Machine Learning and Statically for Evaluating the Classification of Medical and Lab Services

Muna R. Harbi[1,],Loay E. George[2]

[1.]Department of Physics, General Directorate of Thi-Qar Education, Ministry of Education, Thi-Qar, Iraq.
[2]Assistant of University President for Scientific Affairs / University of Information Technology and Communication., Baghdad, Iraq.

Abstract
Artificial intelligence is a realistic choice for developing accurate outcome predictors, notably in health research. It is often referred to as a component of artificial intelligence, such as machine learning. AI requires specialized hardware and software to write and train machine learning algorithms. The methodology has regularly provided considerable insight into service models by processing an enormous amount of semi-structured, multi-domain medical data that is already available. Characterizing the results for allocating risk is where AI in healthcare can be used to improve determined disease models and provide opportunities for personalization and treatment discovery in Primary health centers and services distribution on population in all towns. This study aims to conduct an analysis for pattern identification and classification of medical services using Machine learning methodology for the Thi-Qar governorate, mainly distinguished by its great size (about 13738.67 km²), divided into 15 administrative units according to Iraq's formal administrative division. The utilized Databases were collected during the period from 2019 to June 2021. As a result, the maps produced represent medical services relative to population density and record the results. Thus, the services are categorized by machine learning algorithms and Programming in the R language. After attaining the results of the study area, the study concludes that the Governorate's medical services are divided into five categories: weak (26 percent) and strong (13 percent), and below the average category of health care, on the other hand, accounts for 33% of the total (5 administrative units). The Governorate's health services are included in the average and above-average categories (13 percent).

Keywords:Spatial Analysis, Machine Learning, Service Classification, GIS, plot matrix.

1. Introduction

Machine Learning (ML) is a technique for teaching machines to handle data more efficiently [1].Sometimes, we cannot interpret the extracted information after viewing the data. In that case, machine learning is used. These algorithms are used for various purposes, including data mining, image processing, predictive analytics, and so on[2]. It is essential to identify the patient and service variables that most accurately predict resource utilization to make effective financial allocations among service providers and inform clinical care [3]. Case-mix classification supports how to most effectively manage, compare, and pay for health care services [4].
The observations' locations and the relevant variables' values are increasingly included in datasets. It can be more effective to take into consideration the spatial aspect of the data, and in some circumstances, it may even be necessary for consistency[5].Numerous R language packages are evidence of this growing interest in spatial data [6]. Simple or multinomial logistic regression is used instead of linear regression when the response variable is a qualitative (dichotomic or

polychromic) variable, such as a population's access to healthcare or found absence of disease[7].Numerous techniques are available in R's basic installation for assessing the statistical presumptions used in a regression study [8].

Binary data's discontinuous nature for logistic regression models makes it challenging to understand such displays [9]. Three techniques for the diagnostic inspection of logistic regression models are developed through modifications and extensions of linear model displays [10]. Plots of the local mean deviation are helpful in identifying the general lack of fit. Plots of empirical probabilities aid in identifying lone departures from the fitted model [11]. This study aims to elaborate, systematically analyze, and classify recent advances related to the medical services analysis of physics and ML. Also, the importance of the research comes from the need to prepare a geographic information system (GIS) to study the spatial variation of distribution using all parameters in the study area and classification through ML.

This systematic and meticulous research's primary objective is to map and connect the knowledge landscape of AI and GIS analysis and classification approaches in the service and management domains. In order to achieve this goal, the current study aims to I extract the inductive topical framework to portray the AI research field in services and management, and more specifically for the highly focal domain of the "services model"; (ii) analyze and explain the key services and management latent themes and sub-themes in the research field of AI; and (iii) highlight the trend of services and management studies in the IoT field to detect novelty and emergence.

2. Literature Review

The proposed work focuses on developing a parasitic elements-based MIMO antenna with low mutual coupling using the Machine Learning (ML) technique for lower sub-6 GHz 5G applications. This design solves and analyzes the complexity of the optimum position determination of parasitic elements and ground plane dimensions of a composite MIMO antenna structure using an ML algorithm. The experimental results show good agreement with the simulated results. The proposed antenna has a radiation efficiency of more than 80%[2].So far, international service comparison has failed to deliver accurate data for health planning in various sectors, including health services[12].

Given that no datasets include these annotations, a search for characteristics is offered. The chosen characteristics are more effective at resolving HAR. The authors use an evolutionary algorithm for such a search. They start by populating action classes with random binary representations. Second, they assess a population by employing deep structures with sigmoid activation functions. Performance during validation serves as evolution fitness. On the populations, the authors use non-local mutations. They come to the conclusion that HAR performs better when attribute representations are used. Randomly selecting an attribute representation performs on par with directly categorizing human activities. The lack of a semantic definition for the attributes was a flaw in this method[13].

