Journal of Economics and Administrative Sciences
2024; 30(142), pp. 425-448

P-ISSN 2518-5764
E-ISSN 2227-703X

Journal of Economics and Administrative Sciences (JEAS)

Available online at http://jeasiq.uobaghdad.edu.iq
DOI: https://doi.org/10.33095/ahnw8r72

The Fuzziness Models with The Proposed New Conjugate Gradient Method for The Classification of High-Dimensional Data in Bioinformatics

Omar Qusay Alshebly *
Department of Statistics
College of Administration and Economics
University of Baghdad, Baghdad , Iraq
omarqusay@uomosul.edu.iq
***Corresponding author**

Suhail Najm Abdullah
Department of Statistics
College of Administration and Economics
University of Baghdad, Baghdad, Iraq
dr.suhail.najm@coadec.uobaghdad.edu.iq

Received: 18/10/2023 Accepted:24/12/2023 Published Online First: 30 /8/ 2024

Abstract:

The development of the subject of bioinformatics may be attributed to the exponential growth of biological data, namely the huge amount of high-dimensional gene expression data. The discipline of bioinformatics efficiently tackles issues in molecular biology through the use of optimization, computer science, and statistical methods. The present study introduces a new optimization strategy, namely the proposed conjugate gradient method (PNCG), for the purpose of learning a fuzzy neural network model using the Takagi-Sugeno approach. This study presented a novel algorithm that addresses the issue of delayed convergence seen in the Polak Ribière Polyak (PRP) and Liu-Storey (LS) techniques by using the PRP method. The research used simulated and real datasets to empirically evaluate the suggested method. The results of the study demonstrated that the suggested method exhibited superior performance compared to other well-established methods. The experiment included the use of three publicly available datasets related to cancer. The findings indicate that the proposed approach is both extremely efficient and feasible, hence exhibiting a significant degree of efficacy in terms of average training dataset time, average training dataset accuracy, average testing dataset accuracy, average training dataset mean squared error (MSE), and average testing dataset MSE.

Paper type: Research paper
Keywords: Bioinformatics, Optimization, Fuzzy Neural Networks, Accuracy, Conjugate Gradient, Gene Expression Data.

P-ISSN 2518-5764
E-ISSN 2227-703X

1.Introduction:

Bioinformatics is widely recognized as a crucial and influential discipline that has emerged via the use of advanced mathematical techniques, computerization, and statistical modelling in the area of molecular biology. Its primary objective is to accurately understand illnesses, their underlying causes, and potential treatment strategies. The advent of DNA microarray technology in recent years has proven to be a significant breakthrough in the fields of biology, medicine, bioinformatics, and genetics(Honrado et al, 2006; Noor Alhuda and Basad, 2022; Bhola and Singh, 2018). The microarray datasets exhibit a significant sparsity and feature a large number of dimensions. The challenges connected with these datasets include redundancy, noise, and complexity(Kim and Kim, 2019).

The objective of fuzzy modelling is to identify a set of localized relationships between input and output variables in order to characterize a given process. In comparison to conventional modelling techniques that rely on differential equations, fuzzy modelling offers improved capability in representing nonlinear processes (Chakraborty et al, 2014). Therefore, the task of identifying a fuzzy model may be seen as a challenge of using input-output (I-O) data to determine the model's parameters(Sugeno and Kang, 1998). Generally speaking, the identification process of a fuzzy logic system (FLS) or a fuzzy neural (neuro-fuzzy) network system model may be categorized into two main steps. The identification of structure and parameters(Taniguchi et al, 2001). The primary focus of structure identification is to determine the configuration of a fuzzy model, including the determination of the number of fuzzy rules and the membership functions linked to the premise and consequent fuzzy sets inside each rule. There are several methodologies for the identification of structural elements. One approach involves the use of clustering techniques to derive rules from the given input-output dataset in order to establish the first rule base. The fuzzy rule base of a fuzzy logic system (FLS) is frequently generated by the use of several clustering techniques, such as the K-means algorithm (Bezdek et al, 1999), fuzzy c-means (FCM) (Gu and Wang, 2018), and the mountain clustering method (Yager and Filev, 1994). The fundamental concept behind the detection of structure by clustering methods is to partition the provided samples into distinct clusters, with each cluster representing a distinct rule. In other words, the quantity of rules is equivalent to the quantity of clusters. The use of clustering methodology for structure identification necessitates the pre-collection of data, rendering it unsuitable for real-time online structure identification. Numerous scholarly investigations have been conducted (Juang and Lin, 1998; Shahparast et al, 2019) that center on the use of fuzzy neural networks for dynamic system modeling. Furthermore, the Bayesian TSK fuzzy model, as presented in reference (GU et al, 2016), has the capability to determine the appropriate number of fuzzy rules without relying on any previous expert knowledge or experience. This research focuses on the use of clustering methods for the purpose of identifying structures in order to address stable classification difficulties.

1.1 Literature review:

The "Gradient based Neuro-Fuzzy learning algorithm" (GNF) has been widely used for parameter identification in neuro-fuzzy systems, much like the training process of feed forward neural networks. The work conducted by Ichihashi and Türksen (1993) and Mendel (2013) influenced by the generalized neuro-fuzzy framework for neuro-fuzzy systems.

Wu et al (2010) proposed a modification of the GNF (MGNF). in order to prevent the singularity, the form of the error function has been modified to include the reciprocals of the widths of Gaussian membership functions as independent variables. As a result, the weight sequence formulae have been modified in an easy way. This conversion helps the convergence analysis of the MGNF algorithm.

P-ISSN 2518-5764
E-ISSN 2227-703X

De aguiareight et al (2016) proposed conjugate gradient methods and have used them to improve the performance of the type-1 fuzzy logic system in order to tackle classification challenges. The simulation results obtained indicate that the rate of convergence of conjugate gradient (CG) methods is higher compared to that of the gradient descent technique. Furthermore, the classification results attained by the CG-based fuzzy logic system (FLS) demonstrate superior quality in comparison to those produced by the optimized FLS using the gradient descent technique.

Gao et al (2018) discussed CG-based neuro-fuzzy network model showed competitive performance in regression issues. Furthermore, significant advancements have been made in the field of PRP conjugate gradient based fuzzy neural networks, particularly in the areas of weak and strong convergence. In addition, in the context of gradient-based parameter identification methods, an essential consideration is the determination of an ideal learning rate during the training procedure.

Ahmad et al (2020) conducted a novel adaptation of the variational iteration algorithm-II was introduced in order to compute the numerical solutions for certain nonlinear equations, such as the Korteweg-de Vries (KdV), modified Korteweg-de Vries (mKdV), and combined KdV-mKdV equations.

Last but not least, Salleh et al (2022) provided a new conjugate gradient approach that exhibits characteristics of simplicity, efficiency, and robustness. The approach proposed by Liu and Storey effectively addresses convergence issues and fulfills both convergence qualities and descent criteria. The numerical findings provide evidence of its higher performance compared to other CG approaches, namely in terms of the number of iterations required and the value of CPU time used.

The problem addressed in this article is the difficulty involved in constructing and classifying high-dimensional cancer data using both fuzzy logic (FL) and hybrid technology (ANFIS). The complexity comes from the architectural design, which is based on the use of IF-THEN statements for rule development. The use of these methods in the construction and classification of data is time-intensive work, largely owing to the substantial number of steps in rule formulation involved. Moreover, the task at hand involves the establishment of these regulations based on an adequate scientific basis, given that harmful diseases are not restricted strictly to certain genes or preset proportions. To address this issue, an approach referred to as the conjugate gradient was developed through the use of fuzzy modeling and numerical optimization techniques.

The objective of this study is to propose a new algorithm based on the Polak-Ribière-Polyak (PRP) method for training a fuzzy neural network model. The application aims to maximize the accuracy of cancer disease prediction while simultaneously reducing computational time and achieving the lowest mean square error for training and testing datasets. A comparison study is conducted to evaluate the performance of the provided method.

1.2 Artificial Neural Networks:

Recent scientific research has shown that the cerebral cortex is comprised of a staggering 22 billion neurons and an astonishing 220 trillion synaptic connections. The study of this particular neuron, which plays a crucial role in neuronal function, serves as a source of inspiration for neuroscientists, computer engineers, and psychologists who are driven to replicate the human mind via the construction of a computational framework for supercomputers that emulates the intricate workings of the human brain (Awad et al, 2022).

Artificial neural networks (ANNs) are computational systems designed to simulate and replicate the functioning of natural neural networks seen in humans and other organisms. A network of synchronized and interconnected linkages creates the connection between these pieces. The artificial neural network is a dynamic system that exhibits adaptability by modifying its structure in response to the information it encounters during the learning phase(Ashour and Dahhan, 2022).

P-ISSN 2518-5764
E-ISSN 2227-703X

Figure 1 serves as a visual representation of the fundamental principles behind the functioning of artificial neural networks.

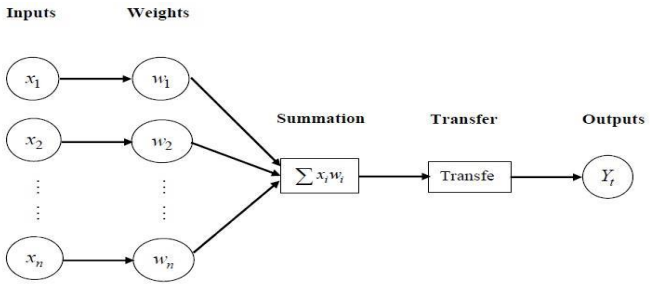


Figure 1: Artificial Neural Networks

The following can be determined from Figure (1):

1- The Input Layer contains x values.

2- The Hidden Layer slide contains the values of w and the resulting operations.

3- The Output Layer contains the value y (DA Silva et al, 2017).

