

## Research Article

# An Application of Fuzzy Multiple Linear Regression in Biological Paradigm

Saima Mustafa <sup>1</sup>, Shumaila Ghaffar,<sup>1</sup> Murrium Bibi,<sup>1</sup> Muhammad Ghaffar Khan <sup>2</sup>,  
Qaisara Praveen,<sup>3</sup> Harish Garg <sup>4</sup>, and Mahamane Saminou <sup>5</sup>

<sup>1</sup>Department of Mathematics and Statistics, PMAS Arid Agriculture University, Rawalpindi, Pakistan

<sup>2</sup>Institute of Numerical Sciences, Kohat University of Science and Technology, Kohat, Pakistan

<sup>3</sup>Division of Continuing Education, PMAS-Arid Agriculture University, Rawalpindi, Pakistan

<sup>4</sup>School of Mathematics, Thapar Institute of Engineering and Technology, Deemed University, Patiala 147004, Punjab, India

<sup>5</sup>Department of Mathematics and Computer Science, University of Agadez, Agadez, Niger

Correspondence should be addressed to Mahamane Saminou; mahamanesaminou@gmail.com

Received 16 March 2022; Revised 17 June 2022; Accepted 28 June 2022; Published 14 September 2022

Academic Editor: Zakia Hammouch

Copyright © 2022 Saima Mustafa et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The regression model is generally utilized in several fields of study because of its applications. Regression is an extremely incredible approach; it builds up a connection between dependent and independent variables. We have addressed a powerful computational model by utilizing dengue information joined with fuzzy multiple linear regression. Information is accumulated on dengue fever through the survey. This paper is centered on the comparison of the crisp method with fuzzy multiple linear regression, and then, the utilization of a fuzzy multiple regression method is explained after the comparison. We have used multiple regression and then converted the said technique into three fuzzy cases. The effectiveness of the fuzzy multiple regression model is measured by numerical computation and comparison of both techniques. 2020 Mathematics Subject Classification. Primary 30C45; 30C50; 30C80; Secondary 11B65, 47B38.

## 1. Introduction

Regression is a method for determining the statistical relationship between two or more variables where a change in a dependent variable is associated with a change in one or more independent variables [1]. Multiple linear regression describes the relationship of one dependent variable with more than one independent variable. This is a statistical technique that is used in examining how multiple independent variables and dependent variables are related. Certain assumptions need to be fulfilled for achieving better results from multiple regression such as linearity, normality, no multicollinearity, and homoscedasticity. Fuzzy set theory was first introduced by Zadeh [2] in 1965, and it is a technique used to handle vague, uncertain, imprecise, or unclear information. This technique is appropriate in the case of vague information [1]. Many recent developments of fuzzy and its applications have been explored by different

researchers [3–5]. They have applied different fuzzy tools, and the applications of these tools have been explored in different fields such as decision-making and logistics processes. Fuzzy regression was proposed by the Japanese researcher Tanaka (1982) [6]. The model of fuzzy linear regression has been treated from diverse points of view which depend upon the type of data given in the input and data try to be achieved from the output [7]. Fuzzy regression can be used in very complex systems in the real world such as economy, marketing, finance, ecology, and industry, see [7]. Dengue fever is a viral illness caused by mosquito bites that are responsible for infecting approximately 96 million infected people in America yearly [8]. In recent decades, dengue has increased in geographic incidence and distribution. The impact of climatic factors on transmission was investigated by researchers in [9, 10]. Zadeh et al. [2] gave the idea of fuzzy set theory to deal with the vagueness and uncertainties occurring in decision making. It is one of the

applications which is in multisensing paradigm-based urban air quality monitoring and hazardous gas source analyzing [11]. Multiple linear regressions are often used to explain the relationship of multiple independent variables. This method is particularly appropriate for disease models and is widely used in the health sciences. In this paper, we discuss the problems that occur in different situations of multiple linear regression. First, it involves a linear relationship between the independent and dependent variables which do not give clear results based on the given information. Second, it specifies the distribution of errors normally distributed between the observed and expected values which cause impreciseness. Third, it assumes the data do not contain multicollinearity. Another difficulty can be caused by nonpreciseness and vague observations that occur frequently in practice. Due to these facts, we used the term “fuzzy” in multiple linear regression which overcome all these difficulties. This paper focuses on multiple linear regressions in a biological paradigm. Here is the numerical computation depending on the dengue study.

