https://doi.org/10.31272/jae.i153.1610
https://admics.uomustansiriyah.edu.iq
P-ISSN: 1813–6729 E-ISSN: 2707-1359

JAE
OPEN ACCESS

Variable Selection in High-Dimensional Solar Radiation Data Using a Hybrid Approach Based on LASSO, Elastic Net, and Fuzzy Logic

Israa Khamees Ahmed [1], Osamah Basheer Shukur [2]
[1, 2]Dept. of Statistics and Informatics, College of Computer Science & Mathematics, University of Mosul, Mosul, Iraq.

Article Information

Article History:
Received: 10 / 05 / 2026
Revised: 06 / 08 / 2026
Accepted: 20 / 08 / 2026
Available Online: 01 / 09 / 2026
Pages no : 138 – 151

Keywords:
Elastic Net (EN), LASSO, Fuzzy-LASSO, Fuzzy-EN

Correspondence:
Researcher name:
Osamah Basheer Shukur

Email:
drosamahannon@uomosul.edu.iq

Abstract

Methods for selecting and predicting variables in high-dimensional time series data face numerous challenges due to the sheer number of variables. This makes variable selection difficult and reduces prediction accuracy. Climate variables, for example, are characterised by complexity, nonlinearity, and high correlations among variables, which limit the effectiveness of traditional models in handling high-dimensional and uncertain data. This study aims to build a hybrid predictive framework that combines conventional regularisation methods. LASSO regression and Elastic Net (EN) regression are methods used to reduce the number of explanatory variables for high-dimensional data, which will be used in this study to predict and reduce the number of explanatory climate variables, and between fuzzy logic, by building hybrid LASSO-Fuzzy and EN-Fuzzy models to improve the accuracy of prediction and the efficiency of variable selection by solving the problem of high dimensions, uncertainty and non-linearity, which are among the problems that the data are affected by, High-dimensional solar radiation time series data will be used for the application using the proposed methods. The importance of the hybrid approach lies in reducing the number of variables in high-dimensional solar radiation data, thereby reducing prediction error and improving the model's interpretability. The results showed that hybrid methods outperformed conventional models such as LASSO and EN. Hybrid models have achieved the lowest predictive error rates and more stable performance in selecting and minimising influential variables in high-dimensional data. It can be concluded that the LASSO-Fuzzy and EN-Fuzzy hybrid models improve prediction accuracy for high-dimensional data by effectively reducing dimensionality and selecting the most influential variables.

1. Introduction

Selecting important variables is a fundamental issue in data analysis, especially for high-dimensional time series data, where the number of variables exceeds the number of observations, such as solar radiation data. This leads to statistical problems such as multicollinearity, high variance, and low predictive power of conventional models. To address these problems, several variable selection methods have been proposed, such as the Least Absolute Shrinkage and Selection Operator (Lasso) method, proposed by Tibshirani [1]. It is an important method for variable selection, as it adds an $\ell 1$ penalty to the sum of squared residuals (SSR), helping simplify the model. The Elastic Net method for selecting important variables, proposed by[2], combines two penalty terms: the first penalty term ($\ell 1$) found in Lasso and the second penalty term ($\ell 2$) found in Ridge regression, to take advantage of the benefits of both terms. Elastic Net has been used to predict house prices in the United States [3]. However, these methods may face challenges in dealing with high-dimensional data because they suffer from uncertainty and high correlation between variables. Therefore, a hybrid framework combining conventional models and Fuzzy Logic [4] was proposed to represent data more flexibly through membership functions. Fuzzy Logic was extended to include regression models when Tanaka [5]introduced fuzzy regression, which represents conventional regression coefficients as fuzzy numbers. To build a Fuzzy Inference System (FIS), the Subtractive Clustering method [6] was used to fuzzify the variables selected by conventional models and generate membership functions. Although fuzzy regularisation techniques such as fuzzy LASSO and fuzzy EN have been reported in previous studies, the novelty of the present work does not lie in introducing these techniques themselves. Instead, this study proposes an integrated hybrid framework that combines conventional regularisation methods (LASSO and EN) with a fuzzy inference system generated using subtractive clustering for variable selection and prediction in high-dimensional solar radiation time-series data. Furthermore, the proposed framework provides a comprehensive comparative evaluation of conventional and hybrid models, demonstrating the approach's effectiveness in improving prediction accuracy while reducing the dimensionality of complex environmental data. The importance of this study lies in the fact that high-dimensional climatic datasets have been described as complex data with many explanatory variables and often exhibit strong correlations. Due to these problems, conventional regression techniques may suffer from multicollinearity and low predictive accuracy. To address these problems, conventional LASSO and EN methods have been integrated by hybridising them with fuzzy logic. Although these methods are effective, they rely on the assumption that the data are accurate and clear, which is unrealistic because many practical applications, such as climatic datasets, are characterised by uncertainty, ambiguity, and inaccuracy. This type of data cannot be effectively handled using conventional methods because these methods do not account for the ambiguity and uncertainty inherent in such data, which may reduce prediction accuracy.

2. The research aims

This study aims to develop a hybrid approach combining conventional regularisation methods and fuzzy logic. It seeks to achieve several objectives:

1- To reduce and select the important variables based on LASSO and Elastic Net.

2- To improve the predictive accuracy of the conventional regularisation methods by integrating them with fuzzy logic to create hybrid approaches like Fuzzy-LASSO and Fuzzy-EN, which can handle uncertainty and ambiguity in high-dimensional data.

3- To compare the predictive accuracy of conventional regularisation methods with the proposed hybrid approaches.

3. LASSO method

The LASSO (Least Absolute Shrinkage and Selection Operator) method is an important regularisation method used in estimating the coefficients of linear regression models, especially when dealing with high-dimensional data where the number of explanatory variables exceeds the number of observations. This method was proposed by [1]. LASSO reduces the shrinkage of certain coefficients to zero to improve prediction accuracy. The Lasso estimator is obtained using the following equation:

$$\hat{\beta}_{LASSO} = arg \min_{\beta_0,\beta} \left(\frac{1}{2N} \sum_{i=1}^{N} \left(y_i - \beta_0 - x_i^T \beta\right)^2 + \lambda \sum_{j=1}^{K} \left|\beta_j\right| \right) \quad (1)$$

The equation above represents the objective function for LASSO, where the first term is the loss function, the sum of squared residuals (SSR). The second part of the equation, lambda sum from j equals 1 to cap K of , absolute value beta sub j,, end absolute value, represents the ℓ1-norm (penalty term), which imposes a penalty on the βj model so that some coefficients are exactly equal to zero. The values of β_0 and β are constants. The purpose of this is to select important variables and reduce the model's complexity. λ represents the tuning parameter, which controls the balance between the SSR and ℓ1-norm parts, and its value is ($\lambda \geq 0$). The tuning parameter (λ) was automatically selected using 10-fold cross-validation in MATLAB.LASSO is a powerful method for selecting important variables. It is used to reduce dimensionality by excluding zero and unimportant variables from the model, retaining only those with a meaningful effect. The mechanism of LASSO is to shrink the regression coefficients of predictive variables highly correlated with the error term to zero or near zero [7] because the ℓ1 penalty can set some coefficients exactly to zero. The set of variables selected is known as the effective set, which helps simplify the model and improve its interpretability.