The most typical method is to use the plot () function on the object that the l m returns (). This results in four graphs that can be used to assess the model fit. Using this method to apply to simple linear regression[14].Scatterplot3d is an R package for the visualization of multivariate data in a three-dimensional space,With the scatterplot3d () function, you can provide symbols, axes, colors, lines, grids, and angles, among other things[15].Visual diagnostic displays can be particularly helpful for identifying and assessing aberrant aspects in the fit of a model to data in ordinary linear regression[16].

The conduct of studies intended to ascertain the influence of one or more variables on a certain response variable is quite common in the health sciences. Depending on the situation, single or multiple regression procedures are utilized when the response variable is numerical[17].The formula for the variance of τ∗ proposed in Pimentel's thesis cannot be applied when the parameters pij attain the values 0 or 1: the proof relies on the delta method which can be applied only when functions are differentiable. At the extrema of the interval [0,1] no function is differentiable, therefore, whenever any of the parameters pij attains the values 1 or 0, the standard R cor. test function for Kendall's correlation is used to obtain p-values, when applicable, as the function does not work in case of vectors with zero variance[18]. It should be emphasized that cor. test () computes the test for correlation between paired samples, whereas cor () computes the correlation coefficient[19][20]. It provides the correlation's p-value and significance level (or correlation coefficient) [21].

3. Data Collection

The University of Baghdad's College of Science sent facilitation guidance to the Thi-Qar Health Department/Public Health Department, the Thi-Qar Planning Department, and the Thi-Qar Municipality Department in order to facilitate a task. This guidance was used to facilitate data collection, analysis, and examination. The names of primary health center facilities were given to me. 127 centers lack coordinates, and some of them lack addresses as well. Accessing a comprehensive database for work and analysis took roughly four months.

3.1 Location of the study

The research area in the Thi-Qar Governorate is shown in Figure 1. They have fifteen districts and eight sub-districts that make up their administrative division. With a total area of (13738.95) km^2, it is the seventh-largest governorate in Iraq in terms of land area. It contains 15 administrative divisions, 900 km² of aquatic bodies, and historical sites[22].

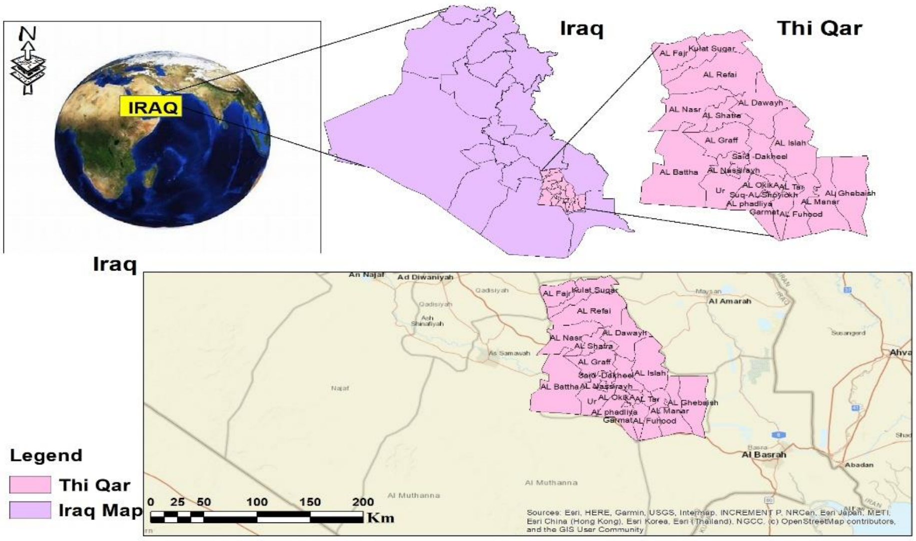


Figure 1.Location of the study area. Thi-Qar province in Iraq.

3.2Data collection technique

This study was conducted over a period of four months. The preparation phase included preparing research tools related to research planning, the field in which the study should be conducted, and obtaining the necessary authorization letters from the College of Science Physics Determent /Baghdad University. The permission letter was submitted to the responsible office of the Thi-Qar Health Department/Public Health Department, Thi-Qar Branch, The Directorate of Urban Planning, and the Directorate of Municipalities in Thi-Qar, where the study was conducted, and data was collected. The following table shows the data collected for this study as follows:

Table .1 Data Collection for this study

Code	x	y	Density	Doctors__Citizens	NO of Centers	NO of Pharmacy	NO of Lab	Lab technician
FO1	664365	3428691	86.385	0.000143	2	2	2	15
NA2	619845	3434597	756.2109	0.01376	26	26	26	137
SU3	640118	3418051	288.7095	0.001051	8	8	8	62
BN4	650014	3417416	200.3059	0.000388	4	4	4	33
GB5	690262	3427368	30.89463	0.000132	2	2	2	7
AS6	652458	3449329	43.33489	0.000843	2	2	2	8
DK7	635525	3444566	120.1251	0.000408	2	2	2	7
DW8	631292	3484783	110.5151	0.000416	3	3	3	23
GR9	617798	3464145	192.1055	0.000699	3	3	3	25
BA10	583442	3442650	27.59923	0.000438	2	2	2	7
NR11	605468	3489679	115.4886	0.000329	4	4	4	32
SH12	610072	3475258	680.6883	0.001068	8	8	8	169
RI13	605997	3509104	130.7338	0.000746	4	4	4	52
KA14	601602	3526169	159.714	0.000343	3	3	3	3
FJ15	591018	3531038	133.5139	0.000088	5	5	5	5

3.3 The Steps workflow

Names of health centers and coordinates (x, y) of each center, number of health staff, number of citizens registered in each health center, and population density 2021. many programs were used mapping and identifying, and classification areas.

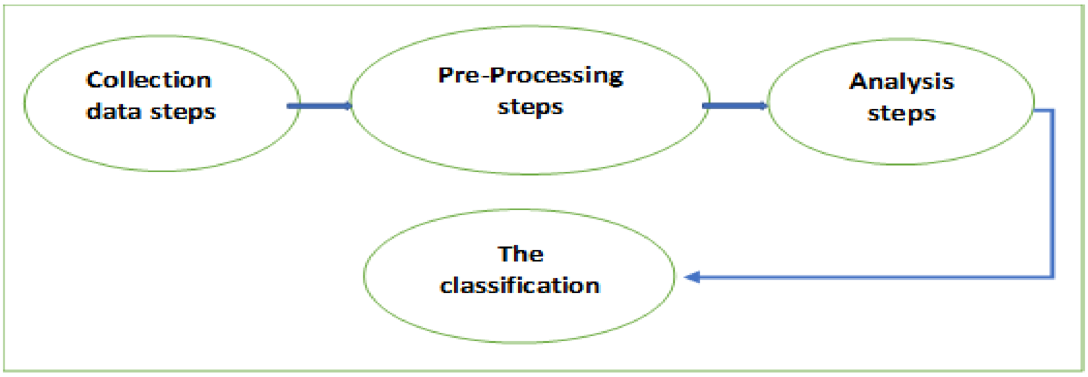


Figure 2.Workflow steps of the study's data.

4. Materials and Methods

This study aims to elaborate, systematically analyze systematically, and classify recent advances related to the medical services analysis of physics and machine learning. Firstly, comprehensive and determined algorithms in data type analysis were chosen, including the following: scatterplot matrix, Correlogram, the plot () function, and Logistic regression.

4.1 Steps for programming with Python and R

The scatter plot enables one to obtain a visual comparison of the two variables in the data set and determines what kind of relationship there might be between the two variables. The scatter plot is interpreted by evaluating the data in the following ways: a) strength (strong, moderate, or weak), b) trend (positive or negative), and c) shape (linear, non-linear, or none). Pearson's Correlation Coefficient, also known as Pearson's r, is the statistical test to determine the strength of the relationship. Curves or points created from data are recorded in the figure matrices. The most fundamental application of the Pandas scatters matrix technique is used to investigate the correlation coefficients for the variables in a dataset and to perform a correlation matrix in Python.as in Figure 3.