## 2. Fuzzy Sets and Numbers

The fuzzy set is defined as a set that contracts with vague boundaries [2]. The set consists of fuzzy logic and linguistic variables. Fuzzy sets are represented as

$$A = \frac{\mu_1}{X_1} + \frac{\mu_2}{X_2} + \dots + \frac{\mu_n}{X_n}, \quad (1)$$

where  $\mu_i$  and  $n$  express the membership function of  $X_i$ ,  $i = 1, 2, \dots, n$  in  $A$ , and the union is denoted by the plus sign [12]. Fuzzy sets are scientific, mathematical models of unclear quantitative or qualitative data, as often as possible, which are generated from natural languages. The model depends upon the generalization of the characteristics function of a set and the classical concept.

## 3. Multiple Linear Regression Model

Sir Francis Galton [1], an English Victorian, introduced the term regression. The general parametric equation is

$$\begin{aligned} Y &= f(X) + \epsilon, \\ Y &= \tau_0 + \tau_1 X_1 + \tau_2 X_2 + \dots + \tau_q X_q + \epsilon, \end{aligned} \quad (2)$$

where  $Y$  and  $X$  represent the dependent and independent variables. The coefficients  $\tau_1, \tau_2, \dots, \tau_q$  represent slopes, and  $\epsilon$  is the random error.

A fuzzy regression technique was first proposed by Tanaka [6]. We have considered the following cases of dependent and independent variables.

Case 1

$$\tilde{z}_i = \tau_0 + \tau_1 \tilde{k}_1 + \epsilon_i. \quad (3)$$

Case 2

$$\tilde{z}_i = \tilde{\tau}_0 + \tilde{\tau}_1 k_1 + \epsilon_i. \quad (4)$$

Case 3

$$\tilde{z}_i = \tilde{\tau}_0 + \tilde{\tau}_1 \tilde{k}_1 + \epsilon_i. \quad (5)$$

In the above models, here,  $\tau_0$  and  $\tau_1$  are the intercept and slope of the regression line respectively.  $Z_i$  are the fuzzy responses. In case 1, the parameters  $\tau_0$  and  $\tau_1$  are crisp parameters, and  $K_i$  are fuzzy. In Case 2, the parameters  $\tau_0$  and  $\tau_1$  are fuzzy but  $K_i$  are crisp. In case 3, the predictor and parameters are all fuzzy.

Consider the multiple fuzzy regression model which can be generalized as follows:

$$\tilde{y}_i = \tilde{\tau}_0 + \tilde{\tau}_1 \tilde{x}_{j1} + \tilde{\tau}_2 \tilde{x}_{j2} + \dots + \tilde{\tau}_p \tilde{x}_{jp} + \epsilon_i. \quad (6)$$

Using the centered values of the crisp predictor, the above equation can be written in matrix form as follows:

$$\tilde{y} = \tilde{x} \tilde{\tau} + \epsilon, \quad (7)$$

where  $\tilde{y}$  is a  $(n \times 1)$  fuzzy vector,  $\tilde{x}$  is a  $(n \times p)$  matrix of  $p$  fuzzy predictors, and  $\tilde{\tau}$  is a  $(p \times 1)$  vector of unknown  $p$  fuzzy parameters. As a result of the lack of linearity of  $F_c(R^p)$ ,  $\epsilon$  is reduced to nonfuzzy random variable (FRV) $\epsilon$ .

$$\begin{aligned} y_{ja} &= x_{j1a} \tau_{1a} + x_{j2a} \tau_{2a} + \dots + x_{jpa} \tau_{pa}, \\ y_{jb} &= x_{j1b} \tau_{1b} + x_{j2b} \tau_{2b} + \dots + x_{jpb} \tau_{pb}, \\ y_{jc} &= x_{j1c} \tau_{1c} + x_{j2c} \tau_{2c} + \dots + x_{jpc} \tau_{pc}, \end{aligned} \quad (8)$$

where  $y_{ja}, y_{jb}, y_{jc}$  are left, middle, and right values, respectively.