4. Elastic Net

The Elastic Net is a method of regularised linear regression that combines Lasso and Ridge regression. [8]. This method was invented by the scientists Zou and Hastie. [2].It is used to overcome the problem of multiple linear correlation [7] by selecting the important variables. EN can overcome the drawbacks of Lasso regression and ridge regression because they combine the advantages of each. They also overcome the problem of overfitting. [9] EN formula:

$$\beta_{\text{Elastic}}\,(\lambda_1,\lambda_2) = \ell(\beta) + \lambda_1 \sum_{j=1}^{p} |\beta_j| + \lambda_2 \sum_{j=1}^{p} \beta_j^2 \tag{2}$$

Lasso penalty limitℓ1-norm:

$$\lambda_1 \sum_{j=1}^{p} |\beta_j| \tag{3}$$

This limit shrinks the number of coefficients to zero, allowing the model to perform variable selection and yield an interpretable model. Ridge penalty limit ℓ2-norm :

$$\lambda_2 \sum_{j=1}^{p} \beta_j^2 \tag{4}$$

This limit works by approximating the coefficients without zeroing them out, which helps address multicollinearity and ensures model stability and proper weight allocation across correlated variables. The EN is characterised by greater flexibility in controlling the mixing parameter (α) through an alternative formulation.

$$\hat{\beta}_{ENET} = arg\,\min_{\beta}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(y_i - \beta_0 - x_i^T\beta\right)^2 + \lambda P_\alpha(\beta)\right) \tag{5}$$

When the value of alpha $\alpha = \frac{\lambda_1}{\lambda_1+\lambda_2}, 0 \leq \alpha \leq 1$

Value $P_\alpha(\beta)$

$$\mathrm{P}_\alpha(\beta) = \frac{(1-\alpha)}{2}\|\beta\|_2^2 + \alpha\|\beta\|_1 = \sum_{j=1}^{p}\left(\frac{(1-\alpha)}{2}\beta_j^2 + \alpha|\beta_j|\right) \tag{6}$$

When $\alpha = 1$, the EN will be equivalent to the ℓ1-norm Lasso regression, and when $\alpha = 0$, the EN will be equivalent to the ℓ2-norm linear regression. In this study, the mixing parameter was fixed at $\alpha = 0.5$ throughout all experiments to provide an equal balance between the ℓ1-norm and ℓ2-norm penalties, while λ was automatically selected using 10-fold cross-validation.

5. Fuzzy Regularisation Techniques

Dealing with uncertainty and nonlinearity is important in various technological fields. Dealing with systems of uncertainty reveals unpredictable behaviour due to several factors, such as input variables or incomplete knowledge of the system's dynamics. Nonlinear systems are characterised by input-output relationships that do not follow a linear pattern, as in high-dimensional time series, making them difficult to handle and model using conventional linear methods. [10]. Therefore, the fuzzy logic presented by [4] was proposed to represent uncertainty and ambiguity through membership functions. Fuzzy logic was extended to incorporate statistical models, leading to the development of fuzzy linear regression models, which assume that the coefficients are fuzzy numbers rather than constants. Conventional models, such as Lasso and EN, assume that the data are accurate and that the coefficients are constant, which is unrealistic because many practical applications, such as high-dimensional time series data, are characterised by uncertainty and ambiguity.

Therefore, hybrid models that combine fuzzy logic with conventional regularisation methods were developed to improve prediction accuracy. The foundation of fuzzy logic lies in fuzzy reasoning systems. These systems operate on a rule-based model, transforming inputs into logical outputs. The aim is to simplify the handling of unpredictable systems.

5.1 Fuzzy LASSO

The conventional LASSO of the multi-regression model has been extended to a fuzzy environment (Hesamian & Akbari, 2019) for coefficient estimation, where the explanatory variables in the multi-regression model are not fuzzy, and the coefficients and the response variable are fuzzy. [11].

$$\tilde{\tilde{\boldsymbol{\beta}}} = \arg\min_{\tilde{\boldsymbol{\beta}}} \sum_{i=1}^{n} D^2(\tilde{y}_i, \tilde{y}_i^*) \tag{7}$$

The Lasso method was extended to a fuzzy environment to estimate the coefficients of the multiple regression model, through several steps outlined below:

1- Modelling input and objective variables using the LASSO regression model.
2- Obtain a reduced number of actually influential and significant explanatory variables along with their coefficient values. Also, obtain the output variable $\hat{Y}_{LASSO}$.
3- Weight the final variables in LASSO through the values of their significant coefficient and specify them as inputs for the Fuzzy Inference System (FIS).
4- Construct the fuzzy inference system (FIS) using the MATLAB R2021a Fuzzy Logic Toolbox. A first-order Sugeno FIS was generated using the Subtractive Clustering algorithm. Gaussian membership functions (gaussmf) were adopted for all input variables, and the weighted average (wtaver) method was employed for defuzzification. The cluster influence range was tuned by evaluating three values (0.3, 0.5, and 0.7), and the optimal model was selected according to the minimum training RMSE. Finally, the generated FIS was used to obtain the final predicted outputs$\hat{Y}_{FIS}$.