The practical use of these networks is in the potential to employ algorithms that modify the synaptic weights between artificial neurons, hence inducing an appropriate response (Wu and Mclarty, 2001).

$$net_{pj}^{L+1} = \sum_{i=1}^{n^L} w_{ij}^{L}\, out_i^{L} + bias_j^{L+1} \quad (1)$$

$$out_{pi}^{L+1} = f\left(net_{pj}^{L+1}\right) = \frac{1}{1+e^{-\beta\, net_{pj}^{L+1}}} \quad (2)$$

whereas:-

net_{pj}^{L+1}**:** Total number of entries multiplied by unit j's weight in layer (L + 1).

out_{pi}^{L+1}**:** The output of the layer (L + 1).

w_{ij}^{L}: Weights of layer L.

$bias_j^{L+1}$: Error value between layers.

1.3 Fuzzy Logic:

Theories and techniques that use fuzzy sets, which are groups without definite limits, as a convenient alternative to the crisp set, which no longer meets the requirements of the new mathematical and logical understanding in our contemporary scientific thought, This concept originated in 1965 by the scientist of Azerbaijani origin, Lotfi Zadeh(Zadeh, 1965), from the University of California, which he developed to use in a better way to process data by applying a more human-like way of thinking in data programming (Sivanandam et al, 2007).

1.3.1 Membership Function:

For a set A, we define a membership function $\mu_A(\mathrm{x})$ such as:

$$\mu_A(\mathrm{x}) = \begin{cases} 1, & \text{iff } x \in A \\ 0, & \text{iff } x \notin A \end{cases}$$

We can say that the function $\mu_A(\mathrm{x})$ maps the elements in the universal set X to the set {0,1}.

$\mu_A(\mathrm{x}) : \mathrm{X} \longrightarrow \{0,1\}$.

The basis condition membership function is that its range lies between zero and one (Klir et al, 1997).

P-ISSN 2518-5764
E-ISSN 2227-703X

1.3.2 Neuro-Fuzzy:

Neuro-fuzzy hybridization is made possible by a hybrid intelligent system that combines the cognitive reasoning approach of fuzzy systems with the ability to adapt to new information and the interconnected structure of neural networks. The combination of neuro-fuzzy techniques is often referred to as a fuzzy artificial neural network (FANN) or a neuro-fuzzy system (NFS). The phrase "neuro-fuzzy system" is often used to refer to a computational approach that combines the reasoning style of fuzzy systems with the use of fuzzy sets and a linguistic model composed of a collection of IF-THEN fuzzy rules. Neuro-fuzzy systems possess a notable advantage in their capacity as universal approximations, enabling them to effectively generate interpretable IF-THEN rules (Khudhur and Abbo, 2020). Both neural networks and fuzzy systems mimic the human thought process, and they have some common qualities in their applications. For example, they solve various issues such as pattern recognition, regression, classification, and other tasks common to these two technologies, and each of them has its disadvantages and advantages. The purpose of combining these two technologies is to try to hide the disadvantages present in both technologies by complementing the shortcomings of one of these two technologies with the advantages of the other.

Another advantage of neuro-fuzzy systems is their capacity to address the conflicting demands of interpretability and accuracy within fuzzy modeling. In practical application, one of the two traits predominates. The subject of study in fuzzy modeling within the domain of neuro-fuzzy may be broadly categorized into two distinct areas. The first area, known as linguistic fuzzy modeling, places a strong emphasis on interpretability. This particular aspect of research was primarily pioneered by the Mamdani model. On the other hand, the second area, referred to as precision fuzzy modeling, prioritizes precision in its approach. The Takagi-Sugeno model has played a significant role in leading advancements in this area.

The adaptive neuro-fuzzy inference system architecture, which is abbreviated (ANFIS), has been proposed as a tool to build a set of fuzzy rules with appropriate membership functions.

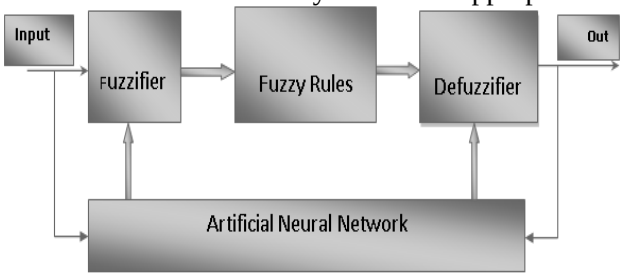


Figure 2: Architecture of The Adaptive Neuro-Fuzzy Inference System(ANFIS)

The ANFIS model converts the characteristics of input into membership functions and converts these functions into rules and then properties for output, which in turn become membership functions for output and then the decision accompanying the final output of the system.

1.3.3 Mathematical Modeling of Zero-Order Takagi–Sugeno Inference System:

The neuro-fuzzy model is a representation of a fuzzy inference system that may be conceptualized as an adaptive network. The neuro-fuzzy model examined in this study is the zero-order Takagi-Sugeno inference system. The topological structure is seen in Figure 3. The network consists of four layers, including of m input nodes denoted as $x = (x_1, x_2, \dots, x_m) \in \mathbb{R}^m$, and a single output node represented as y.

Initially, a comprehensive depiction of the zero-order Takagi-Sugeno inference system was presented. The primary area of our investigation pertains to a fuzzy rule basis, which is precisely defined as follows. (Del Campo et al, 2008; Chaturvedi et al, 2008; Tang et al, 2007):-

Rule i: IF x_1 is A_{1i} and x_2 is A_{2i} and … and x_m is A_{mi}THEN y is y_i (3)

P-ISSN 2518-5764
E-ISSN 2227-703X

where i $(i = 1,2,\ldots,n)$ corresponds to the i^{th} fuzzy rule, n is the number of the fuzzy rules, y_i is a real number, Ali is a fuzzy subset of xl, and $Ali(xl)$ means a Gaussian membership function of the fuzzy decision ''$xl\ is\ Ali$" can be defined:-

$$A_{li} = \frac{\exp(-(x_l - a_{li})^2}{\sigma_{li}^2} \quad (4)$$

where ali is the center of $Ali(xl)$, and σli is the width of $Ali(xl)$.

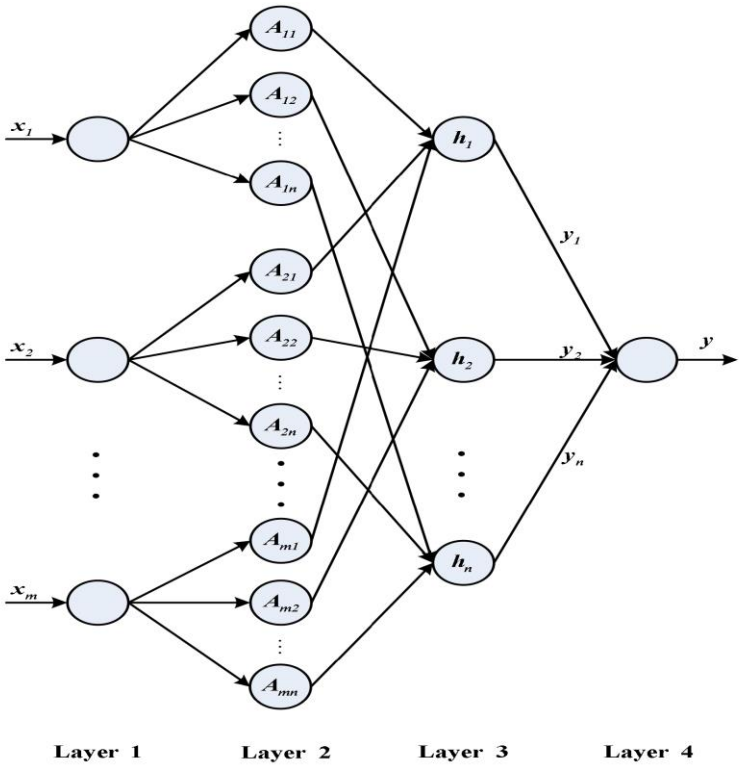


Figure 3: Structure of The Zero-Order Takagi–Sugeno Inference System (Gao et al, 2019).

In the context of a specific observation $x = (x_1, x_2, \ldots, x_m)$, the zero-order Takagi-Sugeno inference system specifies the functions associated with the nodes in this particular model as follows:-

Layer1 (input layer): Each individual neuron within this particular layer is responsible for representing a single input variable. The input variables are immediately propagated to the subsequent layer.

Layer2 (membership layer): Within this particular layer, it can be seen that each individual node functions as a memory unit, representing the membership function of a linguistic variable. In this study, the membership functions of the nodes are implemented using Gaussian functions. The weights that establish the connection between Layer 1 and Layer 2 may be regarded as representing the centers and widths of the Gaussian membership function, respectively.

Layer3 (rule layer): This layer consists of rule nodes, each of which represents a rule term. For $i = 1,2,\ldots,n$ The agreement of the ith antecedent part is computed as:-

$$h_i = h_i(x) = A_{1i}(x_1)A_{2i}(x_2) \ldots A_{mi}(x_m) = \prod_{l=1}^{m} A_{ii}(x_l) \quad (5)$$

The link weights connecting layers2 and 3 are fixed as constant 1.

P-ISSN 2518-5764
E-ISSN 2227-703X

Layer4 (output layer): The function of this layer is to do a summed-weight defuzzification procedure. The output of this layer corresponds to the ultimate outcome, denoted as y, and is obtained by taking a linear combination of the outcomes produced by Layer3.

$$y = \sum_{i=1}^{n} h_i y_i \quad (6)$$

The weights y_i that are linked to the output layer are often known as conclusion parameters.
Remark (1): In the original neuro-fuzzy models (Jang, 1993; Luo et al, 1998) the ultimate consequence y is determined by the use of the gravity approach, which is described as follows:-

$$y = \frac{\sum_{i=1}^{n} h_i y_i}{\sum_{i=1}^{n} h_i} \quad (7)$$

In order to facilitate the learning process, it is sometimes advantageous to acquire the fuzzy consequence without the need to calculate the center of gravity. Consequently, the denominator in equation (7) is excluded (Chaturvedi et al, 2008). One additional benefit of this procedure is its straightforward hardware implementation (Juang and Chen, 2006). Therefore, we will use the structure denoted by equation (6) in all our subsequent analyses and dialogues.
The error function is formally defined as:-

$$E(w) = \frac{1}{2}\sum_{j=1}^{J}(y^j - o^j)^2$$

The variable o^j represents the intended output for the j^{th} training dataset x_j, y_j represents the associated fuzzy reasoning outcome, and J represents the total number of training dataset. The objective of network learning is to identify the optimal weight matrix W* that minimizes the expected error E(W*) to a desired threshold m. The gradient descent technique is sometimes used to address this optimization issue (Sahiner and Ibrahem, 2018; 2019; Abdollahi and Fatemi, 2021).