The  $\tilde{\tau}_{ja}$  is as follows:

$$\tilde{\tau}_{ja} = (x_{ja}^T x_{ja})^{-1} (x_{ja}^T y_{ja}). \quad (9)$$

On the same lines, the above equations can be simplified as

$$\begin{aligned} \tilde{\tau}_{jb} &= (x_{jb}^T x_{jb})^{-1} (x_{jb}^T y_{jb}), \\ \tilde{\tau}_{jc} &= (x_{jc}^T x_{jc})^{-1} (x_{jc}^T y_{jc}). \end{aligned} \quad (10)$$

Consider the multiple fuzzy regression model is generalized as follows:

$$\tilde{y}_i = \tilde{\tau}_0 + \tilde{\tau}_1 \tilde{x}_{i1} + \tilde{\tau}_2 \tilde{x}_{i2} + \dots + \tilde{\tau}_p \tilde{x}_{ip} + \epsilon_i, \quad (11)$$

where  $\tilde{y}$  is a  $(n \times 1)$  fuzzy vector,  $x$  is a  $(n \times p)$  matrix of  $p$  crisp predictors, and  $\tilde{\tau}$  is a  $(p \times 1)$  vector of unknown  $p$  fuzzy parameters. As a result of the lack of linearity of  $F_c(R^p)$ ,  $\epsilon$  is reduced to a nonfuzzy random variable (FRV) $\epsilon$ .

$$\begin{aligned} y_{ja} &= x_{j1} \tau_{1a} + x_{j2} \tau_{2a} + \dots + x_{jp} \tau_{pa}, \\ y_{jb} &= x_{j1} \tau_{1b} + x_{j2} \tau_{2b} + \dots + x_{jp} \tau_{pb}, \\ y_{jc} &= x_{j1} \tau_{1c} + x_{j2} \tau_{2c} + \dots + x_{jp} \tau_{pc}, \\ y_{jd} &= x_{j1} \tau_{1d} + x_{j2} \tau_{2d} + \dots + x_{jp} \tau_{pd}, \end{aligned} \quad (12)$$

where  $y_{ja}, y_{jb}, y_{jc}, y_d$  is lower, middle-lower, middle-upper, and upper values correspondingly.

The  $\hat{\tau}_{ja}$  is as follows:

$$\hat{\tau}_{ja} = (x^T x)^{-1} (x^T y_{ja}). \quad (13)$$

On the same lines above, the equation is simplified as

$$\begin{aligned} \hat{\tau}_{jb} &= (x^T x)^{-1} (x^T y_{jb}), \\ \hat{\tau}_{jc} &= (x^T x)^{-1} (x^T y_{jc}), \\ \hat{\tau}_{jd} &= (x^T x)^{-1} (x^T y_{jd}). \end{aligned} \quad (14)$$

## 4. Numerical Computation

This research depends on the primary data of dengue fever patients collected from two public sector hospitals in Rawalpindi named (Benazir, Holyfamily), respectively. After taking the data, we applied principal component analysis to eliminate insignificant variables and consider those variables that have a significant impact on our study. We have implemented the Keiser-Meyer-Olkin and Bartlett's test for the reduction of data.

### 4.1. Normality Test

$H_0$ : the data are normal.

$H_1$ : the data are normal.

If the  $p$  value is greater than 0.05, then this is normal.

Table 1 indicates that all the  $p$  values are greater than 0.05, so we conclude that our data are normal.

Figure 1 indicates that all the  $p$  values are greater than 0.05 so we concluded that our data are normal.

**4.2. Multicollinearity.** Here, using the VIF (variance inflation factor) values, we have seen that each value of VIF is below 10, and the assumption is fulfilled.