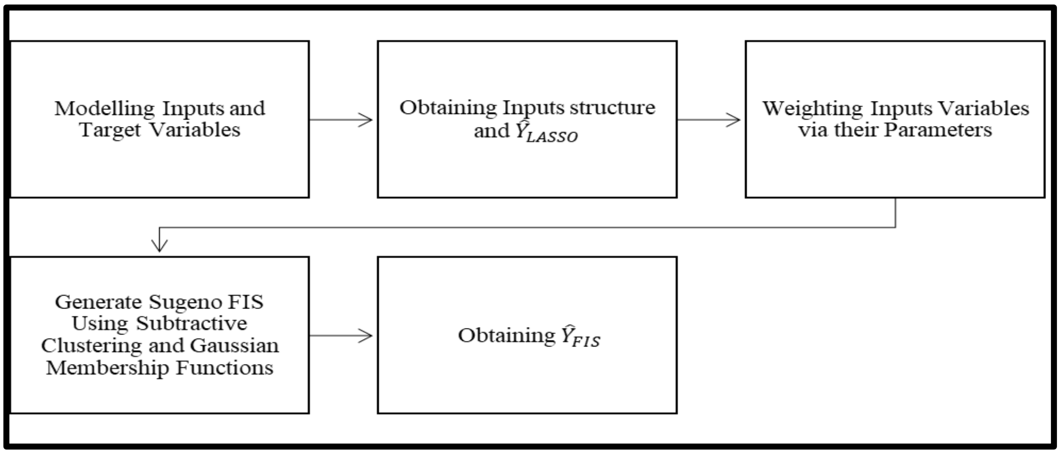

Figure (1): illustration of Fuzzy-LASSO framework

5.2 Fuzzy -EN

Variable selection is a fundamental technique in machine learning and data analysis. It plays a crucial role in selecting important variables from high-dimensional datasets. To accurately select variables and improve model prediction accuracy[12]In an EN, a combination of Lasso and ridge regression was used. However, conventional regularisation techniques can face challenges when dealing with data characterised by uncertainty and ambiguity. Therefore, a hybrid model was developed that integrates the EN with fuzzy logic to represent uncertainty and ambiguity in selecting important variables, aiming to improve prediction accuracy. In an EN, variables are fuzzy by following several steps:

1- Modelling input and objective variables using the EN regression model.
2- Obtain a reduced number of actually influential and significant explanatory variables along with their coefficient values. Also, obtain the output variable $\hat{Y}_{EN}$.
3- Weight the final variables in EN through the values of their significant coefficient and specify them as inputs for the Fuzzy Inference System (FIS).
4- Construct the fuzzy inference system (FIS) using the MATLAB R2021a Fuzzy Logic Toolbox. A first-order Sugeno FIS was generated using the Subtractive Clustering algorithm. Gaussian membership functions (gaussmf) were adopted for all input variables, and the weighted average (wtaver) method was employed for defuzzification. The cluster influence range was tuned by evaluating three values (0.3, 0.5, and 0.7), and the optimal model was selected according to the minimum training RMSE. Finally, the generated FIS was used to obtain the final predicted outputs$\hat{Y}_{FIS}$.

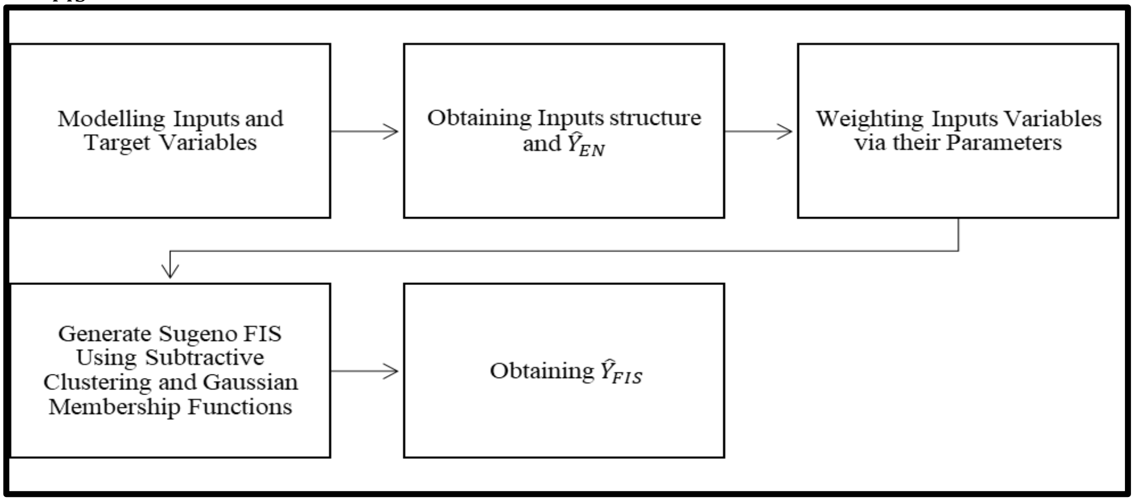

Figure (2): Illustration of Fuzzy-EN framework

6. Data used in the study

The study relied on daily solar radiation data for the city of Mosul in Nineveh Governorate, obtained from the Nineveh Governorate Meteorological Centre. The data covers the period from 1 January 1994 to 31 December 2024. This constitutes a daily time series comprising 11,323 observations. The data was divided into two sets: a training set and a test set, comprising approximately 80% training data and 20% test data, with the split taking into account the start and end dates of each month. The training set runs from January 1, 1994, to October 31, 2018, comprising 9,070 observations, while the test set runs from November 1, 2018, to December 31, 2024, comprising 2,253 observations. The predictor matrix consisted exclusively of lagged solar radiation observations generated from the original daily time series. To investigate the effect of dimensionality, several lag structures were considered (p = 20, 50, 100, 250, 500, and 1000), where each value of p represents the number of predictor variables (lagged observations) included in the model. No additional meteorological variables, such as temperature, humidity, rainfall, or wind speed, were incorporated in this study. Furthermore, the dataset contained no missing observations. Before model estimation, the predictor variables were standardised using the default settings of MATLAB's lasso function to ensure that all predictor variables were evaluated on a common scale during the regularisation process.