2.Material and Methods:

2.1 The Proposed New Conjugate Gradient (PNCG) Method:

This study focuses on the utilization of the Polak Ribière Polyak (PRP) method to enhance the efficiency of fuzzy neural networks in the field of data classification. The study also involves doing a comparative analysis of this algorithm with other existing optimization techniques.

$$w_{k+1} = w_k + \alpha_k d_k, \quad k \geq 1 \quad (8)$$

The step-size, denoted as α_k, is determined using a line search, whereas d_k represents the search direction provided by the algorithm.

$$d_{k+1} = \begin{cases} -g_1, & k = 1 \\ -g_{k+1} + \beta_k d_k, & k \geq 1 \end{cases} \quad (9)$$

where β_k is a parameter.

The parameters that will be used in our study are $\beta^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}$,(Polak and Ribiere, 1969). $\beta^{LS} = \frac{g_{k+1}^T y_k}{g_k^T d_k}$,(Liu and Storey, 1991). Where work be carried out on the development of a PRP method and comparison with the previous two methods. Also, $g_k = \nabla E(w_k)$denotes the gradient of the function of error $E(w)$ in regard to w, k the number of iterations, and let $y_k = g_{k+1} - g_k$.
In this proposal, we provide an alternative conjugate gradient technique for the purpose of data classification. This approach is primarily based on the Polak Ribière PolYak (PRP) algorithm. As a result, we derive a novel formula:-

P-ISSN 2518-5764
E-ISSN 2227-703X

$$-g_{k+1}+\beta_k d_k=-\lambda g_{k+1}+\beta_k^{PRP} d_k$$
$$-g^T{}_{k+1}g_{k+1}+\beta_k g^T{}_{k+1}d_k=-\lambda g^T{}_{k+1}g_{k+1}+\beta_k^{PRP} g^T{}_{k+1}d_k$$

$$\beta_k^{PNCG}=\begin{cases}\frac{(1-\lambda_k)g^T{}_{k+1}g_{k+1}}{g^T_{k+1}d_k}+\beta_k^{PRP}, & if \quad g^T_{k+1}d_k\neq 0\\ \beta_k^{PRP}, & if \quad g^T_{k+1}d_k=0\end{cases}$$

Where $\lambda \in [0,1]$.

$$d_{k+1}=-g_{k+1}+(\frac{(1-\lambda)g^T{}_{k+1}g_{k+1}}{g^T{}_{k+1}d_k}+\frac{g^T_{k+1}y_k}{g^T_k g_k})d_k \quad (10)$$

2.1.1 The Descent Property of Proposed New Conjugate Gradient Method:
we have to show the descent property for our proposed new conjugate gradient scheme, denoted by β_k^{PNCG}. In the following:
Theorem (1)

The search direction d_{k+1} and β_k^{PNCG} given in equation:

$$d_{k+1}=-g_{k+1}+\beta_k^{PNCG}d_k$$

$$\beta_k^{PNCG}=\begin{cases}\frac{(1-\lambda_k)g^T{}_{k+1}g_{k+1}}{g^T_{k+1}d_k}+\beta_k^{PRP}, & if \quad g^T_{k+1}d_k\neq 0\\ \beta_k^{PRP}, & if \quad g^T_{k+1}d_k=0\end{cases}$$

$$d_{k+1}=-g_{k+1}+(\frac{(1-\lambda)g^T{}_{k+1}g_{k+1}}{g^T{}_{k+1}d_k}+\frac{g^T_{k+1}y_k}{g^T_k g_k})d_k$$

Where $\lambda \in [0,1]$.It will hold for all $k \geq 1$.
Proof:- The proof is established by the use of mathematical induction.

1- If $k=1$ then $g_1^T d_1<0,\ d_1=-g_1 \rightarrow < 0$.
2- Let the relation $g_k^T d_k<0$ for all k.
3- We prove that the relation is true when $k=k+1$ by multiplying the equation (10) in g_{k+1}^T we obtain:

$$g_{k+1}^T d_{k+1}=-g_{k+1}^T g_{k+1}+\left(\frac{(1-\lambda)g_{k+1}^T g_{k+1}}{g_{k+1}d_k}+\frac{y_k^T g_{k+1}}{g_k^T g_k}\right)g_{k+1}^T d_k$$

Let $\delta=\frac{(1-\lambda)g_{k+1}^T g_{k+1}}{g_{k+1}d_k}, v=\frac{y_k^T g_{k+1}}{g_k^T g_k}$
$g_{k+1}^T d_{k+1}=-g_{k+1}^T g_{k+1}+(\delta+v)g_{k+1}^T d_k$
Let $g_{k+1}^T d_{k+1}>0$ and $\delta > v$
Then
$g_{k+1}^T d_{k+1}\geq 0$.

P-ISSN 2518-5764
E-ISSN 2227-703X

2.1.2 Study of Global Convergence:

In this study, we will demonstrate the global convergence of the CG technique using the parameter β^{PNCG}. Assumptions are necessary in order to establish the convergence of the newly suggested algorithm.

Assumptions (1)

1-Assume f is bound below in the level set $S = \{w \in R^n : E(w) \leq E(w_0)\}$; in some initial point.

2- E is continuously differentiable and its gradient is Lipchitz continuous, there exists $L > 0$ such that:-(Laylani and Khudhur, 2018)

$$\| g(x) - g(y) \| \leq Lx - y \| \ \forall x, y \in N \qquad (11)$$

However, assuming Assumption (1), it becomes evident that there is a positive constant B such that

$$\|w\| \leq B, \forall w \in S \qquad (12)$$

$$\| \nabla E(w) \| \leq \overline{\gamma}, \forall w \in S \qquad (13)$$

Lemma (1)

Given that Assumption (1) is satisfied and equation (12) is valid, we may proceed with our analysis. Consider any conjugate gradient approach described in equations (8) and (9), where d_k represents a suitable direction and α_k is determined using the Strong Wolfe Line Search (SWLS) technique. If

$$\sum_{k>1} \frac{1}{\| d_{k+1} \|^2} = \infty$$

then we have

$$\liminf_{k\to\infty} \| g_k \| = 0$$

Further information may be found in references (Abdullah et al, 2019; Gao et al, 2019).

Theorem (2)

Given that Assumption (1), Equation (8), and the descent condition are satisfied. Let us consider a conjugate gradient scheme expressed in the following form:

$$d_{k+1} = -g_{k+1} + \beta_k^{PNCG} d_k$$

The value of α_k is determined by the application of the strong Wolfe line search condition, as explained in more detail in references (Al-Baali, 1985; Zhang and Zhou, 2008; Gao et al, 2019). If the goal function is evenly distributed across set S, then:

$$lim_{n\to\infty}(inf \| g_k \|) = 0$$

Proof:-

$$d_{k+1} = -g_{k+1} + \beta_k^{PNCG} d_k$$

$$\beta_k^{PNCG} = \frac{(1-\lambda) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \frac{y_k^T g_{k+1}}{g_k^T g_k}$$

$$\| d_{k+1} \| = \| -g_{k+1} + \left(\frac{(1-\lambda) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \frac{y_k^T g_{k+1}}{g_k^T g_k}\right) d_k \|$$

P-ISSN 2518-5764
E-ISSN 2227-703X

$$\| d_{k+1} \| \leq \| g_{k+1} \| + \| \frac{(1-\lambda) g_{k+1}^T g_{k+1}}{g_{k+1}^T d_k} + \frac{y_k^T g_{k+1}}{g_k^T g_k}) \| d_k \|$$

$$\| d_{k+1} \| \leq \| g_{k+1} \| + \| \frac{(1-\lambda) \| g_{k+1} \|^2}{\| d_k \| \| g_{k+1} \|} + \frac{\| y_k^T \| \| g_{k+1} \|}{\| g_k \|^2} \| \| d_k \|$$

$$\| d_{k+1} \| \leq (1 + \| \frac{(1-\lambda) \| d_k \|}{\| d_k \|} + \frac{\| d_k \| \| y_k^T \|}{\| g_k \|^2}) \| g_{k+1} \|$$

$$\psi = \| \frac{(1-\lambda) \| d_k \|}{\| d_k \|} + \frac{\| d_k \| \| y_k^T \|}{\| g_k \|^2} \|$$

$$\| d_{k+1} \| \leq (1 + \psi) \| g_{k+1} \|$$

$$\sum_{k \geq 1} \frac{1}{\| d_{k+1} \|^2} \geq (\frac{1}{(1+\psi)^2}) \frac{1}{\lambda^2} \Sigma 1 = \infty$$

2.2 Metrics and Evaluation of Performance:

Metrics of performance are utilized to assess the effectiveness of classification and regression algorithms. To assess the prediction capability of the suggested approach. The performance of the model in classification problems will be evaluated through the utilization of performance metrics below:

1-Accuracy:

The accuracy can be characterized as the proportion of accurate predictions made relative to the total number of predictions made. The accuracy metric evaluates the proportion of accurate predictions made by the classifier in relation to the total number of data points.

$$Accuracy = (TP + TN) / N$$

(14)

Whereas:-

TN: The scenario arises when the predicted class and the actual class of a data point are both 0.

TP: The scenario occurs when the predicted class and the actual class of a given data point are both equal to 1.

N: Total number of samples. (Dash, 2020; Abd Algafore and Hashem, 2016).