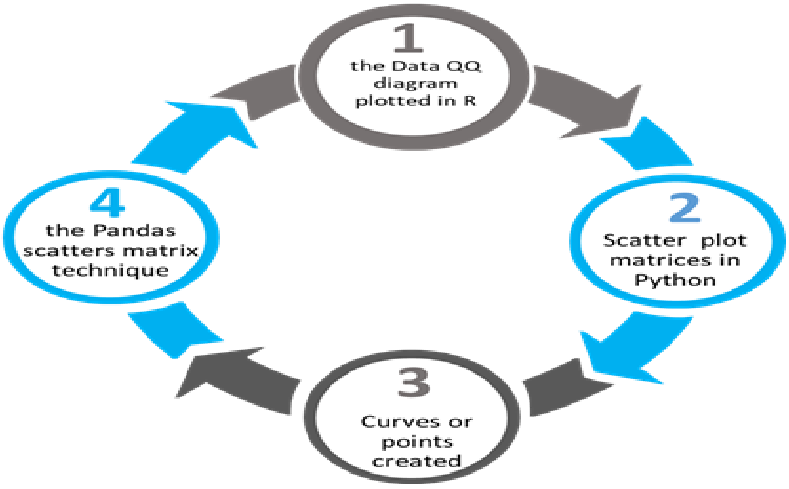


Figure 3. Phases of programming using Python and R

4.2 Logistic regression models

The probabilities of events are estimated using logistic regression models as functions of independent variables. Let x (j = l, k) represent the values of the k independent variables for case I and let y represent the value of the dependent variable for case i. Assume that Y is a binary variable measuring group membership. Using the coding y = 1 if case I belongs to that group and 0 otherwise, let p represent the likelihood that y = 1. Given by p/ are the probabilities that y = 1. (l-p). The natural logarithm of p is equal to the log odds, or logit of p. (l-p). The log chances are calculated using logistic regression as a linear combination of the independent variables[23].

$$Logit(p) = B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_k X_k \quad (1)$$

Where; 0, 1, and 2 are the subscripts for the lack of the medium's ability to subscribe at the time[24].

The sigmoid function is referred to as an activation function for logistic regression and is defined as:

$$F(x) = \frac{1}{1+e^{-x}} \quad (2)$$

Where: e = base of natural logarithms, value = numerical value one wishes to transform.

The following equation represents logistic regression[24]:

$$Y = \frac{e^{(b_0 + b_1 x)}}{1+e^{(b_0 + b_1 x)}} \quad (3)$$

Where, x = input value, y = predicted output, b0 = bias or intercept term, b1 = coefficient for input (x)

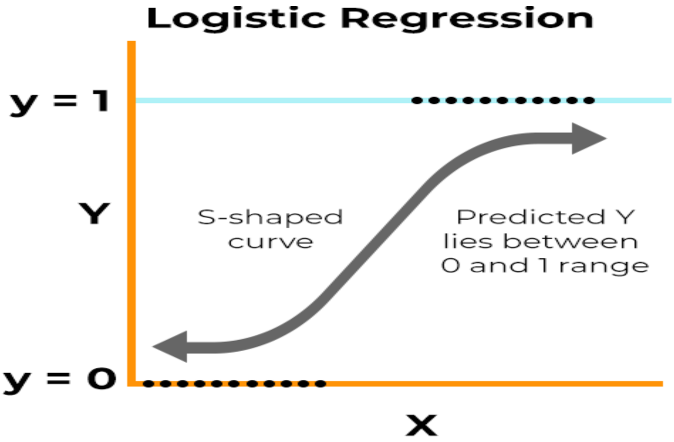


Figure 4. Represents the Logistic Regression[24].

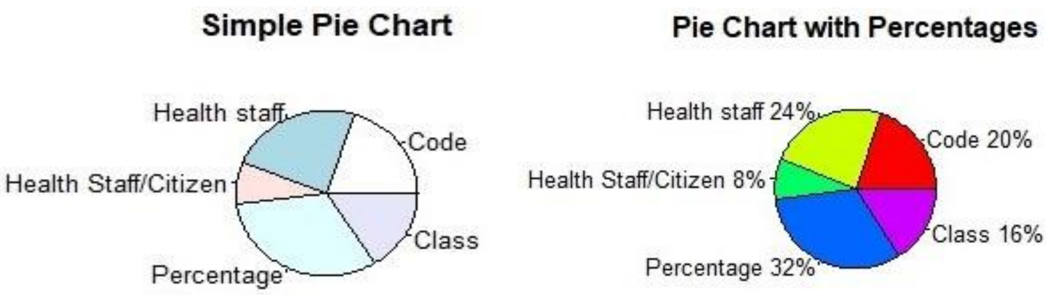
4.3 Euclidean Distance

Physical or monetary distances can be used to express distances in a GIS project. While the physical distance determines the straight-line or Euclidean distance, the cost distance determines the cost of going across that distance. In real-world settings, Understanding the variations between the two categories of distance metrics is essential for practical applications[25]. The distances and the input features' relationship can be calculated from the following equation:

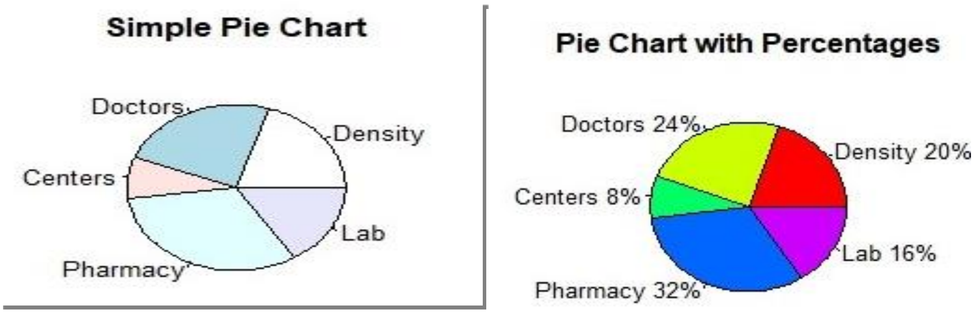
$$\text{Euclidean distance (Function)} \quad d(p_1, p_2) = \left(\sum_{i=1}^{v} (p_1(i) - p_2(i))^2\right)^{1/2} \qquad (4)$$