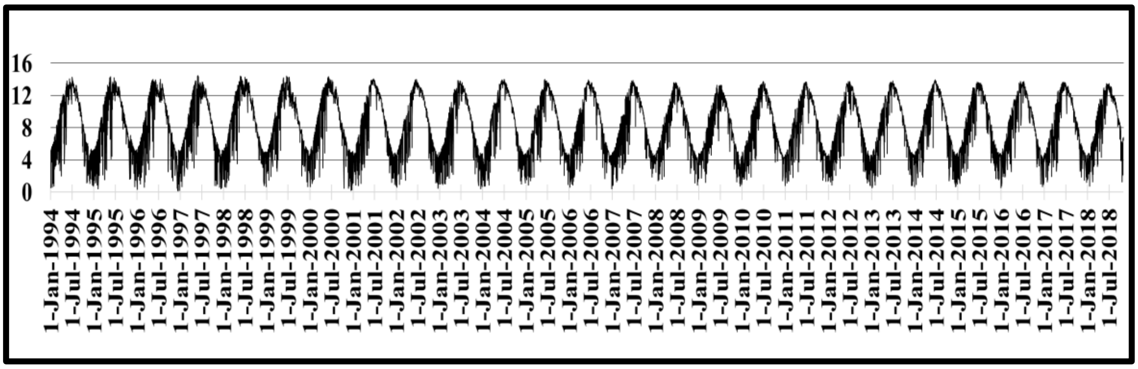

Figure (3): Time series of the actual values of the variable Y

This section presents the results from the models used in the study. Conventional regularisation methods, LASSO and EN, were used to select variables that influenced the results, and these were compared with the proposed fuzzy methods. Subtractive clustering was used to generate a fuzzy inference system (FIS), and the system's performance was compared using the RMSE error criterion. The following tables illustrate the evaluation of the conventional regularisation methods for both LASSO and EN and the proposed Fuzzy-LASSO and Fuzzy-EN methods at different values of the number of original variables P.

7. Results of LASSO method

The tables show the number of variables selected using conventional regularisation methods and the RMSE criterion for both training and test data.

Table (1) The variables selected by LASSO and their corresponding RMSE values.

Model	Number of original variables (p)	Number of selected variables (h)	RMSE (Training)	RMSE (Test)	MAE (Train)	MAE (Test)	R^2 (Train)	R^2 (Test)	MAPE Train	MAPE Test
LASSO	20	14	1.4474	1.4668	1.0080	0.9769	0.8379	0.8587	25.24%	23.83%
LASSO	50	14	1.4374	1.4561	0.9956	0.9637	0.8401	0.8611	24.78%	23.36%
LASSO	100	18	1.4139	1.4327	0.9722	0.9440	0.8453	0.8658	24.38%	22.97%
LASSO	250	18	1.4067	1.4254	0.9744	0.9469	0.8469	0.8676	24.57%	23.15%
LASSO	500	20	1.3931	1.4116	0.9651	0.9353	0.8498	0.8698	24.35%	22.97%
LASSO	1000	19	1.3943	1.4130	0.9664	0.9377	0.8495	0.8696	24.43%	23.02%

From the results in the table, we observe that the RMSE for the training data gradually decreases as the number of original variables (P) increases. This indicates an improvement in the model's ability to represent the data as the number of variables increases. We also note that the model selected a small number of variables compared to the original number, reflecting LASSO's ability to perform shrinkage while simultaneously selecting important variables. The additional evaluation metrics further support these findings. As the number of original variables increased, the MAE and MAPE values generally decreased, while the R^2 values increased for both the training and testing datasets. The consistency of these evaluation metrics confirms improved predictive performance of the LASSO model.

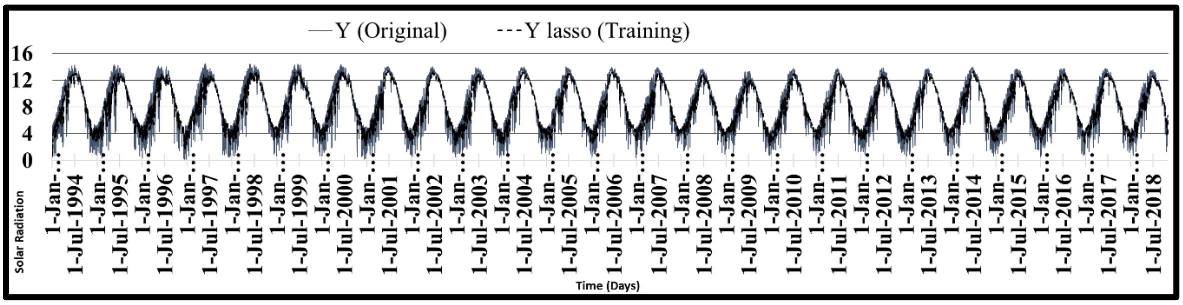

Figure (4): Comparison between observed and predicted solar radiation values using the LASSO model for the training dataset

Figure 4 shows that when P=20, the LASSO model tracks the actual time-series values well, but it may show some deviations at the peaks, indicating that the model tends to simplify due to its shrinkage property.

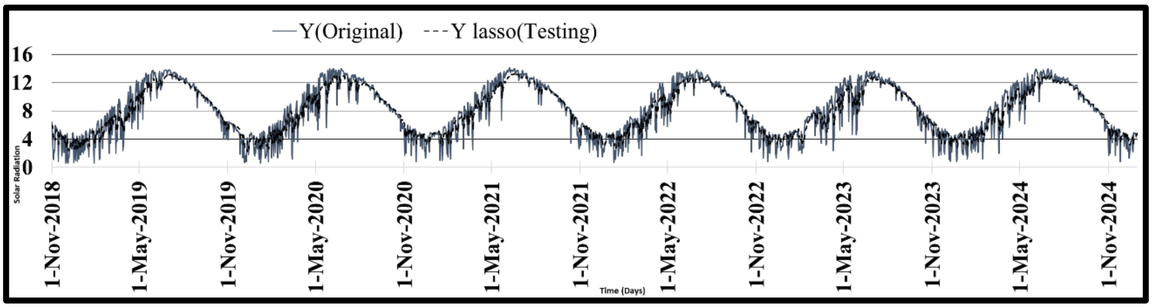

Figure (5): Comparison between observed and predicted solar radiation values using the LASSO model for the (Test dataset)

8. Results of Fuzzy-LASSO

After the variables were selected using the conventional Lasso method, the selection transitioned to fuzzy variables to build a fuzzy inference system using the Subtractive Clustering algorithm. Several values of the cluster influence range **r** were tested, and the optimal value was chosen based on the lowest RMSE value of the training data. The following steps were followed in the fuzzing process:

1- initialisation of the data for the explanatory variables, which are time lags [13], with different numbers.

$$X = [X_{t-1} \quad X_{t-2} \quad \dots\dots \quad X_{t-n}] \tag{8}$$

Where $X_{t-1} \quad X_{t-2} \quad \dots\dots \quad X_{t-n}$ represent lagged observations of the solar radiation variable, which serve as the explanatory variables throughout this study

2- Applying LASSO regression to the inputs (1) to obtain a reduced number of actually influential and significant variables with their coefficient values.

3- Obtain the output variable $\hat{Y}_{LASSO}$ and calculate Error criteria such as root mean square error (RMSE).