2- Mean Square Error (MSE):

$$MSE = (y - \hat{y})^2 / N \qquad (15)$$

Whereas:-

y:real output; $\hat{y}$:estimation output. (Khalil and Rada, 2023).

3. Results and Discussion:

Both fuzzy logic (FL) and hybrid technology (ANFIS) exhibit complexity in their construction and classification of high-dimensional cancer data. This complexity arises from their architecture, which relies on the formulation of rules in the form of IF-THEN statements. The process of constructing and classifying data using these techniques is time-consuming due to the extensive rule development required. Additionally, the challenge lies in establishing these rules on a solid scientific foundation, as pathological conditions are not confined to specific genes or fixed proportions.

P-ISSN 2518-5764
E-ISSN 2227-703X

In order to address this issue, we used fuzzy modeling and numerical optimization techniques to develop the approach known as conjugate gradient. It is associated with fuzzy neural networks through the Takagi Sugino method of fuzzy neural network inference system, which was discussed and analyzed in our study of the classification of high-dimensional data in bioinformatics.

We have evaluated and compared our proposed supervised learning algorithm, PNCG, described in previous against with two other high-related existing algorithms ,Polak Rebire Polyak (PRP) and Louis Story (LS), on classification for the simulation and real datasets. The microarray data typically exhibits a high degree of correlation, and the use of simulation data enables the assessment of proposed methodologies on such correlated data. Initially, the dataset is partitioned into a training dataset comprising 70% of the total data and a test dataset comprising 30% of the total data.

The computations were conducted by averaging the results from 100 random partitioning iterations. All starting parameters for training these three separate algorithms are similar, save for the unique distinction of the applied optimization techniques. In order to examine the impact of varying numbers of rules, we have selected three distinct examples for each dataset and conducted a comprehensive analysis to evaluate their performance. Also, the learning rate chosen was equal to 0.001.The λ values used were between zero and one, and the best value chosen was 0.5. All the implementations of the study on simulation and real data applications are carried out using R and Matlab.

Figure 3 shows the workflow diagram of the study algorithm.

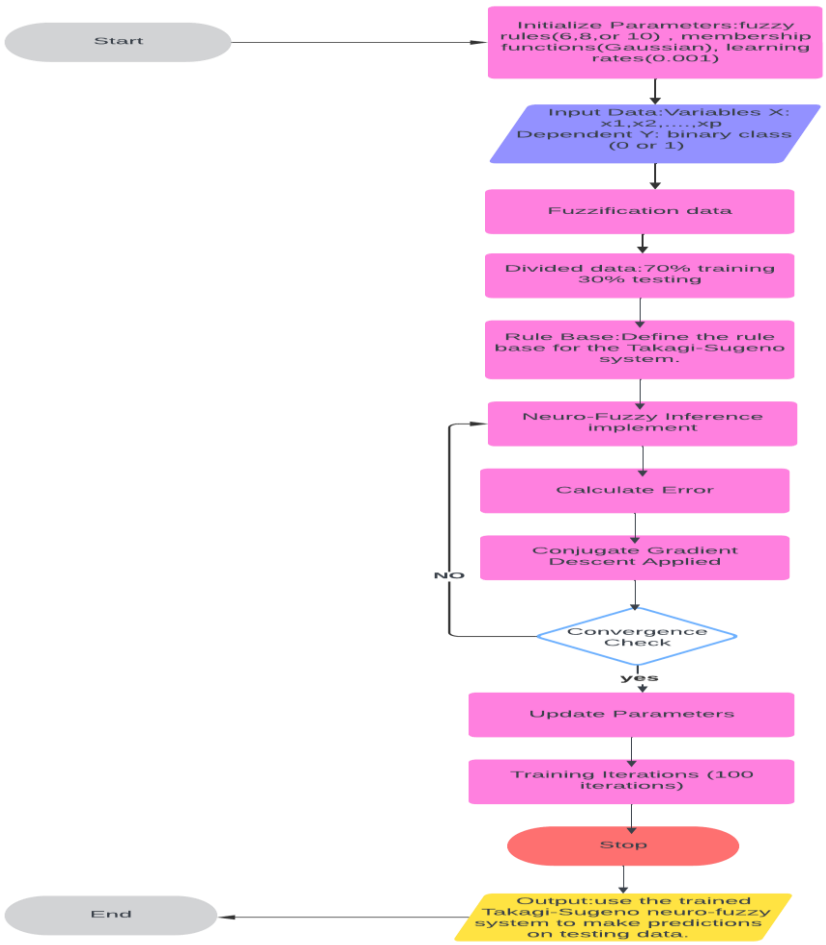


Figure 4: Diagram of the Algorithm steps

P-ISSN 2518-5764
E-ISSN 2227-703X

3.1 Results of Simulation:

The utilization of the multivariate normal distribution within an auto-regressive correlation structure enables the generation of modeling data. The correlation matrix is a mathematics concept that is described formally as follows:

$$\sum = \begin{pmatrix} \rho_{11} & \rho_{12} & \cdots\cdots\cdots & \rho_{1g} \\ \rho_{21} & \rho_{22} & \cdots\cdots\cdots & \rho_{2g} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots\cdots & \cdots\cdots \\ \rho_{s1} & \rho_{s2} & \cdots\cdots\cdots & \rho_{sg} \end{pmatrix} \quad (16)$$

The equation that is used to produce the binary response variables is as follows:-

$$\pi_i(y_i = 1|x_i) = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)} \quad (17)$$

The variable of interest is denoted as a set of significant variables(x_i), while the response variable y_i is produced through Bernoulli experiments, as demonstrated in Equation (17). A total of 100 datasets were generated, with a sample size of n=100 and a variable dimension of d=1000. The regression coefficients were derived from a uniform distribution with a lower bound of 2 and an upper bound of 4 (Pi and Halabi, 2018). The equation that represents the correlation between the i^{th} and j^{th} variables is $\rho_{ig} = \rho^{|i-j|}$ where i and j are integers ranging from 1 to g. by employing this high-dimensional configuration, we obtained data exhibiting correlations of 0.1, 0.4, and 0.7.

Table1: Average Performance Comparison for Classification Simulation datasets with 100 iterations

Correlation	Methods	No. of Rules	Average training dataset time	Average training dataset accuracy	Average testing dataset accuracy	Average training dataset MSE	Average testing dataset MSE
0.1	LS	*4*	38.7679	0.5300	0.4800	0.8021	0.8223
	PRP	*4*	37.6188	0.7414	0.5767	0.3831	**0.4837**
	PNCG	*4*	**26.8867**	**0.7643**	**0.5900**	**0.3506**	0.4888
	LS	*6*	26.2600	0.5229	0.4933	0.7317	0.7288
	PRP	*6*	26.4048	0.4871	0.5000	0.9571	0.9578
	PNCG	*6*	**25.8895**	**0.7686**	**0.6567**	**0.3331**	**0.4575**
	LS	*10*	42.5881	0.4914	0.5700	1.0357	1.0333
	PRP	*10*	43.3269	0.4912	**0.5710**	1.0042	1.0040
	PNCG	*10*	**42.1267**	**0.5457**	0.4800	**0.6748**	**0.6957**
0.4	LS	*4*	**18.1882**	0.4757	0.4100	1.0132	1.0129
	PRP	*4*	18.1941	0.7886	**0.7333**	0.2956	0.3844
	PNCG	*4*	18.8063	**0.8000**	0.7200	**0.2944**	**0.3810**
	LS	*6*	26.6003	0.5029	0.4900	0.7044	0.7134
	PRP	*6*	26.7262	0.5043	0.4800	0.9740	0.9783
	PNCG	*6*	**26.5204**	**0.7271**	**0.7200**	**0.3461**	**0.3611**
	LS	*10*	43.6363	0.4843	**0.5333**	1.1958	1.1959
	PRP	*10*	44.1183	0.4840	0.5330	1.1655	1.1610
	PNCG	*10*	**42.7262**	**0.5257**	0.4600	**0.7802**	**0.7800**
0.7	LS	*4*	**11.3203**	0.5314	0.4767	1.0193	1.0212
	PRP	*4*	11.7304	0.6514	0.6367	0.8399	0.8407
	PNCG	*4*	16.9947	**0.8743**	**0.8433**	**0.2009**	**0.2172**
	LS	*6*	**19.9105**	0.5086	0.5133	0.8238	0.8245
	PRP	*6*	20.4252	0.4914	0.4867	1.0629	1.0637
	PNCG	*6*	27.1979	**0.7671**	**0.7533**	**0.4039**	**0.4040**
	LS	*10*	27.2106	0.5000	**0.5067**	1.0215	1.0214
	PRP	*10*	**13.2280**	0.4943	0.4933	1.0008	1.0008
	PNCG	*10*	47.8960	**0.5086**	0.4967	**0.9771**	**0.9768**

P-ISSN 2518-5764
E-ISSN 2227-703X

Table 1 displays the calculated values of average training dataset time, average training dataset accuracy, average testing dataset accuracy, average training dataset mean squared error (MSE), and average testing dataset MSE for all the approaches. To establish the effectiveness of the suggested new conjugate gradient approach in conjunction with other famous methods, we performed a study using simulated data consisting of three separate correlation values and three rules(4,6,10). we demonstrated the superior performance of PNCG across all metrics. we find that the greater the number of rules, the greater the actual execution time.

Upon examination of the simulation results shown in the aforementioned table, it is evident that the most favorable outcomes for the classification of high-dimensional data were achieved when the number of rules equaled 6. In this scenario, the suggested approach exhibited superior performance compared to the PRP and LS methods. In the dataset under consideration, the correlation coefficient between variables was determined to be 0.1, while the total number of rules amounted to 6. The average duration of execution was estimated to be around 26 minutes. Furthermore, the accuracy rate for the training data was found to be 77%, whereas for the test data it was 66%. The MSE for the training dataset were calculated to be 0.33, while for the test dataset the value was approximately 0.45.

