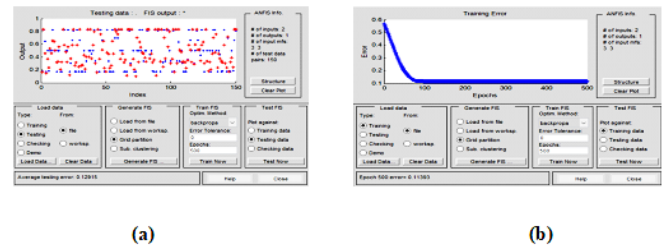


Figure (11): a: The Peakness training system, b: Peakness testing system

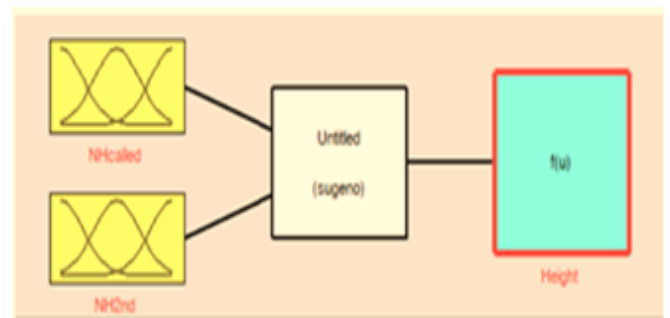


Figure (12): The Height Subsystem's Input /Output Variables

We have applied several options and the most appropriate system was adopted after the training and testing process on the data to obtain the lowest possible error rate as shown in the Table (2) and in Figure (13):

Table (2): The parameters of Height subsystem

Type of MF of inputs	Triangular
Type of MF of outputs	Constant
Type of network	back propagation
The number of epochs	550

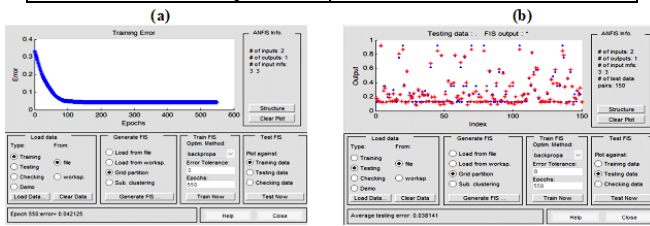


Figure (13): a: The Height training system, b: Height testing system

The Spacing Neuro-Fuzzy Analysis system

The system consists of two inputs (S_{called} , S_{2nd}) to obtain a Spacing value as shown in Figure 14.

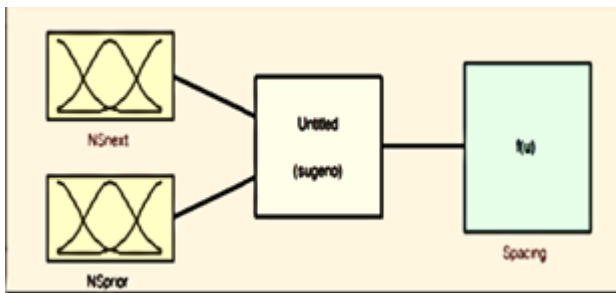


Figure (14): The Spacing Subsystem's Input/Output Variables

We have applied several options and the most appropriate system was adopted after the training and testing process on the data to obtain the lowest possible error rate as shown in the Table (3) and in Figure 15:

Table (3): the parameters of spacing subsystem

Type of MF of inputs	Gaussian
Type of MF of outputs	Constant
Type of network	back propagation
The number of epochs	600

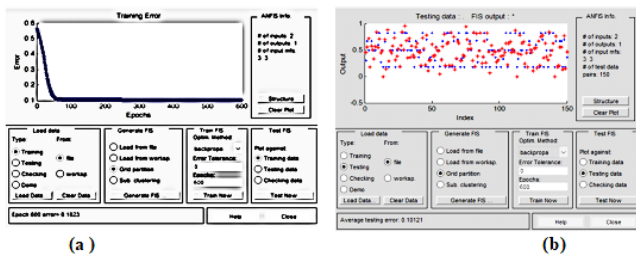


Figure (15): a: The Spacing training system, b: Spacing testing system

8- The Neuro-Fuzzy Analysis main system:

In the main system, we use the three variables inputs obtained from the three subsystems which are (Peakness, Height, and Spacing) as shown in Figure 16:

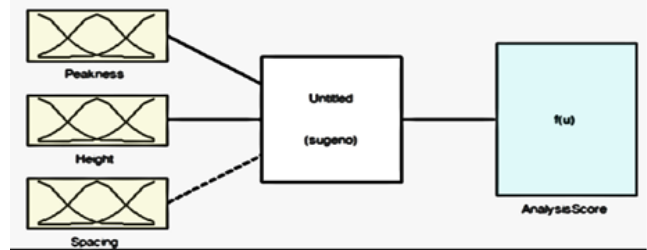


Figure (16): The main system's Input/Output Variables

We have applied several options and the most appropriate system was adopted after the training and testing process on the data to obtain the lowest possible error rate as shown in the Table (4) and in Figure(17):

Table (4): the parameters of main system

Type of MF of inputs	Trapezoidal
Type of MF of outputs	Constant
Type of network	back propagation
The number of epochs	750



Figure 17: a: The Spacing training system, b: Spacing testing system

9-Results

By creating the best design, then training and testing the system, we were able to accomplish high performance results. The rules are developed using the Neuro-Fuzzy approach. Depending on the data set entered into the model when the system is tested, the IF-THEN principles are automatically produced in a way which enables to provide the outcome with reach to prosperity in the right analysis. On our data for each system, we ran a number of ANFIS models in an effort to determine which model had the lowest average testing error and should be relied upon. Our method is illustrated using a segment of a six-base DNA sequence (CCGAT).

We collected the values of characteristics (height, spacing, and peakness) from these three ANFIS subsystems, with a view to compute analysis amount of every base's in base calling of DNA. The core ANFIS

system's three inputs were then comprised of these three values.

10- The Peakness subsystem:

After the system was approved, the data was applied to the system and the values were deduced as shown in the Table5.

Table (5): The peakness value

NP called	NP 2nd	Peakness
-----------	--------	----------

After the system was approved, the data was applied to the system and the values were deduced as shown in the Table 6.

Table (6): The Height value

NHcalled	NH2nd	Height
0.45	0.56	0.135
0.8	0.36	0.559
0.35	0.34	0.198
0.5	0.34	0.281
0.8	0.3	0.586

In Figure (18) shows the rules and the values that were applied in it, where the first column is the first input, the second column is the second input, and the third column is the output.

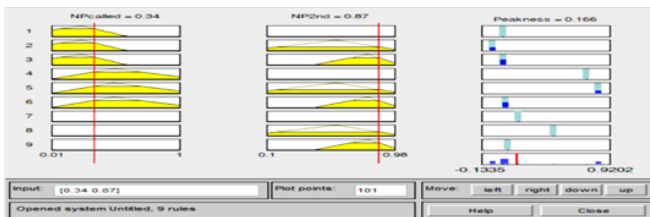


Figure (18): The Peakness view rules

In Figure (19) shows the rules and the values that were applied in it, where the first column is the first input, the second column is the second input, and the third column is the output.

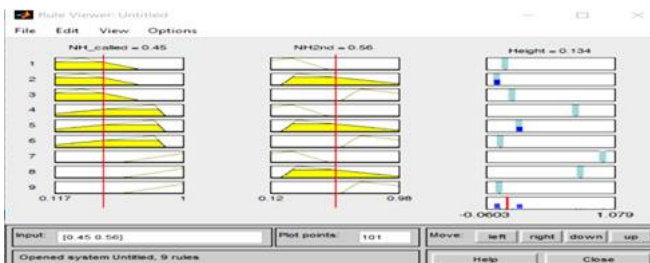


Figure (19): The Height view rules

12- The Spacing subsystem

0.34	0.87	0.166
0.76	0.76	0.372
0.45	0.56	0.5
0.34	0.5	0.326
0.5	0.47	0.5

In Figure (18) shows the rules and the values that were applied in it, where the first column is the first input, the second column is the second input, and the third column is the output.

11- The Height subsystem

After the system was approved, the data was applied to the system and the values were deduced as shown in the Table 7.

Table 7: The Spacing value

NSnext	NSprior	Spacing
0.36	0.7	0.449
0.45	0.5	0.584
0.35	0.46	0.704
0.12	0.45	0.822
0.65	0.34	0.5295

In Figure (20) shows the rules and the values that were applied in it, where the first column is the first input, the second column is the second input, and the third column is the output.



Figure (20): The Spacing view rules

13- The Analysis score of the main system

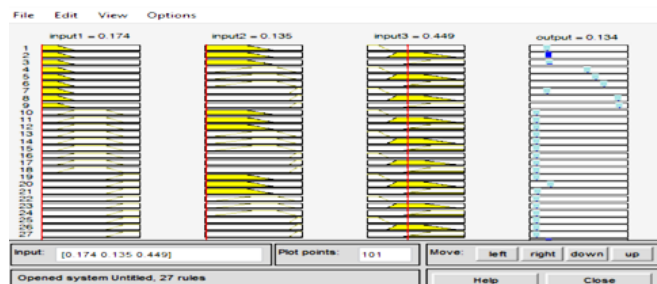
The data obtained from the three subsystems (Peakness, Height, and Spacing) are used as inputs to the main system to construct the analysis score of DNA base calling, as shown in Table 8.