4- Weight the final variables in LASSO through the values of their significant coefficient and specify them as inputs for the Fuzzy Inference System (FIS).

5- Determining the Membership Function for input and output variables, in addition to determining one of the Fuzzy Inference System (FIS) algorithms, and the Subtractive Clustering algorithm was used.

6- Applying the Fuzzy Inference System (FIS) and obtaining the final outputs $\hat{Y}_{FIS}$ and calculating one of the error criteria, RMSE, to compare the outputs between LASSO before and after fuzzing.

Table (2) below shows the Fuzzy -LASSO results, RMSE values for both training and test data, as well as the membership functions, rules, and best r

Table (2) The RMSE values, membership functions, and the number of rules for the Fuzzy-LASSO model.

h	Best r (Training)	MF (Training)	Rules (Training)	RMSE (Training)	RMSE (Testing)	MAE (Train)	MAE (Test)	R^2 (Train)	R^2 (Test)	MAPE Train	MAPE Test
14	0.3	7	7	1.4169	1.4365	0.9772	0.9540	0.8446	0.8621	24.04%	22.73%
14	0.3	7	7	1.4169	1.4365	0.9772	0.9540	0.8446	0.8621	24.04%	22.73%
18	0.5	5	5	1.4173	1.4368	0.9768	0.9563	0.8445	0.8623	24.00%	22.78%
18	0.5	5	5	1.4173	1.4368	0.9768	0.9563	0.8445	0.8623	24.00%	22.78%
20	0.5	6	6	1.3921	1.4107	0.9598	0.9343	0.8500	0.8673	23.84%	22.59%
19	0.7	5	5	1.4033	1.4224	0.9695	0.9442	0.8476	0.8655	23.89%	22.66%

It can be observed from Table 2 that, based on the RMSE criterion for the training data, the error value remained the same at P=20,50 as well as at P=100,250. However, when the number of variables was increased to P=500, the RMSE declined, reaching the lowest training error, indicating improved performance of the fuzzy model in representing the data. Although the improvement in RMSE compared with the conventional LASSO model is relatively small at P=500, incorporating the fuzzy inference system enables the proposed model to capture nonlinear relationships and uncertainty that the conventional LASSO model cannot adequately represent. The additional evaluation metrics support the RMSE results. The MAE and MAPE values generally decreased, while the R^2 values increased, particularly at P=500, indicating improved predictive performance for both the training and testing datasets. The consistency of these evaluation metrics further confirms the effectiveness of the proposed Fuzzy-LASSO model.

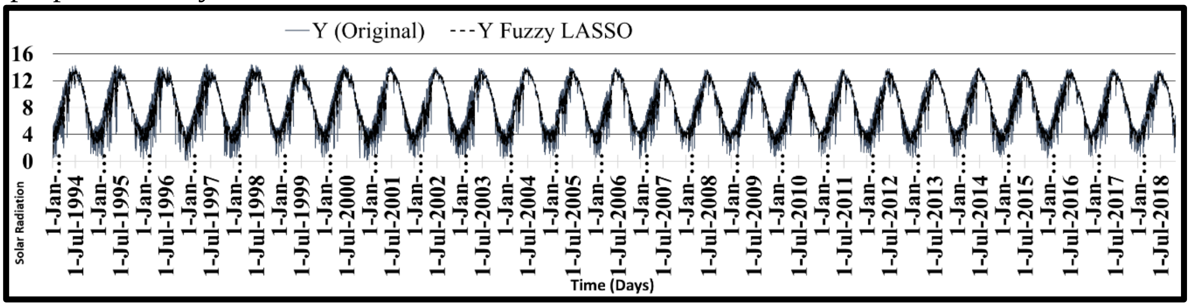

Figure (6): Comparison between observed and predicted solar radiation values using the Fuzzy LASSO model for the training dataset

Figure (6) shows that when the value of P=20 is used in a Fuzzy-LASSO model, the model's fit improves after fuzzing, as the estimated values become closer to the actual values, reflecting the fuzzy model's ability to deal with uncertainty.

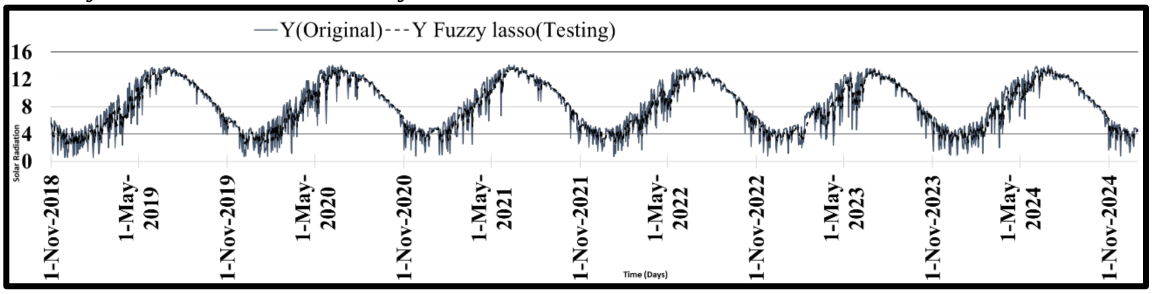

Figure (7): Comparison between observed and predicted solar radiation values using the (Fuzzy LASSO) model (Test dataset)

The conventional LASSO model was compared with the proposed Fuzzy-LASSO model using the RMSE criterion on the training data, as shown in Tables 1 and 2. The results showed that the Fuzzy-LASSO model outperformed the conventional LASSO model, as it recorded the lowest error criterion values. The outperformance of the fuzzy model is attributed to its ability to handle high-dimensional data and uncertainty, indicating that integrating fuzzy logic with conventional methods improves prediction accuracy, particularly in high-dimensional settings.

9. Results of EN Model

The table below shows the results of applying EN, including the selected variables and RMSE values for both the training and test data.

Table (3) The variables selected using EN and the RMSE values.