On the other hand, when the correlation coefficient was 0.4 and the number of rules was 6,the average execution time was approximately 27 minutes, and the average accuracy of the training and testing datasets (73%, 72%), respectively .The average error squares for the training and testing data are 0.34 and 0.36. In addition, the results for the correlation coefficient equal to 0.7 also showed superiority of the proposed method with all the performance measures used in our study.

In conclusion, our analysis demonstrates that the approach suggested has shown superior performance compared to the other methods used across the majority of the findings. However, it is worth noting that there is a slight increase in execution time associated with this method.
Figures 5, 6, and 7 show the average accuracy of the training dataset and all the rules used for the simulation dataset over three correlation coefficient values, with of 100 iterations.

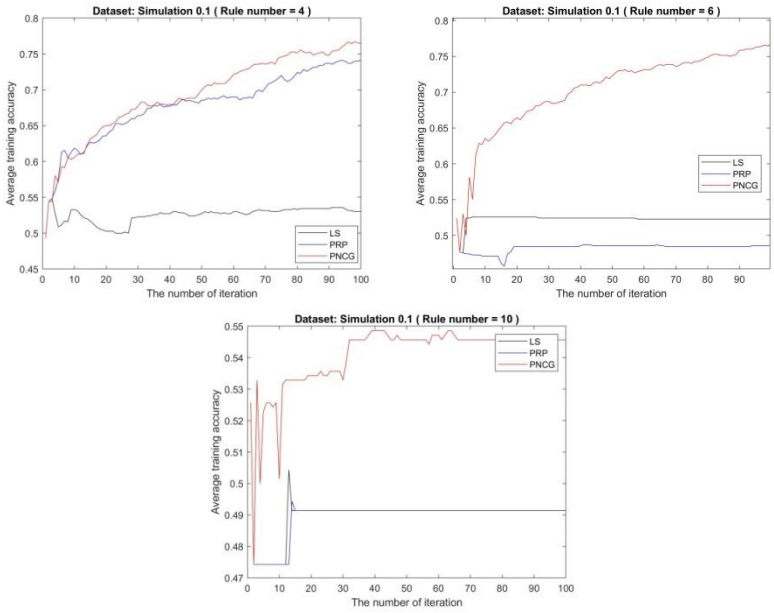


Figure 5: The average training accuracy for simulation dataset with ρ=0.1 and three rules(4,6,10)

P-ISSN 2518-5764
E-ISSN 2227-703X

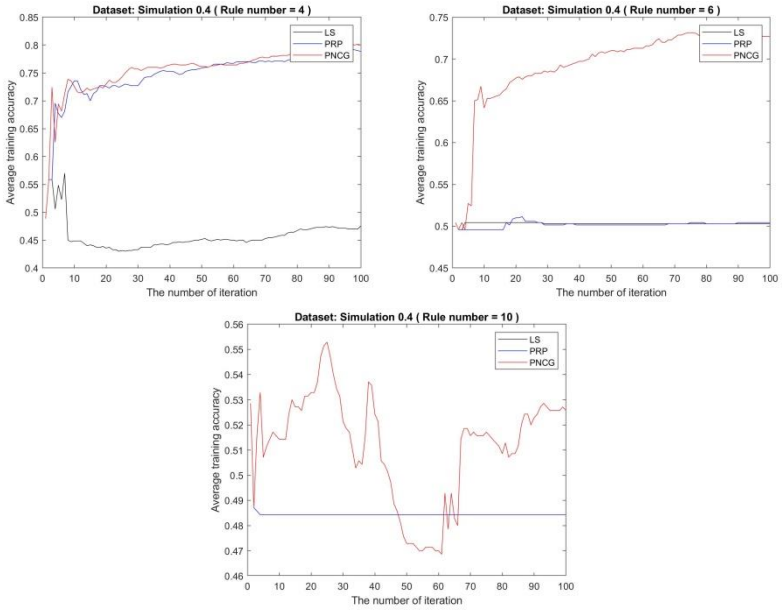


Figure 6: The average training accuracy for simulation dataset with ρ=0.4 and three rules(4,6,10)

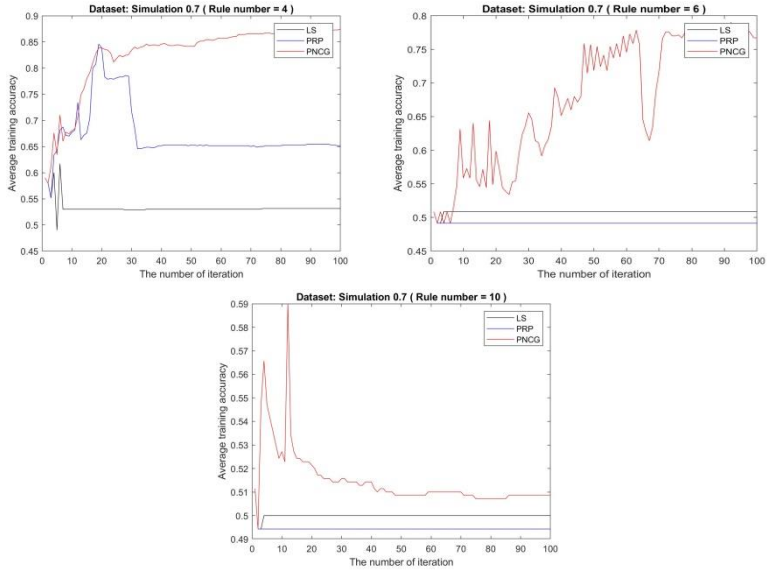


Figure 7: The average training accuracy for simulation dataset with ρ=0.7 and three rules(4,6,10)

3.2 Real Dataset Collection:

The datasets used in the present work included microarray gene expression data. The datasets included in this work are linked to three publicly accessible gene expression datasets characterized by a large number of dimensions. The datasets include colon cancer (Alon et al, 1999), leukemia data (Golub et al, 1999), and prostate cancer (Singh et al, 2002). The datasets being analyzed consist of a response variable that has been classified into two different categories (0,1). The decision to use the prior dataset in this study was driven by the absence of available data equivalent to that within the specific geographic context of Iraq.

Table 2 to describe the real data used in our study

P-ISSN 2518-5764
E-ISSN 2227-703X

Table 2: Presents a comprehensive summary of the three gene expression datasets

Dataset	No. of samples(observations)	No. of genes(variables)	Class
Colon Cancer	62	2000	40 tumor / 22 normal
Leukemia	73	7129	48 ALL / 25 AML
Prostate Cancer	102	12600	52 tumor / 50 normal

The provided dataset is preprocessed, ensuring that the gene expressions have been normalized and that there are no missing values.

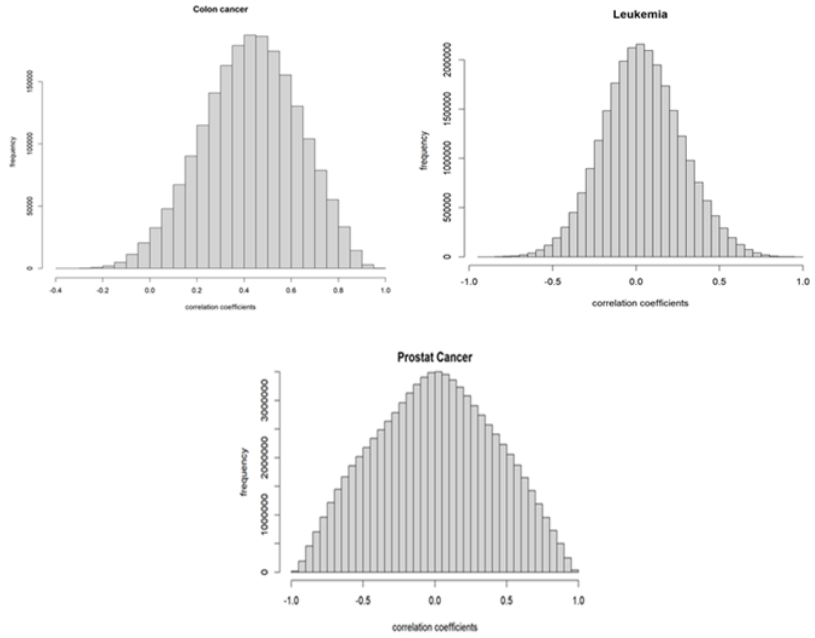


Figure 8: The histogram of pairwise correlation coefficients of a real dataset

The histogram in Figure 8 provides a summary of the pairwise correlation. The average correlation coefficient among genes in colon cancer is 0.44. The evidence suggests a strong link between genes. The correlation between genes (variables) in cases of leukemia and prostate cancer is also multi-correlated and complex due to a large number of variables. This is attributed to the use of complete gene expression data. From this, we conclude that the high-dimensional data of the genetic matrix have high correlations between them.