4.4 Pie charts

Pie charts are made using the function pie (x, labels), where x is a non-negative numeric vector representing the area of each slice and labels are a character vector representing the slice labels pi to the input data. seems to be especially helpful to visualize the consensus using a pie chart. The pie graphic and plain text are more effective when used across analysis party lines and have better recall[26].The graphs that resulted are as shown in Figure 5(a, b, c, d).



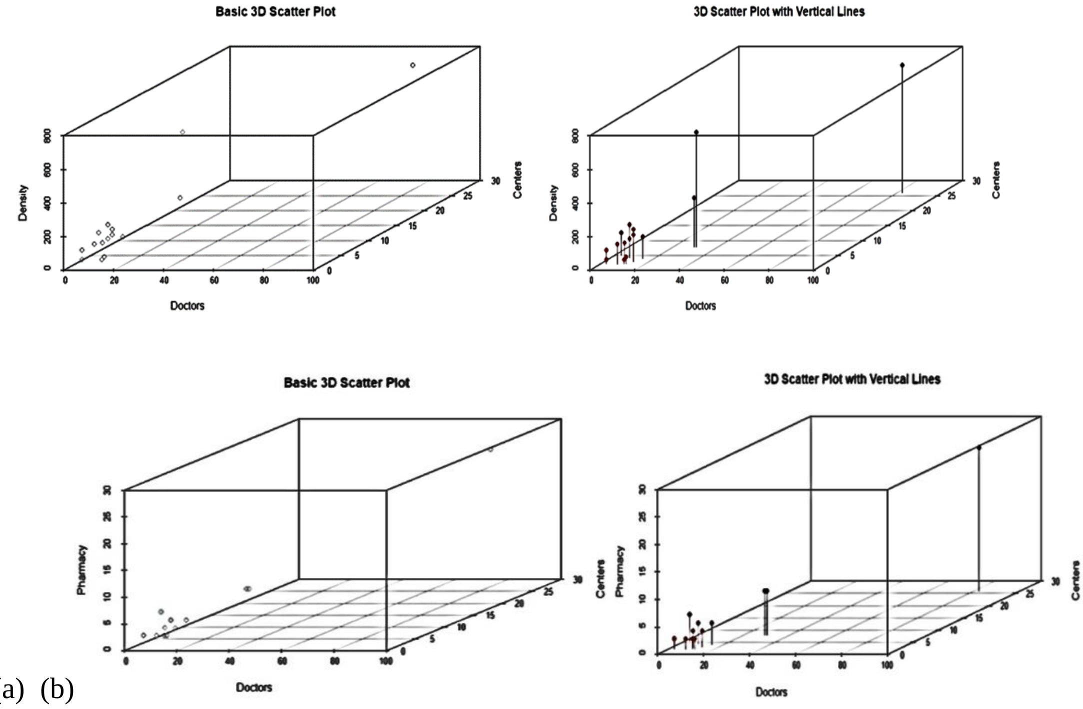
(a) (b)



(c) (d)

Figure 5(a, b, c, d). Pie chart applied to the datain this study**.**

Scatter plots and scatter plot matrices visualize the interaction of three quantitative variables at once. In this case, it uses a 3D scatter plot. The relationship between automobile mileage, service weight, and Density.



(a) (b)

(c) (d)

Figure 6. Principal component analysis 3D scatter plot of services per area: loading of 15 variables of health centers factors.

5. Results and Discussions

The primary benefit of using ML is that once an algorithm learns what to do with data, it can do so automatically. In this paper, it divides machine learning algorithm applications for medical services into five categories. Also, the use of Machine learning in this study and the alternative hypothesis, i.e., true correlation is equal to (95) as the highest estimate. It is highly challenging to compare health services between nations, particularly those that attempt to provide long-term care.

5.1 Description of the sample

The demographic compositions of the Full Analysis Samples are given in Table 1. these distributions were broadly similar to those of other large data sets of all Thi-Qar Area. although older children may have been slightly overrepresented in our sample 5 old to a child one-day old in addition to every person registered in the medical center in each district of the study areas. This gave better results.