Model	Number of original variables (p)	Number of selected variables (h)	RMSE (Training)	RMSE (Test)	MAE (Train)	MAE (Test)	R^2 (Train)	R^2 (Test)	MAPE TRAIN	MAPE Test
EN	20	17	1.4402	1.4587	1.0172	0.9875	0.8387	0.8586	25.10%	23.74%
EN	50	18	1.4285	1.4469	0.9942	0.9650	0.8426	0.8626	24.59%	23.24%
EN	100	22	1.4057	1.4243	0.9942	0.9676	0.8442	0.8634	25.03%	23.64%
EN	250	21	1.3989	1.4174	0.9914	0.9646	0.8464	0.8663	24.99%	23.55%
EN	500	21	1.3886	1.4070	0.9868	0.9582	0.8485	0.8676	24.95%	23.54%
EN	1000	20	1.3879	1.4062	0.9867	0.9576	0.8486	0.8679	24.97%	23.49%

The table shows the results of applying EN to select variables at different values of the original variable P. It can be observed from the table that the RMSE values for the training data gradually decrease with the increase in the number of variables, as they decreased from 1.4402 at P=20 to 1.3879 at P=1000, and the lowest error value was at the value of 1000, which indicates the ability of EN to handle high-dimensional data. The additional evaluation metrics are consistent with the RMSE results. In general, the MAE and MAPE values decreased, whereas the R^2 values increased as the number of original variables increased for both the training and testing datasets. These results provide further evidence of the improved predictive performance of the EN model and confirm the reliability of the obtained results.

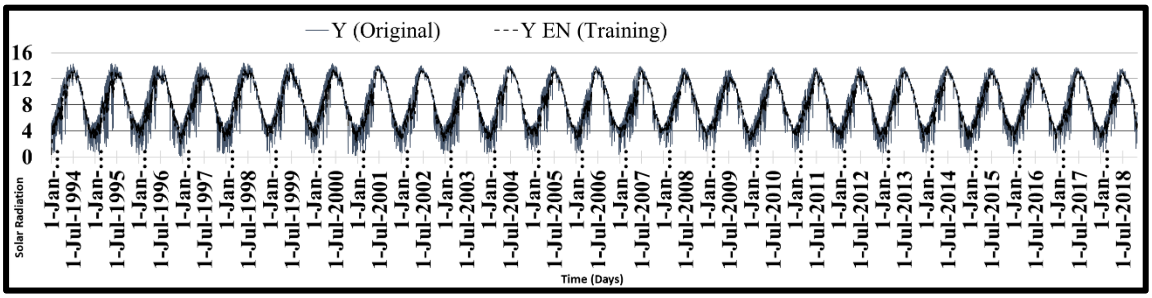

Figure (8): Comparison between observed and predicted solar radiation values using the model (EN) training dataset

Figure 8 shows the EN model with P = 20, which fits the actual values better than LASSO because it combines shrinkage and selection.

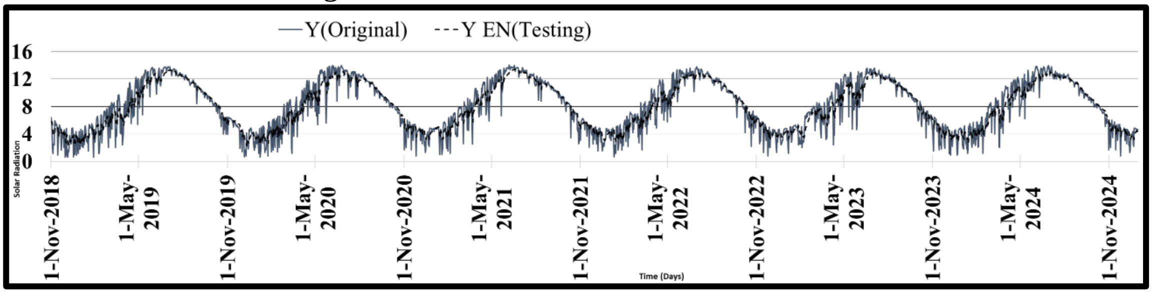

Figure (9): Comparison between observed and predicted solar radiation values using the (Fuzzy EN) model (Test dataset)

10. Results of Fuzzy EN

After the variables were selected using the conventional EN method, the selected variables were converted to fuzzy sets to build a fuzzy inference system using the Subtractive Clustering algorithm. Several values of the cluster influence range r were tested, and the optimal value was chosen based on the lowest RMSE value of the training data.

The following steps were followed in the fuzzing process:

1- Initialisation of the data for the explanatory variables, which are time lags [13], with different numbers.

$$X=[X_{t-1} \quad X_{t-2} \quad \ldots\ldots \quad X_{t-n}] \tag{9}$$

2- Applying EN regression to the inputs (1) to obtain a reduced number of actually influential and significant variables with their coefficient values.
3- Obtain the output variable $\hat{Y}_{EN}$ and calculate Error criteria such as root mean square error (RMSE).
4- Weighting the final variables in EN through the values of their significant coefficient and specifying them as inputs for the Fuzzy Inference System (FIS).
5- Determining the Membership Function for input and output variables, in addition to determining one of the Fuzzy Inference System (FIS) algorithms, and the Subtractive Clustering algorithm was used.
6- Applying the Fuzzy Inference System (FIS) and obtaining the final outputs $\hat{Y}_{FIS}$ and calculating one of the error criteria, RMSE, to compare the outputs between EN before and after fuzzing.

Table (4) below shows the Fuzzy -EN results, RMSE values for both training and test data, as well as the membership functions, rules, and best r

Table (4) shows the RMSE values, membership functions, and number of rules for the Fuzzy-EN model.

H	Best **r** for training	MF (Training)	Rules (Training)	RMSE (Training)	RMSE (Testing)	MAE (Train)	MAE (Test)	R^2 (Train)	R^2 (Test)	MAPE TRAIN	MAPE Test
17	0.3	6	6	1.4148	1.4334	0.9782	0.9567	0.8445	0.8617	23.95%	22.63%
18	0.5	5	5	1.4129	1.4316	0.9768	0.9563	0.8445	0.8623	24.00%	22.78%
22	0.5	5	5	1.4102	1.4295	0.9740	0.9580	0.8450	0.8614	23.93%	22.77%
21	0.3	7	7	1.4076	1.4263	0.9722	0.9550	0.8457	0.8612	23.91%	22.71%
21	0.7	5	5	1.3985	1.4171	0.9692	0.9443	0.8478	0.8653	23.89%	22.68%
20	0.7	5	5	1.3877	1.4162	0.9695	0.9429	0.8477	0.8657	23.90%	22.65%

The table presents the results of fuzzifying the selected variables using conventional EN. The number of fuzzy rules and membership functions was determined by the radius r, which was selected based on the training data. The error value for the training data decreases gradually from 1.4148 at p=20 to 1.3877 at p=1000, representing the lowest training error value shown in the table. This indicates the fuzzy model's ability to handle high-dimensional data. The additional evaluation metrics are consistent with the RMSE results. In general, the MAE and MAPE values decreased, whereas the R^2 values increased for both the training and testing datasets. These consistent improvements across all evaluation metrics further demonstrate the effectiveness of the proposed Fuzzy-EN model in enhancing predictive performance.