P-ISSN 2518-5764
E-ISSN 2227-703X

Table 3: Average Performance Comparison for Classification Real datasets with 100 iterations

Dataset	Methods	No. of Rules	Average training dataset time	Average training dataset accuracy	Average testing dataset accuracy	Average training dataset MSE	Average testing dataset MSE
Colon Cancer	LS	*4*	10.6473	0.6744	0.6737	0.3848	0.4496
	PRP	*4*	**10.4792**	0.6748	0.6740	0.3844	0.4493
	PNCG	*4*	10.5241	**0.7628**	**0.7037**	**0.3394**	**0.4367**
	LS	*6*	16.0679	0.7860	0.7579	0.3012	0.3875
	PRP	*6*	16.1039	0.7870	0.7474	0.3013	0.3871
	PNCG	*6*	**15.9452**	**0.8572**	**0.8012**	**0.2496**	**0.3382**
	LS	*10*	23.1545	0.8279	0.7579	0.2500	0.3544
	PRP	*10*	**22.8999**	0.8290	0.7585	0.2503	0.3548
	PNCG	*10*	23.1794	**0.8512**	**0.7895**	**0.2207**	**0.3415**
Leukemia	LS	*4*	40.2110	0.8720	0.6909	0.1813	0.3812
	PRP	*4*	**40.1737**	0.8740	0.7010	0.1809	0.3809
	PNCG	*4*	40.6694	**0.9240**	**0.7591**	**0.1465**	**0.3547**
	LS	*6*	62.9072	0.9600	0.8818	0.1298	0.2302
	PRP	*6*	64.6851	0.9610	0.8825	0.1295	0.2299
	PNCG	*6*	**64.0716**	**0.9760**	**0.8918**	**0.1062**	**0.2227**
	LS	*10*	101.6367	0.9720	0.8545	0.0852	0.2181
	PRP	*10*	**99.9752**	0.9740	0.8565	0.0832	0.2161
	PNCG	*10*	100.7546	**0.9800**	**0.8727**	**0.0681**	**0.2087**
Prostate Cancer	LS	*4*	107.0383	0.9127	0.6645	0.2012	0.4465
	PRP	*4*	**105.6452**	0.9155	0.6650	0.2003	0.4460
	PNCG	*4*	107.4495	**0.9296**	**0.6774**	**0.1560**	**0.4342**
	LS	*6*	**149.4426**	0.9183	0.7935	0.1932	0.3350
	PRP	*6*	158.3041	0.9185	0.7938	0.1931	0.3349
	PNCG	*6*	154.6244	**0.9408**	**0.8012**	**0.1499**	**0.3286**
	LS	*10*	218.7635	0.5211	0.4581	0.4969	0.5025
	PRP	*10*	**250.5704**	0.8113	0.6968	0.2796	0.4058
	PNCG	*10*	251.7892	**0.9239**	**0.7742**	**0.1506**	**0.3323**

P-ISSN 2518-5764
E-ISSN 2227-703X

Table 3 presents the computed values of average training dataset time, average training dataset accuracy, average testing dataset accuracy, average training dataset mean squared error (MSE), and average testing dataset MSE for all the methods. In order to evaluate the efficacy of the proposed new conjugate gradient approach when used together with well-known approaches, research was conducted using real data, including three publicly available datasets on gene expression (namely, colon cancer, leukemia, and prostate cancer) and three predefined rules (4, 6, and 10).

After analyzing the real data shown in the aforementioned table, it is apparent that the classification of high-dimensional data yielded the most beneficial results when the number of rules was either 6 or 10. In this particular situation, the suggested method demonstrated a higher level of efficacy in comparison to the PRP and LS methods. In the dataset pertaining to colon cancer, the total number of rules was 6. The mean period of execution was calculated to be around 23 minutes. Additionally, it was determined that the accuracy rate for the training data was 86%, while for the test data it was 80%. The mean squared error (MSE) for the training dataset was determined to be 0.24, but for the test dataset, the estimated value was 0.33. The correlation value of 0.4 seen in the simulation results aligns with the mean correlation coefficient of 0.44 found in the colon cancer data. The results were similar because the simulation sample generated was close to the size of the real data for colon cancer.

In contrast, while using the leukemia dataset and utilizing a set of 10 rules, the mean duration of execution was roughly 100 minutes. Additionally, the average accuracy for the training and testing datasets was 98% and 87%, respectively. The mean squared errors for the training and testing datasets are 0.06 and 0.20, respectively. Furthermore, the findings obtained from the prostate cancer dataset likewise demonstrated the superiority of the suggested method when evaluated using all the performance metrics used in our study. The data pertaining to leukemia and prostate cancer exhibited noteworthy outcomes upon increasing the number of rules to 10. In contrast, the simulated data created did not provide similar results, likely owing to the substantial amount of the aforementioned data. However, it is important to note that the implementation process was very costly.

Figures 9, 10, and 11 show the mean accuracy of the training dataset and the entirety of the rules used for the real datasets, with 100 iterations.

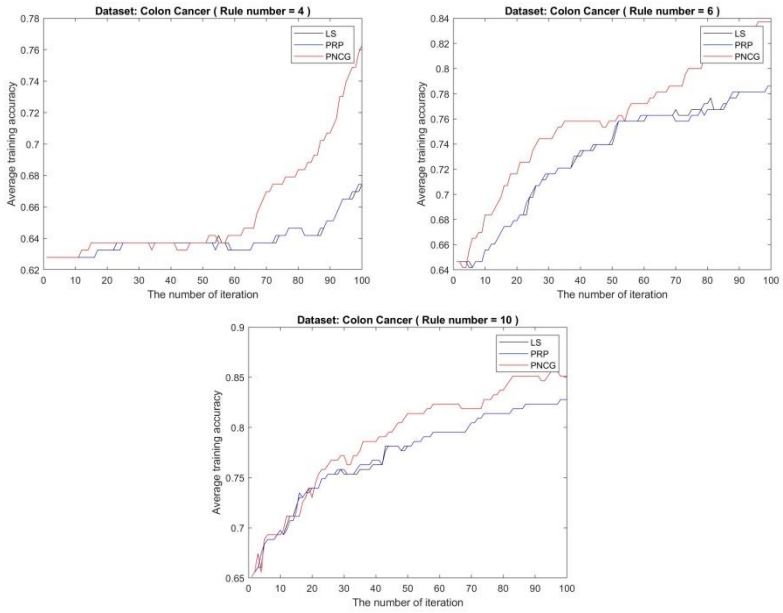


Figure 9: The average training accuracy for Colon Cancer with three rules(4,6,10)

P-ISSN 2518-5764
E-ISSN 2227-703X

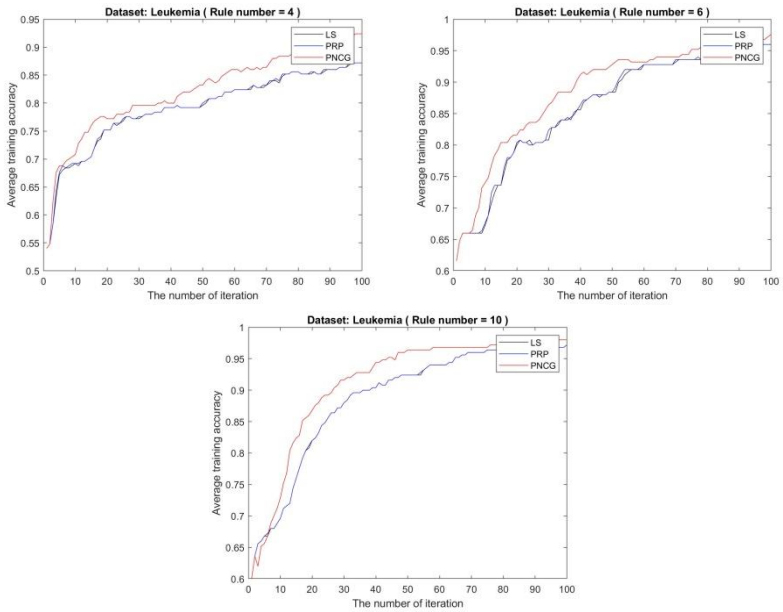


Figure 10: The average training accuracy for Leukemia with three rules(4,6,10)

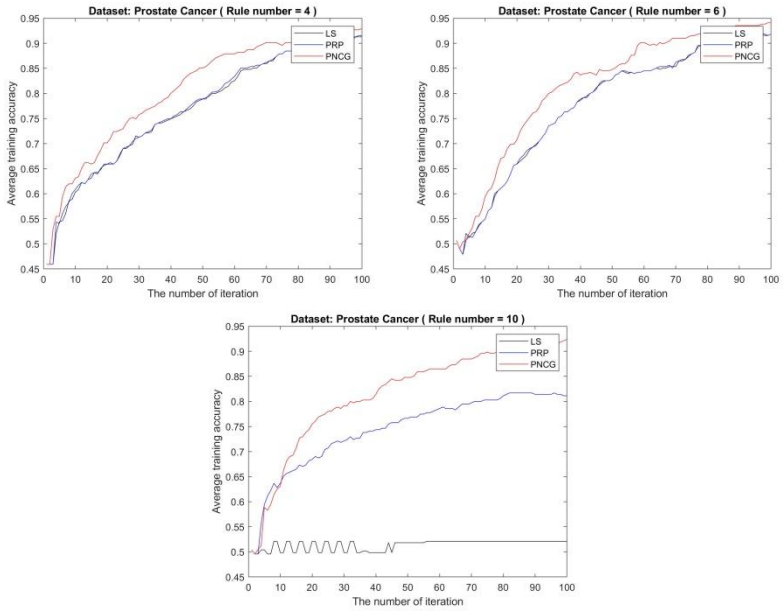


Figure 11: The average training accuracy for Prostate Cancer with three rules(4,6,10)

P-ISSN 2518-5764
E-ISSN 2227-703X

4.Conclusion:

In light of the significant importance of high-dimensional data, which type of gene expression dataset (cancer data) is associated with bioinformatics and what may result in mortality and economic crisis costs to the community, In addition, this is one of the diseases whose incidence has increased in recent years. It was necessary to study and analyse this data by employing fuzzy models based on artificial intelligence technology to classify tumor/normal conditions by using simulated and real data. The present research suggests that the conjugate gradient approach is a favorable alternative to the gradient descent method due to its superior convergence rate. This is achieved by identifying a conjugate descent route and using adaptive learning factors. This study introduces a new method to train the fuzzy neural network system of the 0-th order Takagi-Sugeno (TKS) by using a conjugate gradient method. The comparative analysis of numerical results obtained from simulations and real data sets demonstrates that the proposed method has superior generalization efficiency when compared to existing analogues. Additionally, the obtained data strongly support the effective convergence of the proposed method. Furthermore, it can be inferred that the suggested method has the capability to effectively address optimization functions and may be used in the training of artificial neural networks.

Authors Declaration:

Conflicts of Interest: None

-We Hereby Confirm That All The Figures and Tables In The Manuscript Are Mine and Ours. Besides, The Figures and Images, Which are Not Mine, Have Been Permitted Republication and Attached to The Manuscript.

- Ethical Clearance: The Research Was Approved By The Local Ethical Committee in The University.