Table 8: Analysis value from the main system

Peakness	Height	Spacing	Analysis Score
0.166	0.135	0.449	0.134
0.372	0.559	0.584	0.497
0.5	0.198	0.704	0.144

0.326	0.281	0.822	0.262
0.5	0.586	0.5295	0.515

Figure (21) shows the rules and the values that were applied in it from the three subsystems, where the first column is the Peakness, the second column is the Height, and the third column is the Spacing, and the fourth column is the Analysis score.



Figure(21): The view rules of the main system

14-The Conclusions

Using the MATLAB tool, the implementation of DNA sequence analysis is concerned with obtaining the analysis value for base calling. The model was able to forecast the value in neuro-fuzzy by developing the ANFIS for each of the three subsystems (height, spacing, and peakness) to acquire the three characteristics and the major system to prophesy the analytical amount for every base in DNA. We attempt to use the method in several processing systems and rely on the one that performs the best and has the lowest average testing error.

15- References

[1]Mardis ER., “DNA sequencing technologies,” 2006-2016. Nat Protoc., pp. 213–218, Feb. 2017, doi: 10.1038/nprot.2016.182.

[2] J. A. Reuter, D. v. Spacek, and M. P. Snyder, “High-Throughput Sequencing Technologies,” Molecular Cell, vol. 58, no. 4. Cell Press, pp. 586–597, May 21, 2015. doi: 10.1016/j.molcel.2015.05.004.

[3] H. A. Lewin et al., “Earth BioGenome Project: Sequencing life for the future of life,” Royal Botanic Gardens, vol. 115, 2018, doi: 10.1073/pnas.1720115115/-/DCSupplemental.

[4] D. G. Feitelson and M. Treinin, “The Blueprint for Life? A living cell is the product of complex interactions between DNA and the cellular

environment. Can researchers recreate the environment based on the information contained in the DNA alone?,” 2002.

[5] P. L. Privalov and C. Crane-Robinson, “Forces maintaining the DNA double helix,” European Biophysics Journal, vol. 49, no. 5, pp. 315–321, Jul. 2020, doi: 10.1007/s00249-020-01437-w.

[6] M. Rosello et al., “Precise base editing for the in vivo study of developmental signaling and human pathologies in zebrafish.,” Elife, vol. 10, pp. 1–27, Feb. 2021, doi: 10.7554/eLife.65552.

[7] P. Zheng, S. Wang, X. Wang, and X. Zeng, “Editorial: Artificial Intelligence in Bioinformatics and Drug Repurposing: Methods and Applications,” Frontiers in Genetics, vol. 13. Frontiers Media S.A., Mar. 17, 2022. doi: 10.3389/fgene.2022.870795.

[8] V. Gligorijević and N. Pržulj, “Methods for biological data integration: Perspectives and challenges,” Journal of the Royal Society Interface, vol. 12, no. 112. Royal Society of London, Nov. 06, 2015. doi: 10.1098/rsif.2015.0571.

[9]Haznedar, B., Arslan, M. T., & Kalınlı, A. (2018). Using adaptive neuro-fuzzy inference system for classification of microarray gene expression cancer profiles.

[10] Bamisaye, A., Adesina, M. O., Alfred, M. O., Idowu, M. A., Adeleke, O., & Adegoke, K. A. (2023). Antimicrobial study and biosorption of Pb²⁺ ions onto chitosan-walnut composites: mechanistic studies and neuro-fuzzy modeling approach. Biomass Conversion and Biorefinery, 1-19.

[11] Mu’azu, M. A. (2022). In Situ Skin Friction Capacity Modeling with Advanced Neuro-Fuzzy Optimized by Metaheuristic Algorithms. Geotechnics, 2(4), 1035-1058.

[12] Cao, Y., Pourrostan, T., Zandi, Y., Denić, N., Ćirković, B., Agdas, A. S., ... & Milic, M. (2021). Analyzing the energy performance of buildings by neuro-fuzzy logic based on different factors. Environment, Development and Sustainability, 1-25.

[13]Sharafati, A., Haghbin, M., Tiwari, N. K., Bhagat, S. K., Al-Ansari, N., Chau, K. W., & Yaseen, Z.