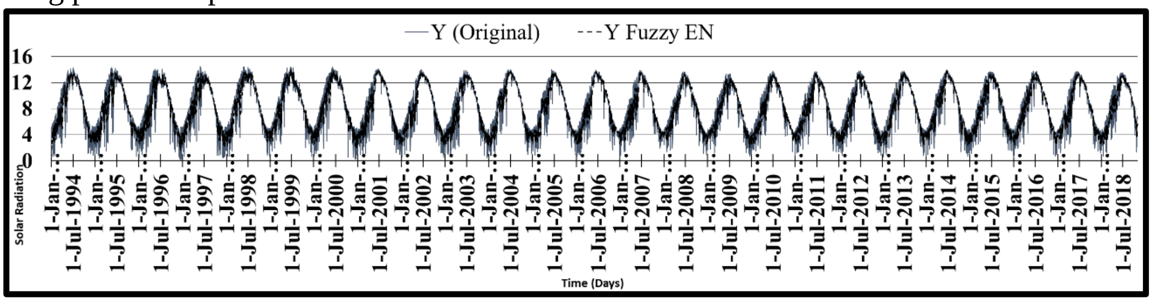

Figure (10): Comparison between observed and predicted solar radiation values using the model (Fuzzy EN) training dataset

Figure (10) shows that when the value of P=20, the Fuzzy-EN model achieves the best fit with the real values, because it combines the flexibility of EN and the ability of the fuzzy model to represent uncertainty.

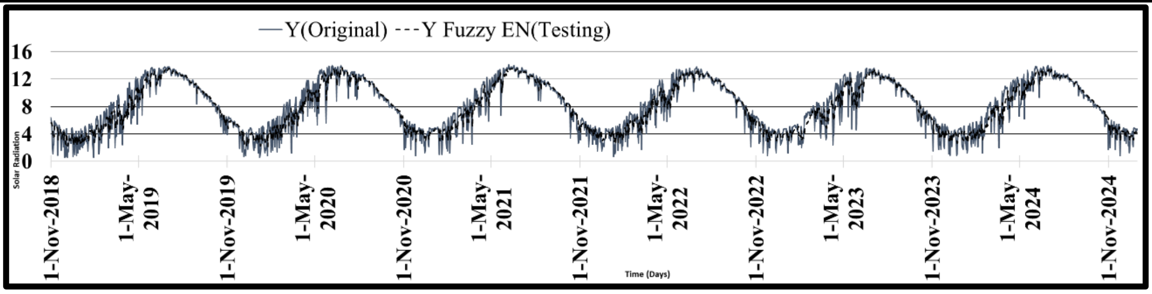

Figure (11): Comparison between observed and predicted solar radiation values using the (Fuzzy EN) model (Test dataset)

11. Discussion

The proposed hybrid models improve prediction accuracy by combining the variable selection capability of LASSO and Elastic Net with the ability of the fuzzy inference system to model nonlinear relationships and uncertainty. Consequently, the hybrid models are particularly effective for high-dimensional datasets characterised by correlated variables and nonlinear patterns, where conventional linear regularisation methods may not fully capture the underlying data structure. However, the proposed framework introduces additional computational cost due to the fuzzy inference stage, and its performance depends on the appropriate selection of the regularisation and clustering parameters.

The proposed hybrid models introduce an additional computational stage through the fuzzy inference system (FIS), resulting in a higher computational cost than the conventional LASSO and EN models. However, this cost is substantially reduced by the variable selection stage, which significantly decreases the number of predictor variables before constructing the fuzzy inference system. Consequently, the fuzzy inference system is applied only to the selected variables rather than to the entire high-dimensional dataset, thereby improving computational efficiency. Furthermore, the proposed framework is scalable to high-dimensional datasets because the dimensionality reduction performed by the regularisation models limits the computational burden of the fuzzy stage. Although the execution time increases slightly due to the additional fuzzy inference layer, the improvement in prediction accuracy justifies the additional computational effort.

12. Conclusions

As shown in the results section above, the hybrid methods Fuzzy-EN and Fuzzy-LASSO outperform the conventional models LASSO and EN. Hybrid methods recorded the Lowest prediction error rates and more stable performance. Furthermore, the hybrid models enhance prediction accuracy for high-dimensional data by effectively minimising the number of significant variables. Additionally, the models improve positively as the number of original variables increases. Based on these results, the Fuzzy-EN model appears better able to handle high-dimensional data by reducing the number of variables and focusing on the most significant ones, compared to the LASSO model. It can also be concluded that integrating fuzzy logic with both the LASSO and Fuzzy-EN models improved prediction accuracy and the hybrid models' ability to handle high-dimensional data, despite the inherent challenges posed by such data.

Despite the promising results achieved by the proposed hybrid models, this study has several limitations. First, the analysis was conducted using solar radiation data from a single geographic location (Mosul), which may limit the generalizability of the findings. Second, only one climatic variable was considered. In addition, the predictive performance depends on the selection of the regularisation and clustering parameters. Finally, the proposed models were not compared with recent machine learning and deep learning algorithms, which could be investigated in future research.

13. Recommendation

Hybrid (fuzzy) regularisation techniques based on fuzzy logic, such as fuzzy LASSO and fuzzy EN, are recommended for high-dimensional data that are inherently uncertain and ambiguous.

These techniques improve variable selection efficiency and model stability by addressing multiple linear dependencies while accounting for data ambiguity.

14. Supplementary Material

(None).

15. Author's Contributions

Israa Khamees Ahmed: Designed the research. The results were analysed using MATLAB. Interpreted the results and contributed to the discussion.

16. Funding

(None).

17. Data Availability Statement

Nineveh Governorate Meteorological Centre.

18. Acknowledgements

The author is very grateful to the University of Mosul/College of Computer Sciences and Mathematics for providing facilities, which helped improve the quality of this work.

19. Conflict of Interest

The authors declare no conflict of interest.