References:

1.Abd Algafore, H. A. and Hashem, S. H. (2016), "Spam filtering based on naïve Bayesian with information gain and ant colony system", Iraqi Journal of Science, Vol. 57 No. 1c, pp.719-727.

2.Abdollahi, F. and Fatemi, M. (2021), "A new conjugate gradient method based on a modified secant condition with its applications in image processing". RAIRO-Operations Research, Vol. 55 Wthout No.**,** pp. 167-187.

3.Abdullah, Z. M., Hameed, M., Hisham, M. and Khaleel, M. A. (2019), "Modified new conjugate gradient method for Unconstrained Optimization", Tikrit Journal of Pure Science, Vol. 24 No. 5, 86-90.

4.Ahmad, H., Seadawy, A. R. and Khan, T. A. (2020), "Study on numerical solution of dispersive water wave phenomena by using a reliable modification of variational iteration algorithm", Mathematics and Computers in Simulation, Vol. 177 Wthout No., pp. 13-23.

5.Al-Baali, M. (1985), "Descent property and global convergence of the Fletcher-Reeves method with inexact line search", IMA Journal of Numerical Analysis,Vol. 5 No. 1**,** pp. 121-124.

6.Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D. and Levine, A. J. (1999), "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays", Proceedings of the National Academy of Sciences, Vol. 96 No. 12**,** 6745-6750.

7.Ashour, M., and Dahhan, I. (2022), "Curve fitting optimization for French electricity exports using recurrent neural networks", Artificial Intelligence & Future Applications. AHFE International Conference. AHFE Open Access, Vol. 68 Without No.,pp. 115-130.

8.Awad, A. J., Ahmed, A. A. and Abdulateef, O. F.(2022), "Estimate and Analysis the Availability of Generator in Electric Power Plant Using ANN" , Al-Khwarizmi Engineering Journal, Vol. 18 No. 2**,** pp. 1-12.

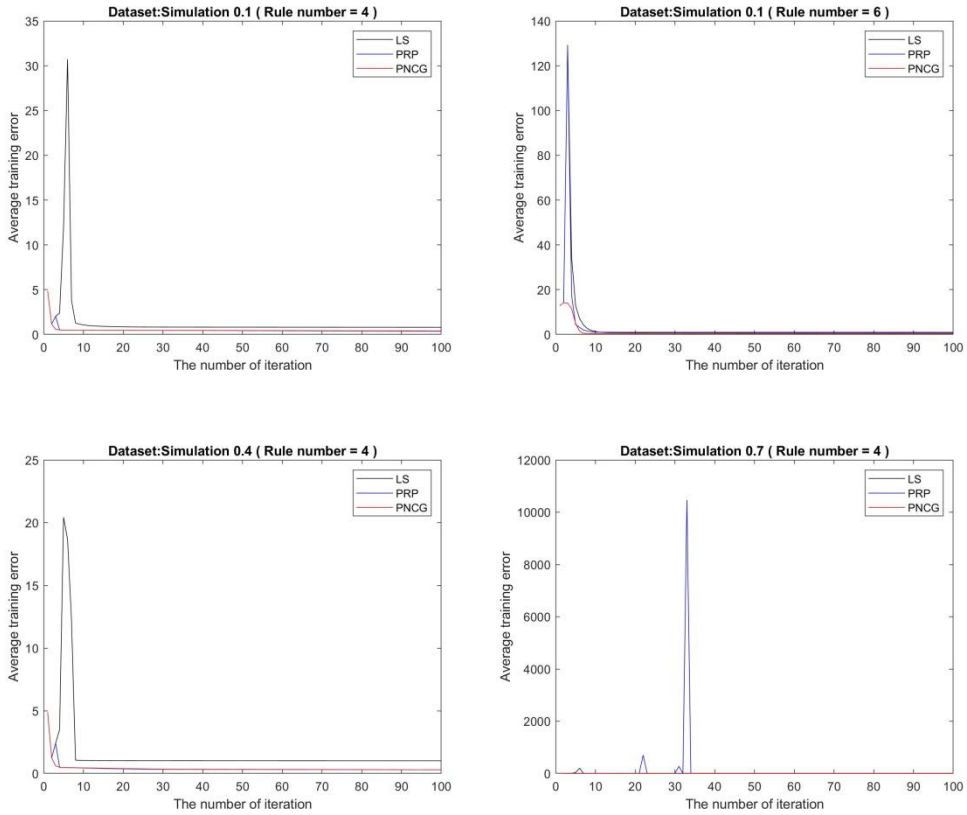
9.Bezdek, J. C., Keller, J., Krisnapuram, R. and Pal, N. (1999), "Fuzzy models and algorithms for pattern recognition and image processing", Springer Science & Business Media.
10. Bhola, A. and Singh, S. (2018), "Gene selection using high dimensional gene expression data: an appraisal", Current Bioinformatics, Vol. 13 No. 3**,** pp. 225-233.
11. Chakraborty, S., Konar, A., Ralescu, A. and Pal, N. R. (2014), "A fast algorithm to compute precise type-2 centroids for real-time control applications" , IEEE transactions on cybernetics, Vol. 45 No. 2**,** pp. 340-353.
12. Chaturvedi, K. T., Pandit, M. and Srivastava, L. (2008), "Modified neo-fuzzy neuron-based approach for economic and environmental optimal power dispatch", Applied Soft Computing, Vol. 8 No. 4**,** pp. 1428-1438
13. Da Silva, I. N., Spatti, D. H., Flauzino, R. A., Liboni, L. H. B. and Dos Reis Alves, S. F. (2017), "Artificial neural networks", Cham: Springer International Publishing.
14. Dash, R. (2020), "A two stage grading approach for feature selection and classification of microarray data using Pareto based feature ranking techniques: A case study", Journal of King Saud University-Computer and Information Sciences, Vol. 32 No. 2**,** pp. 232-247.
15. De Aguiar, E. P., Nogueira, F. M. D. A., Amaral, R. P., Fabri, D. F., Rossignoli, S. C. D. A., Ferreira, J. G., Vellasco, M. M., Tanscheit, R., Vellasco, P. C. D. S. and Ribeiro, M. V. (2016), "EANN 2014: a fuzzy logic system trained by conjugate gradient methods for fault classification in a switch machine", Neural Computing and Applications, Vol. 27 Without No.**,** pp. 1175-1189.
16. Del Campo, I., Echanobe, J., Bosque, G. and Tarela, J. M. (2008), "Efficient hardware/software implementation of an adaptive neuro-fuzzy system", IEEE Transactions on Fuzzy Systems, Vol. 16 No. 3**,** pp. 761-778.
17. Gao, T., Wang, J., Zhang, B., Zhang, H., Ren, P. and Pal, N. R. (2018), "A polak-ribière-polyak conjugate gradient-based neuro-fuzzy network and its convergence", IEEE Access, Vol. 6 Without No.**,** pp. 41551-41565.
18. Gao, T., Zhang, Z., Chang, Q., Xie, X., Ren, P. and Wang, J. (2019), "Conjugate gradient-based Takagi-Sugeno fuzzy neural network parameter identification and its convergence analysis", Neurocomputing , Vol. 364 Without No.**,** pp. 168-181.
19. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R. and Caligiuri, M. A. (1999), "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring", science, Vol. 286 No. 5439**,** pp. 531-537.
20. Gu, X. and Wang, S. (2018), "Bayesian Takagi–Sugeno–Kang fuzzy model and its joint learning of structure identification and parameter estimation", IEEE Transactions on Industrial Informatics, Vol. 14 Without No.**,** pp. 5327-5337.
21. Gu, X., Chung, F.-L., Ishibuchi, H. and Wang, S. (2016), "Imbalanced TSK fuzzy classifier by cross-class Bayesian fuzzy clustering and imbalance learning", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 47 Without No.**,** pp. 2005-2020.
22. Honrado, E., Osorio, A., Palacios, J. and Benítez, J. (2006), "Pathology and gene expression of hereditary breast tumors associated with BRCA1, BRCA2 and CHEK2 gene mutations", Oncogene, Vol. 25 No. 43**,** pp. 5837-5845.
23. Ichihashi, H. and Türksen, I. B. (1993), "A neuro-fuzzy approach to data analysis of pairwise comparisons", International Journal of Approximate Reasoning, Vol. 9 No. 3**,** pp. 227-248.
24. Jang, J.-S. (1993), "ANFIS: adaptive-network-based fuzzy inference system", IEEE transactions on systems, man, and cybernetics, Vol. 23 No. 3**,** pp. 665-685.
25. Juang, C.-F. and Chen, J.-S. (2006), "Water bath temperature control by a recurrent fuzzy controller and its FPGA implementation", IEEE Transactions on Industrial Electronics, Vol. 53 No. 3**,** pp. 941-949.
26. Juang, C.-F. and Lin, C.-T. (1998), "An online self-constructing neural fuzzy inference network and its applications", IEEE transactions on Fuzzy Systems, Vol. 6 No. 1**,** pp. 12-32.