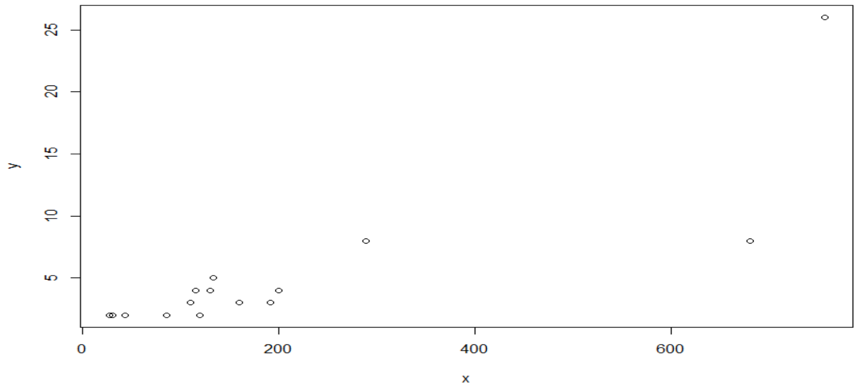


Figure7. Dot plot of Pearson's product analysis of the services of health centers

To define a factor describing how the elements of x are organized, you can add a group option. If so, the options (gcolor) and (cex) determine the color and size of the group's label, respectively.as in figure (7). Pearson's product-moment correlation data: x (No of Centers) and y (No of Medical services) x, y: numeric vectors with the same length, t = 5.7705, df = 13, p-value = 6.485e-05 alternative hypothesis: true correlation is not equal to 0 ,95 percent confidence interval: 0.9483387 0.5937424 sample estimates: correlation 0.8480646 cor. test (x, y). For a given independent variable, logistic regression models estimate the log chances of the outcome occurring vs. the log odds of the outcome not occurring (predictor variable).

In Figure 8. These log odds ratios should be appropriately presented in the summary of results as they are not probabilities but rather functions of the probabilities (p). Odds ratios, which are exponentiated versions of log odds ratios, are frequently provided in summaries.

Figure 8.The ordinary vs logistic regression

Scatter plot matrix of dependent and independent variables for the state's data, including linear and smoothed fits, and marginal distributions (kernel density plots and rug plots) By default.Correlogram showing the relationships between the various factors in the population rate frame and medical service data. Principal components analysis has been used to rearrange the rows and columns.

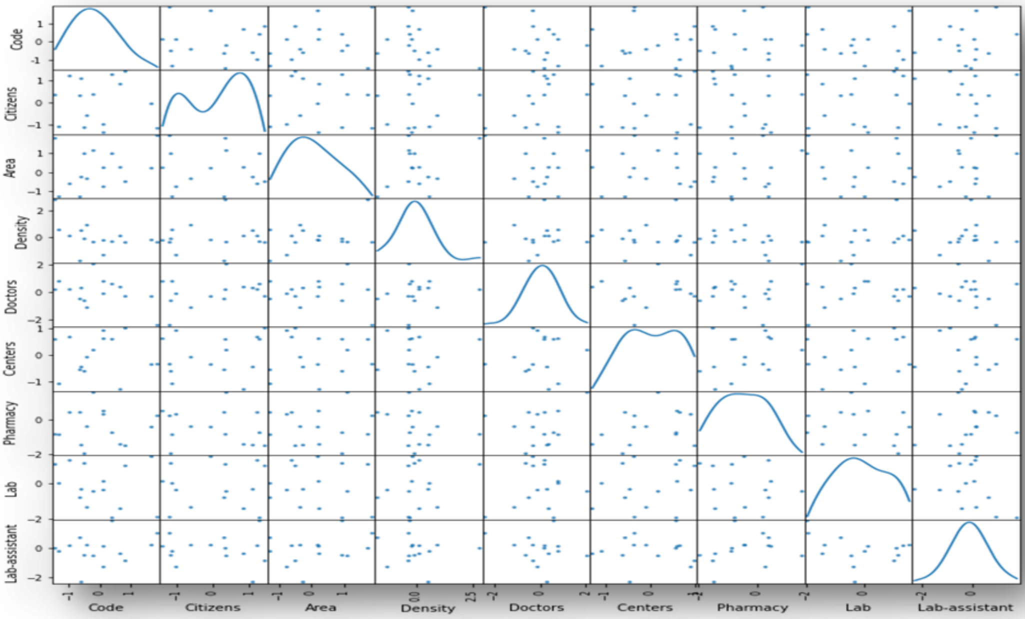


Figure 9. Scatter plot matrix of dependent and independent variables for the state's data.

You can see that the growth rate may be bimodal and that each of the predictor variables is skewed to some extent.