20. Declaration of generative AI use

During the preparation of this work, the authors did not use AI or program translation for grammar checking and language polishing. The authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 58, pp. 267-288, 1996. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[2] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 67, pp. 301-320, 2005.https://doi.org/10.1111/j.1467-9868.2005.00503.x

[3] V. Plakandaras, R. Gupta, P. Gogas, and T. Papadimitriou, "Forecasting the US real house price index," Economic Modelling, vol. 45, pp. 259-267, 2015.https://doi.org/10.1016/j.econmod.2014.10.050

[4] L. A. Zadeh, "Fuzzy sets," Information and control, vol. 8, pp. 338-353, 1965.https://doi.org/10.1016/S0019-9958(65)90241-X

[5] H. Asai, S. Tanaka, and K. Uegima, "Linear regression analysis with fuzzy model," IEEE Trans. Systems Man Cybern, vol. 12, pp. 903-907, 1982.https://doi.org/10.1109/TSMC.1982.4308925

[6] R. Pereira, A. Fagundes, R. Melicio, V. Mendes, J. Figueiredo, and J. Quadrado, "Fuzzy subtractive clustering technique applied to demand response in a smart grid scope," Procedia Technology, vol. 17, pp. 478-486, 2014.https://doi.org/10.1016/j.protcy.2014.10.256

[7] A. Mustikasari and A. D. Pahrany, "Lasso Regression and Elastic Net in Analysing Factors Affecting the Open Unemployment Rate," Jurnal Varian, vol. 8, pp. 151-164, 2025.https://doi.org/10.30812/xey50x64

[8] A. E. Hoerl and R. W. Kennard, "Ridge regression: applications to nonorthogonal problems," Technometrics, vol. 12, pp. 69-82, 1970.https://doi.org/10.1080/00401706.1970.10488635

[9] A. R. Nur, A. K. Jaya, and S. Siswanto, "Comparative analysis of ridge, lasso, and elastic net regularization approaches in handling multicollinearity for infant mortality data in south sulawesi," Jurnal Matematika, Statistika dan Komputasi, vol. 20, pp. 311-319, 2024.https://doi.org/10.20956/j.v20i2.31632

[10] H. H. Tang and N. S. Ahmad, "Fuzzy logic approach for controlling uncertain and nonlinear systems: a comprehensive review of applications and advances," Systems Science & Control Engineering, vol. 12, p. 2394429, 2024.https://doi.org/10.1080/21642583.2024.2394429

[11] L. Kong, "Fuzzy linear regression model based on adaptive lasso method," International Journal of Fuzzy Systems, vol. 24, pp. 508-518, 2022.https://doi.org/10.1007/s40815-021-01156-0

[12] Y. Gao, Q. Wu, Z. Xu, C. Cao, J. Pan, G. Shao, et al., "Principal component analysis with fuzzy elastic net for feature selection,"IEEE Transactionson Fuzzy Systems, vol. 32,pp.6878-6890, 2024.https://doi.org/10.1109/TFUZZ.2024.3466926

[13] N. Tak and D. Inan, "Type-1 fuzzy forecasting functions with elastic net regularization," Expert Systems with Applications, vol. 199, p. 116916, 2022.https://doi.org/10.1016/j.eswa.2022.116916

https://doi.org/10.31272/jae.i153.1610
https://admics.uomustansiriyah.edu.iq
P-ISSN: 1813–6729 E-ISSN: 2707-1359

JAE
OPEN ACCESS

اختيار المتغيرات في بيانات السلاسل الزمنية للاشعاع الشمسي عالية الابعاد باستخدام أسلوب هجين بالاعتماد على LASSO و الشبكة المرنة و المنطق الضبابي

اسراء خميس احمد [1]، أسامة بشير شكر [2]
[1، 2] قسم الإحصاء والمعلوماتية، كلية علوم الحاسب والرياضيات، جامعة الموصل، الموصل، العراق.

المستخلص

ان أساليب اختيار المتغيرات والتنبؤ بها في السلاسل الزمنية عالية الابعاد تواجه مشاكل عديده لانها تحتوي على عدد كبير من المتغيرات مما يؤدي الى صعوبة في اختيار المتغيرات و انخفاض دقة التنبؤ، والمتغيرات المناخية لانها تتسم بالتعقيد و عدم الخطية بالاضافه الى الارتباط العالي بين المتغيرات مما يحد من كفاءه النماذج التقليدية على التعامل مع البيانات عالية الابعاد وعدم التأكدية .تهدف هذه الدراسة الى بناء اطار تنبؤي هجين يجمع بين اساليب الانحدارالمنتظم Regularization التقليدية يعد انحدار LASSO و انحدار EN من الطرق المستخدمة في اختزال عدد المتغيرات التفسيرية للبيانات عالية الابعاد والتي سيتم استخدامها في هذه الدراسة للتنبؤ و اختزال عدد المتغيرات المناخية التفسيرية وبين المنطق الضبابي Fuzzy logic و ذلك من خلال بناء نماذج هجينة LASSO-Fuzzy و EN-Fuzzy بهدف تحسين دقة التنبؤ و كفاءة اختيار المتغيرات من خلال حل مشكلة الابعاد العالية وعدم التأكدية وعدم الخطية التي تعتبر من المشاكل التي تعاني منها البيانات، سيتم استخدام بيانات السلسلة الزمنية للاشعاع الشمسي عالية الابعاد المدينة الموصل للتطبيق باستخدام الاساليب المقترحة، تكمن اهميه الاسلوب الهجين على اختزال عدد المتغيرات لبيانات الاشعاع الشمسي عالية الابعاد مما يقلل من خطا التنبؤ وتحسين قابلية النموذج للتفسير، اظهرت النتائج تفوق الاساليب الهجينة على النماذج التقليدية مثل LASSO و EN وقد سجلت النماذج الهجينة اقل المعدلات لخطأ التنبؤ و اداء اكثر استقراراً في اختيار المتغيرات المؤثرة للبيانات عالية الابعاد و للتقليل منها.
من الممكن استنتاج ان النماذج الهجينة LASSO-Fuzzy و EN-Fuzzy تزيد من دقة التنبؤات للبيانات عالية الابعاد من خلال التقليل الفعلي للمتغيرات و اختيار افضلها المؤثرة فعلياً.

معلومات البحث

تواريخ البحث:
التقديم: 10 / 05 / 2026
المراجعة: 06 / 08 / 2026
قبول النشر: 20 / 08 / 2026
نشر الكتروني: 01 / 09 / 2026
تسلسل الصفحات: 138 - 151

الكلمات المفتاحية:
الشبكة المرنة، لاسو، الشبكة المرنة المضببة، لاسو المضبب.

المراسلة:
أسم الباحث: أسامة بشير شكر الحنون

Email: drosamahannon@uomosul.edu.iq