27. Khalil, H. H. and Rada, S. M. (2023), "An Artificial Intelligence Algorithm to Optimize the Classification of the Hepatitis Type", Journal of Economics and Administrative Sciences, Vol. 29 No. 135**,** pp. 43-55.
28. Khudhur, H. M. and Abbo, K. (2020), "A new conjugate gradient method for learning fuzzy neural networks", J. Multidiscip. Model. Optim, Vol. 3 No. 2**,** pp. 57-69.
29. Kim, S. and Kim, J.-M. (2019), "Two-stage classification with sis using a new filter ranking method in high throughput data", Mathematics, Vol. 7 No. 6**,** pp. 493.
30. Klir, G. J., St. Clair, U. and Yuan, B. (1997), "Fuzzy set theory: foundations and applications", Prentice-Hall, Inc.
31. Laylani, Y. A. and Khudhur, H. M. (2018), "Training feed forward neural network with modified Fletcher-Reeves method", Journal of Multidisciplinary Modeling and Optimization, Vol. 1 No. 1**,** pp. 14-22.
32. Liu, Y. and Storey, C. (1991), "Efficient generalized conjugate gradient algorithms, part 1: theory", Journal of optimization theory and applications, Vol. 69**,** pp. 129-137.
33. Luo, X., Liu, D. and Wan, B. (1998), "An adaptive fuzzy neural inferring network", Fuzzy Systems and Mathematics, Vol. 12 No. 4**,** pp. 26-33.
34. Mendel, J. M. (2013), "General type-2 fuzzy logic systems made simple: a tutorial", IEEE Transactions on Fuzzy Systems, Vol. 22 No. 5**,** pp. 1162-1182.
35. Noor Alhuda Khalid, H. and Basad, A.-S. (2022), "Deep Learning and Machine Learning via a Genetic Algorithm to Classify Breast Cancer DNA Data", Iraqi Journal of Science, Vol. 63 No. 7**,** pp. 3153-3168.
36. Pi, L. and Halabi, S. (2018), "Combined performance of screening and variable selection methods in ultra-high dimensional data in predicting time-to-event outcomes", Diagnostic and prognostic research, Vol. 2 No. 1**,** pp. 1-12.
37. Polak, E. and Ribiere, G. (1969), "Note on the convergence of methods of combined directions", french journal of computer science and operational research. Red Series, Vol.3 No. 16, pp. 35-43.
38. Sahiner, A. and Ibrahem, S. A. (2019), "A new global optimization technique by auxiliary function method in a directional search", Optimization Letters, Vol. 13**,** pp. 309-323.
39. Sahiner, A., Yilmaz, N. and Ibrahem, S. A. (2018), "Smoothing approximations to non-smooth functions", Journal of Multidisciplinary Modeling and Optimization, Vol. 1 No. 2**,** pp. 69-74.
40. Salleh, Z., Alhamzi, G., Masmali, I. and Alhawarat, A. (2022), "A modified liu and storey conjugate gradient method for large scale unconstrained optimization problems", Algorithms, Vol. 14 No. 8**,** pp. 227.
41. Shahparast, H., Mansoori, E. G. and Zolghadri Jahromi, M. (2019), "AFCGD: an adaptive fuzzy classifier based on gradient descent", Soft computing, Vol. 23 Without No.**,** pp. 4557-4571.
42. Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D'amico, A. V. and Richie, J. P. (2002), "Gene expression correlates of clinical prostate cancer behavior", Cancer cell, Vol. 1 No. 2**,** pp. 203-209.
43. Sivanandam, S., Sumathi, S. and Deepa, S. (2007), "Introduction to fuzzy logic using MATLAB", Springer.
44. Sugeno, M. and Kang, G. (1988), "Structure identification of fuzzy model", Fuzzy sets and systems, Vol. 28 No. 1**,** pp. 15-33.
45. Tang, M., Jun Wang, K. and Zhang, Y.(2007), "A research on chaotic recurrent fuzzy neural network and its convergence", International Conference on Mechatronics and Automation, IEEE, pp. 682-687.
46. Taniguchi, T., Tanaka, K., Ohtake, H. and Wang, H. O. (2001), "Model construction, rule reduction, and robust compensation for generalized form of Takagi-Sugeno fuzzy systems", IEEE Transactions on Fuzzy Systems, Vol. 9 No. 4**,** pp. 525-538.

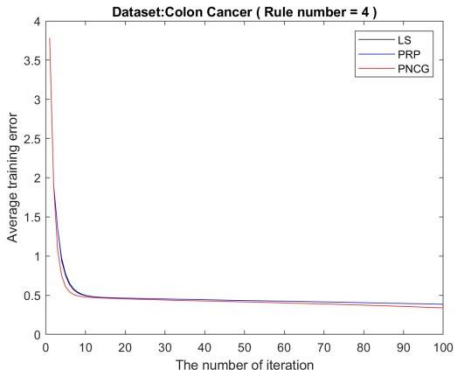
47. Wu, W., Li, L., Yang, J. and Liu, Y. (2010), "A modified gradient-based neuro-fuzzy learning algorithm and its convergence", Information Sciences, Vol. 180 No. 9**,** pp. 1630-1642.
48. Wu-Catherine, H. and Mclarty, J. (2001), "Neural networks and genome informatics", Methods in computational biology and biochemistry, New York: Elsevier**,** pp.1-205.
49. Yager, R. R. and Filev, D. P. (1994), "Generation of fuzzy rules by mountain clustering", Journal of Intelligent & Fuzzy Systems, Vol. 2 No. 3**,** pp. 209-219.
50. Zadeh, L. A. (1965), "Fuzzy sets", Information and control, Vol. 8 No. 3**,** pp. 338-353.
51. Zhang, L. and Zhou, W. (2008), "Two descent hybrid conjugate gradient methods for optimization" , Journal of Computational and Applied Mathematics, Vol. 216 No. 1**,** pp. 251-264.

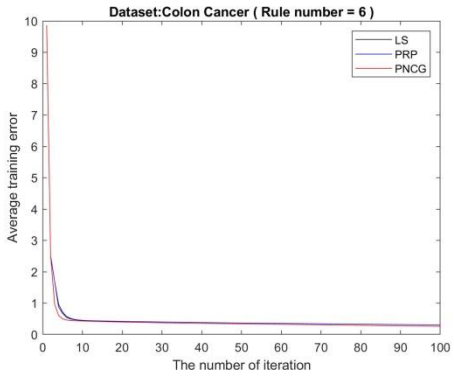
Appendix

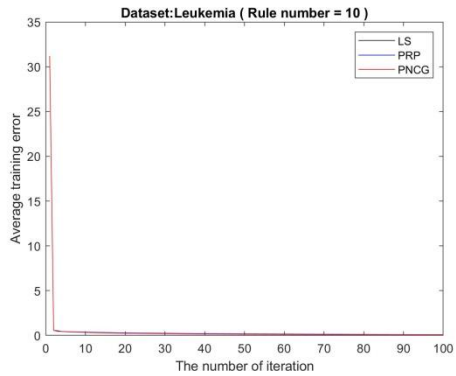
Some of the plotting average training error (MSE) for simulation and real dataset with a different .number of rules

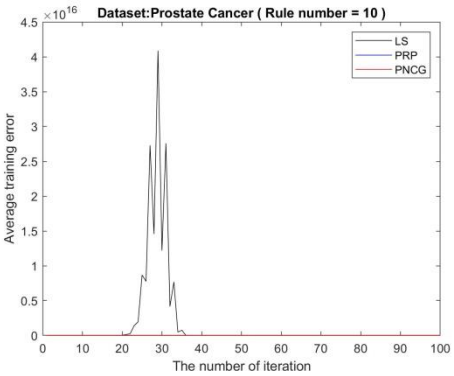


P-ISSN 2518-5764
E-ISSN 2227-703X









Journal of Economics and Administrative Sciences
2024; 30(142), pp. 425-448

P-ISSN 2518-5764
E-ISSN 2227-703X

النماذج الضبابية مع طريقة التدرج المترافق الجديدة المقترحة لتصنيف البيانات عالية الأبعاد في المعلوماتية الحيوية

عمر قصي الشبلي[1]
جامعة بغداد / كلية الادارة والاقتصاد/ قسم الاحصاء بغداد، العراق.
omarqusay@uomosul.edu.iq

سهيل نجم عبدالله[2]
جامعة بغداد / كلية الادارة والاقتصاد/ قسم الاحصاء بغداد، العراق.
dr.suhail.najm@coadec.uobaghdad.edu.iq

Received: 18/10/2023 Accepted:24/12/2023 Published Online First: 30 /8/ 2024

مستخلص البحث:

يمكن أن يعزى تطور موضوع المعلوماتية الحيوية إلى النمو الأسي للبيانات البيولوجية ، أي الكمية الهائلة من بيانات التعبير الجيني عالية الأبعاد. يتعامل تخصص المعلوماتية الحيوية بكفاءة مع المسائل في البيولوجيا الجزيئية من خلال استخدام الأمثلية العددية وعلوم الحاسوب والطرائق الإحصائية. اقترح هذا البحث طريقة التدرج المترافقة الجديدة (PNCG) كطريقة لمسألة الامثلية لتعلم نموذج الشبكة العصبية المضببة على أساس نهج تاكاجي-سوجينو. قدمت هذه الدراسة خوارزمية جديدة تعالج مسألة التقارب المتأخر الذي شوهد في تقنيات بولاك ريبيري بولياك (PRP) و ليو ستوري (LS) باستخدام طريقة PRP. استخدم في البحث مجموعات بيانات محاكاة وحقيقية لتقييم الطريقة المقترحة تجريبيا. أظهرت نتائج الدراسة أن الطريقة المقترحة أظهرت أداء فائقا مقارنة بالطرق الأخرى الموجودة. تضمنت النتائج التجريبية استخدام ثلاث مجموعات بيانات حقيقية تتعلق بالسرطان وهي احد انواع البيانات عالية الأبعاد. حيث اشارت النتائج إلى أن النهج او الطريقة المقترحة فعالة للغاية وممكن تطبيقها ، وبالتالي اظهرت درجة كبيرة من الفاعلية من حيث معدل الوقت لمجموعة بيانات التدريب ، ومعدل الدقة لمجموعة بيانات التدريب ، ومعدل الدقة لمجموعة بيانات الاختبار ، ومعدل متوسط مربعات الخطأ (MSE) لبيانات التدريب ، وكذلك معدل متوسط مربعات الخطأ (MSE) لبيانات الاختبار.

نوع البحث: ورقة بحثية.

المصطلحات الرئيسة للبحث: المعلوماتية الحيوية، الأمثلية ، الشبكات العصبية المضببة ، الدقة ، التدرج المترافق ،بيانات التعبير الجيني.

[1] طالب دكتوراه ,جامعة بغداد ,كلية الادارة والاقتصاد ,قسم الاحصاء ,بغداد ,العراق.
[2] بروفيسور ,جامعة بغداد ,كلية الادارة والاقتصاد ,قسم الاحصاء ,بغداد ,العراق.