The multiple linear regression results are explained in Table 2 and represented by Figure 2.

**4.3. Interpretation.** The estimated values in Table 3 represent the regression coefficient which shows that the values of dengue fever increased by 0.043 units for one element increase in age, decreased 0.104 units for a unit increase in suffering fever, decreased 0.004 units for one element add-in checkup, decreased by 0.045 units for the unit increase in BP ( $U$ ), and decreased 0.088 units for one unit add-in BP ( $L$ ).

These results are represented in Figure 2 which shows the best fitted classical multiple regression model.

**4.3.1. Case 1.** Fuzzy multiple linear regression model with fuzzy independent variables and crisp dependent variable.

The fuzzy multiple linear regression results are explained in Table 4 and represented by Figure 3.

**4.4. Interpretation.** The estimated fuzzy regression coefficient shows that rates of  $Y$  increase by 0.037 units for one

TABLE 1: Normality test.

Variables	$P$ value
Age	0.178
Suffering	0.322
Checkup	0.334
BPU	0.103
BPL	0.212

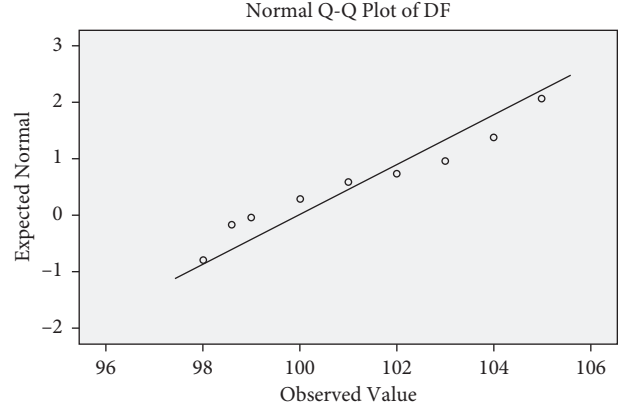


FIGURE 1: Data are normal.

TABLE 2: Multiple regression.

Variables	Constant	SE	$t$	$P$ value
Intercept	108.473	6.350	17.083	0.001
Age	0.043	0.052	0.841	0.403
Suffering fever	-0.104	0.289	-0.360	0.720
Check up	-0.004	0.668	-0.006	0.996
BPU	-0.045	0.060	-0.756	0.452
BPL	0.088	0.075	-1.167	0.246

element add-in  $X_1$ , 0.114 divisions on behalf of the unit increase in  $X_2$ , 0.014 entities for the unit increase in  $X_3$ , 0.064 units for the unit increase in  $X_4$ , and increased 0.065 units for the unit rise in  $X_5$ .

**4.4.1. Case 1.** Fuzzy multiple linear regression model with fuzzy independent variables and crisp dependent variable.

The multiple linear regression results are explained in Table 5 and represented by Figure 4.

**4.5. Interpretation.** The estimated fuzzy regression coefficient indicates that the value of  $Y$  increases by 0.200836 units for a unit increase in  $X_1$ , 0.277264 units for the unit rise in  $X_2$ , 0.024512 units for the unit increasing in  $X_3$ , decreases by 0.09249 units for the unit increase in  $X_4$ , and increases 0.43941 units for the unit increase in  $X_5$ .

## 5. Performance Comparison

The following section describes the performance of multiple linear regression and fuzzy multiple linear regression.