Table -2 Displays a sample of the data training

Code	Health.staff	Health.Staff.Citizen	Percentage	Class
1 AS6	37.2	0.000647	9.518	Fair
2 FO1	27.3	0.000669	9.841	Below Fair
3 BA10	25.4	0.000496	7.296	Below Fair
4 GB5	16.9	0.000278	4.089	Below Fair
5 DK7	19.7	0.000378	5.560	Below Fair
6 FJ15	36.5	0.000548	8.061	Above Fair

Medical service rates rise with population and submitted services, and fall with higher density growth levels and low-level medical labs. At the same time, strong governorates have lower service rates and populations and higher incomes. as shown in Figure 10.

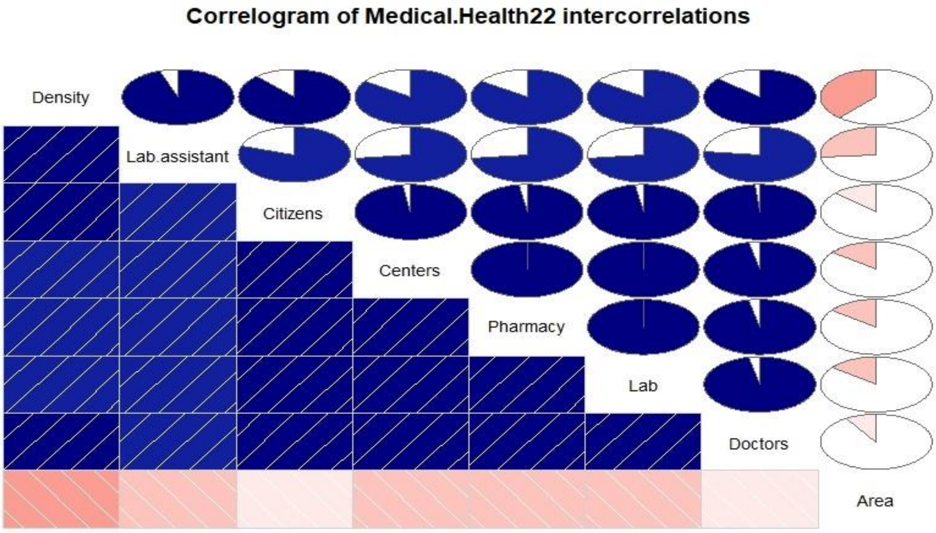


Figure 10. Relationships between the various factors in the population rate frame and medical service data.

The function scatterplot matrix in these packages provides univariate plots for each variable linear regression line and loess smoothed curves for each pair automatic labeling of noteworthy observations.The data can be displayed in a variety of ways using a corrgram (also known as a correlation diagram), Here the major objective is determined by panel functions to determine how relates to the other variables.

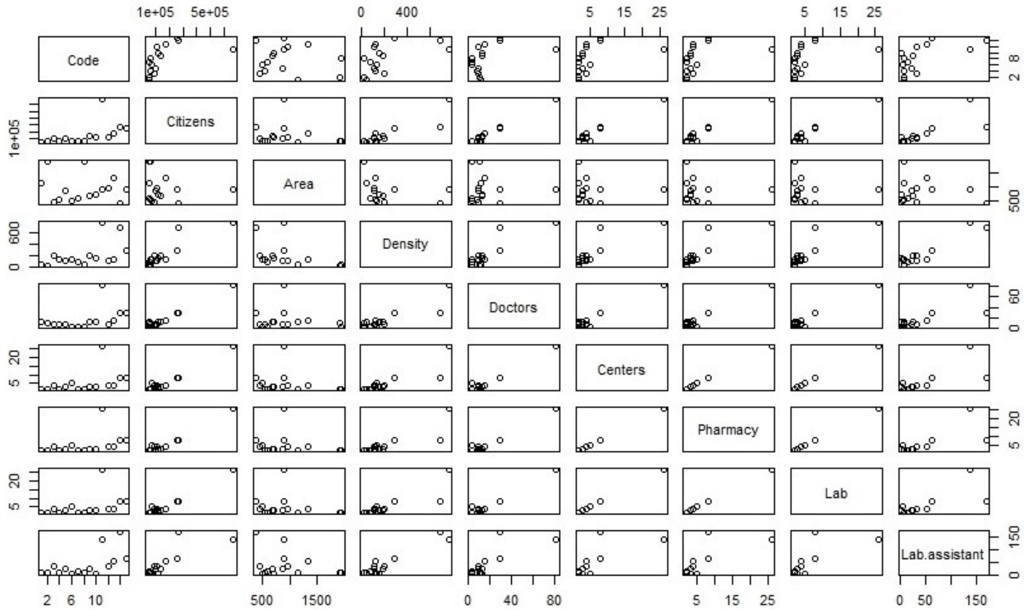


Figure 11. Scatter plot matrix of variables data in python language.