- M. (2021). Performance evaluation of sediment ejector efficiency using hybrid neuro-fuzzy models. *Engineering Applications of Computational Fluid Mechanics*, 15(1), 627-643.
- [14] C. Ledergerber and C. Dessimoz, "Base-calling for next-generation sequencing platforms," *Brief Bioinform*, vol. 12, no. 5, pp. 489–497, Sep. 2011, doi: 10.1093/bib/bbq077.
- [15] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning," *Nat Biotechnol*, vol. 33, no. 8, pp. 831–838, Aug. 2015, doi: 10.1038/nbt.3300.
- [16] Y. Guo, F. Ye, Q. Sheng, T. Clark, and D. C. Samuels, "Three-stage quality control strategies for DNA re-sequencing data," *Brief Bioinform*, vol. 15, no. 6, pp. 879–889, Aug. 2013, doi: 10.1093/bib/bbt069.
- [17] J. Zeng, H. Cai, H. Peng, H. Wang, Y. Zhang, and T. Akutsu, "Causalcall: Nanopore Basecalling Using a Temporal Convolutional Network," *Front Genet*, vol. 10, Jan. 2020, doi: 10.3389/fgene.2019.01332.
- [18] P. Liao, G. A. Satten, and Y. J. Hu, "PhredEM: a phred-score-informed genotype-calling approach for next-generation sequencing studies," *Genet Epidemiol*, vol. 41, no. 5, pp. 375–387, Jul. 2017, doi: 10.1002/gepi.22048.
- [19] M. A. Depristo et al., "A framework for variation discovery and genotyping using next-generation DNA sequencing data," *Nat Genet*, vol. 43, no. 5, pp. 491–501, May 2011, doi: 10.1038/ng.806.
- [20] "Neuro-Fuzzy Systems: Review And Prospects." [Online]. Available: <https://www.researchgate.net/publication/2400542>
- [21] D. Pamučar, D. Božanić, A. Puška, and D. Marinković, "APPLICATION OF NEURO-FUZZY SYSTEM FOR PREDICTING THE SUCCESS OF A COMPANY IN PUBLIC PROCUREMENT," *Decision Making: Applications in Management and Engineering*, vol. 5, no. 1, pp. 135–153, 2022, doi: 10.31181/dmame0304042022p.
- [22] G. Arji et al., "Fuzzy logic approach for infectious disease diagnosis: A methodical evaluation, literature and classification," *Biocybern Biomed Eng*, vol. 39, no. 4, pp. 937–955, Oct. 2019, doi: 10.1016/j.bbe.2019.09.004.
- [23] S. A. Saadat, S. M. Ghamari, H. Mollae, and F. Khavari, "Adaptive neuro-fuzzy inference systems (ANFIS) controller design on single-phase full-bridge inverter with a cascade fractional-order PID voltage controller," *IET Power Electronics*, vol. 14, no. 11, pp. 1960–1972, Aug. 2021, doi: 10.1049/pel2.12162.

الحصول على فريد من خلال تحليل الحمض النووي باستخدام خوارزمية عصبية ضبابية

ثائر خير الله عصمان¹ ، هديل محمد صالح² ، علاء حميد محمد³

¹جامعة الانبار/ كلية الآداب

²جامعة الانبار /مركز التعليم المستمر

³كلية المعارف /قسم هندسة تقنيات الاجهزة الطبية

Thaar.asman@uoanbar.edu.iq

Haddeel.mohammed@uoanbar.edu.iq

Aallaaha12@gmail.com

الخلاصة:

في هذه الدراسة قد تم مناقشة مشاكل التمثيل الدقيق والاختبار الجدير بالثقة للحكم على استدعاء قيمة الحمض النووي. تمت دراسة هذه المشكلة من خلال استخدام نموذج لجمع البيانات وتطوير التصميم المناسب وفقاً لمعايير المشكلة المطروحة. الارتفاع والذروة والتباعد بين المرشح الأول الأكثر ترجيحاً (القيمة) بالإضافة إلى ذروة وارتفاع ثاني أكثر المرشحين ترجيحاً و هي متغيرات الإدخال الثلاثة في نموذجنا. من خلال جمع البيانات من الاستدعاء الأساسي ، يمكن حساب هذه السمات الثلاث في الأنظمة الفرعية الثلاثة واستخدامها في النظام الرئيسي لحساب قيمة التحليل لكل قيمة في تسلسل الحمض النووي. تم استخدام مبرمج ماتلاب لإنشاء هذا النموذج. الحمض النووي هنا يبحث في تطبيق خوارزمية الضبابية العصبية لحل المشكلة. تحقق هذه الأساليب نتائج ناجحة مع أداء عالٍ في مكان العمل وسهولة الوصول إلى بيانات دقيقة عالية.