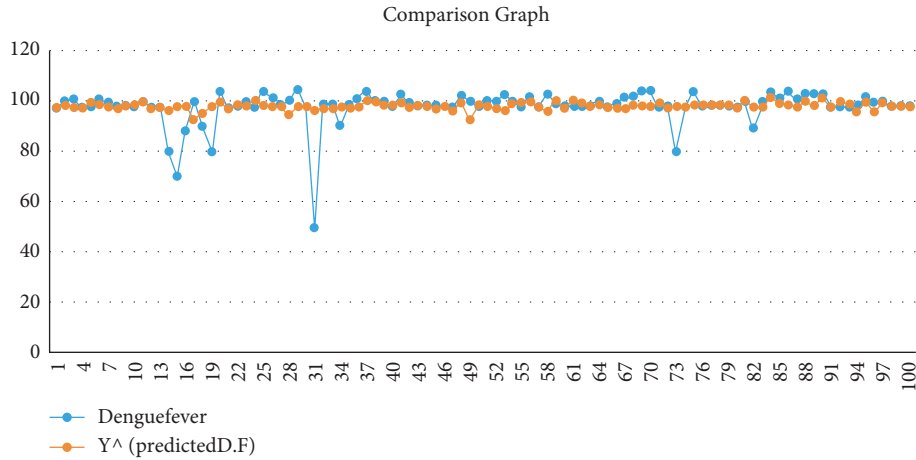


FIGURE 2: Best Fitted Graph of Multiple Linear regression.

TABLE 3: KMO and Bartlett’s test.

Kaiser-Meyar-Olkin measure adequacy	Measure of sampling	[.660]
Bartlett’s test of sphericity	Approx. chi-square	[53.67]
	DF	[10]
	Sig.	[.000]

TABLE 4: Result of fuzzy multiple regression when fuzzy independent variables and crisp-dependent variable in triangular membership function.

Variables	Constant	SE	<i>t</i>	<i>P</i> value
Intercept	0.513	0.063	8.073	0.001
Age	0.037	0.055	0.679	0.499
Suffering fever	0.114	0.050	2.294	0.024
Check up	0.041	0.082	0.499	0.619
BPU	0.064	0.082	0.779	0.438
BPL	0.065	0.044	1.495	0.138

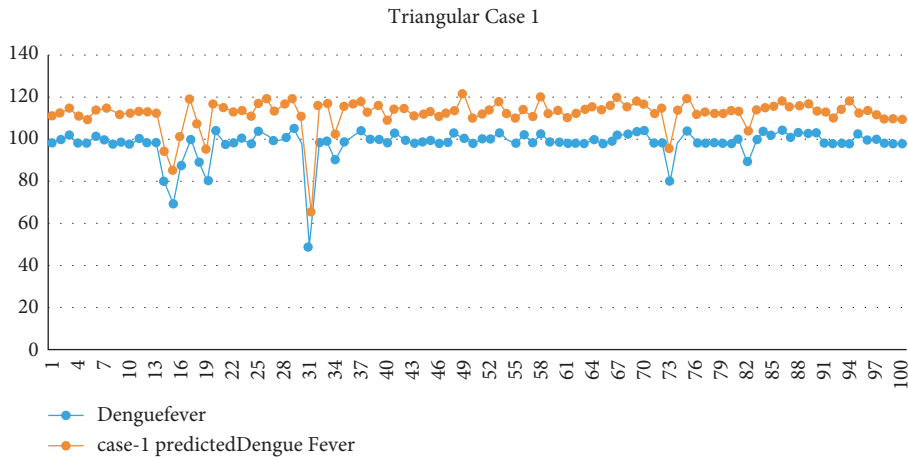


FIGURE 3: Fuzzy multiple regression using triangular membership function.

It is important to know which method from classical multiple linear regression and fuzzy multiple linear regression method is performing best and gives significant results. Comparison between these methods is made by

using different evaluation criteria. Results obtained by using different evaluation techniques are given in Table 6. The empirical analysis shows that the MSE, RMAE, BIC, and RAE in the case of fuzzy multiple linear regression are all

TABLE 5: Result of fuzzy multiple regression when fuzzy independent variables and crisp-dependent variable in trapezoidal membership function.

Variables	Constant	SE	<i>t</i>	<i>P</i> value
Intercept	2.114	0.147	14.420	0.001
Age	0.053	0.082	0.647	0.519
Suffering fever	0.160	0.068	2.358	0.020
Check up	0.099	0.058	1.700	0.092
BPU	-0.156	0.103	-1.514	0.133
BPL	0.289	0.112	2.584	0.011

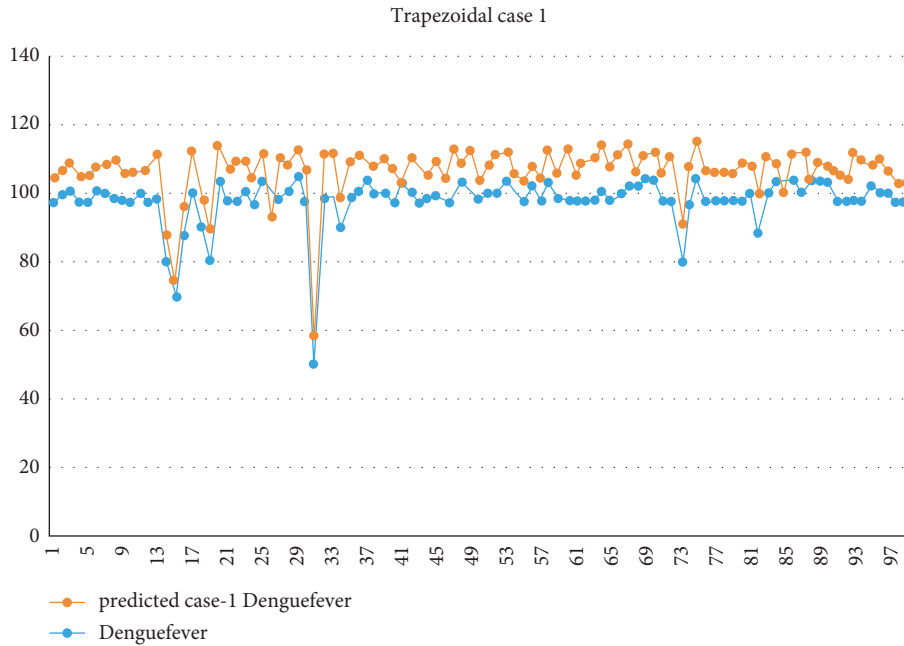


FIGURE 4: Fuzzy multiple regression using trapezoidal membership function.

TABLE 6: Error comparison of multiple regression and fuzzy multiple regression.

Error computation	Multiple regression	Fuzzy regression
MSE	49.03473	0.086875
BIC	944.26	-20.168
RMSE	7.00248	0.294746
RAE	4903.473	8.687522

TABLE 7: Error comparison of multiple regression and fuzzy multiple regression.

Error computation	Multiple regression	Fuzzy regression
MSE	49.03473	0.025824
BIC	944.26	-225.783
RMSE	7.00248	0.1607
RAE	4903.473	2.582449

smaller than the classical multiple linear regression method. This shows that fuzzy multiple regression can smooth the defuzzified forecast, and it is more consistent.

It is important to know which method is performing best and gives significant results among classical multiple linear regression and fuzzy multiple linear regression method. Comparison between these methods is evaluated by using different evaluation criteria. Results obtained by using different evaluation techniques are given in Table 7. The empirical analysis shows that the MSE, RMAE, BIC, and RAE in the case of fuzzy multiple linear regression are all smaller than the classical multiple linear regression method. This shows that fuzzy multiple regression can smooth the defuzzified forecast, and it is more consistent.

We have compared the fuzzy regression results of triangular and trapezoidal membership functions in Figure 4 which shows that the triangular membership function results were lower than that of trapezoidal so the triangular membership function is more efficient compared to the trapezoidal membership function and can be used in further studies for comparison.

## 6. Conclusion

There are numeral classical methods that are used to distribute accurate information, but in a lot of circumstances, accurate quantities cannot be achieved. This paper is based on the basic idea of the multiple linear regression method

and fuzzy multiple regression method. In the proposed work, the fuzzy multiple regression method is done by using the triangular and trapezoidal membership function and evaluating the computation of dengue fever data to express the efficiency of the proposed fuzzy multiple regression with the existing multiple regression model. The realistic result of the mean square error (MSE), Bayesian information criteria (BIC), root absolute error (RAE), and root mean square error (RMSE) of fuzzy multiple regression with triangular and trapezoidal membership function is smaller compared to the multiple linear regression which indicates that the proposed method has a better performance as compared to multiple regression [13–17].

### Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### References

- [1] K. Kumari and S. Yadav, "Linear regression analysis study," *Journal of the practice of Cardiovascular Sciences*, vol. 4, no. 1, pp. 33–36, 2018.
- [2] L. A. Zadeh, "Information and Control," *Fuzzy sets*, vol. 8, no. 3, pp. 338–353, 1965.
- [3] S. Mustafa, I. Fatimah, and Y. B. Jun, "Modelling the logistic processes using fuzzy decision approach," *Hacettepe Journal of Mathematics and Statistics*, vol. 48, no. 2, pp. 552–563, 2018.
- [4] M. Bibi and S. Mustafa, "Modelling the logistic processes Using Fuzzy Soft sets," *Hacettepe Journal of Mathematics and Statistics*, vol. 48, no. 2, pp. 1–79, 2020.
- [5] S. Mustafa, N. Safdar, M. Bibi, A. F. Sayed, M. G. Khan, and Z. Salleh, "A Study of Bipolar Fuzzy Soft Sets and its Sppliation in Decision-Making Problems," *Data-Driven Fuzzy Multiple Criteria Decision Making and its Potential Applications 2021*, vol. 2021, Article ID 5742288, 12 pages, 2021.
- [6] H. Ishibuchi and H. Ishibuchi, "Identification of possibilistic linear systems by quadratic membership functions of fuzzy parameters," *Fuzzy Sets and Systems*, vol. 41, no. 2, pp. 145–160, 1991.
- [7] V. Marza and M. A. SeyyediSeyyedi, "Fuzzy multiple regression model for estimating software development time," *International Journal of Engineering Business Management*, vol. 1, no. 1, p. 9, 2009.
- [8] C. Papadopoulos, M. Spiliotis, I. Gkioukhis, F. Pliakas, and B. Papadopoulos, "Fuzzy linear regression analysis for groundwater response to meteorological drought in the aquifer system of Xanthi plain, NE Greece," *Journal of Hydroinformatics*, vol. 23, no. 5, pp. 1112–1129, 2021.
- [9] J. Saqib and Mustafa, *Fuzzy Estimation of Regression Parameters in the Spline Regression Model*, vol. 5, pp. 1–19, PMAS Arid Agriculture university Rawalpindi, Pakistan, 2016.
- [10] P. Škrabánek and N. Martínková, "Algorithm 1017," *ACM Transactions on Mathematical Software*, vol. 47, no. 3, pp. 1–18, 2021.
- [11] Z. Zhu, B. Chen, Y. Zhao, and Y. Ji, "Multi-sensing paradigm based analyzing," *Journal of Safety Science and Resilience*, vol. 2, no. 3, pp. 131–145, 2021.
- [12] M. Jezewski, R. Czabanski, and J. Leski, "Introduction to fuzzy sets," *Theory and Applications of Ordered Fuzzy Numbers*, vol. 11, pp. 3–22, 2017.
- [13] L. Abdullah and N. D. Khalid, "Prediction of carbon dioxide emission using fuzzy linear regression model: a case of developed and developing countries," *Journal of Sustainability Science and Management*, vol. 9, no. 1, pp. 69–77, 2014.
- [14] S. A. A. Naqvi, B. Jan, S. Shaikh et al., "Changing climatic factors favor dengue transmission in lahore, Pakistan," *Environments*, vol. 6, no. 6, pp. 71–85, 2019.
- [15] A. F. Shapiro, "Fuzzy random variables," *Insurance: Mathematics and Economics*, vol. 44, no. 2, pp. 307–314, 2009.
- [16] B., M. A. Springer and A. O. Sykes, "An introduction to regression analysis," *Austrian Journal of Statistics*, vol. 32, no. 3, pp. 239–257, 2003.
- [17] G. K. Yasuik and Guler, "N A study on multiple linear regression analysis," *Procedia-Social and Behavioral sciences*, vol. 106, pp. 234–240, 2013.