Multiple programming languages and algorithms were employed to create scatter plots of the evaluation and classification of services, and machine-learning techniques were utilized to compare the outcomes to those of the categorization step.as shown in Figure 12.

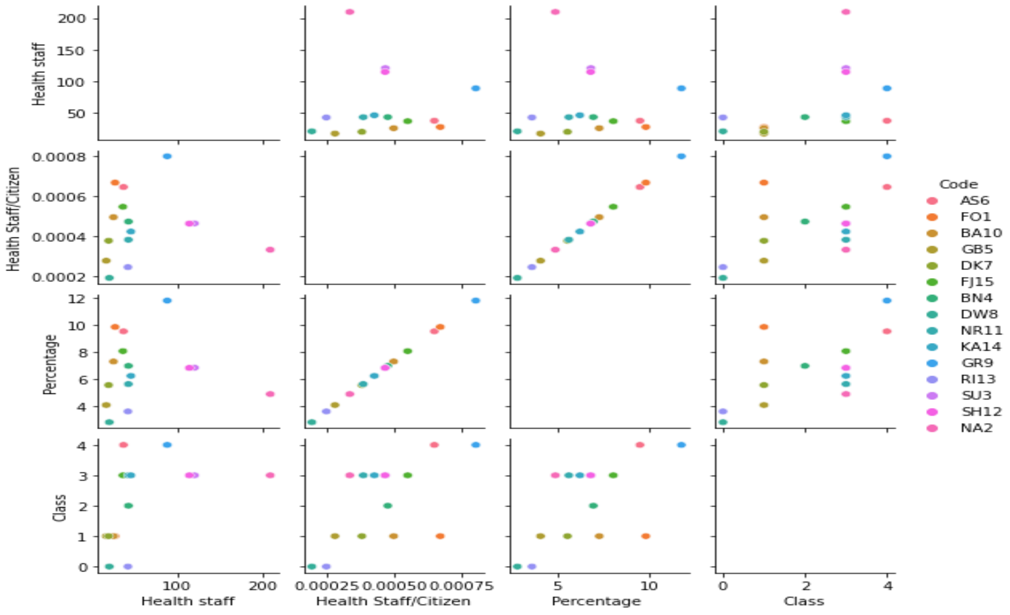


Figure 12. Scatter plot matrices of medical services data with other variables

Figure .13 Here is a chart that illustrates how the classification algorithm is applied to medical and Labs services.

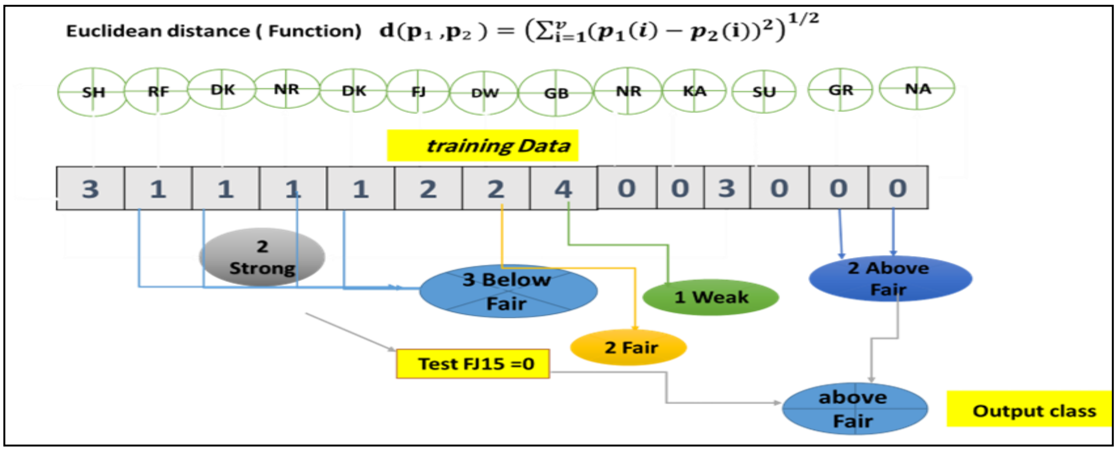


Figure 13. Classification of FJ15 based on tringting data by Regressionalgorithm

5.2 A typical strategy

Numerous techniques are available in R's basic installation for assessing the statistical presumptions used in a regression study. The most typical method is to use the plot() function on the object that the lm returns (). This results in four graphs that can be used to assess the model fit. Applied this strategy of basic linear regression .as in figure (14).To understand these graphs, A probability plot of the standardized residuals against the values that would be anticipated assuming normalcy is shown in the Normal Q-Q plot (top right). The points on this graph should lie on a straight 45-degree line if the normalcy assumption has been satisfied. Because they don't, you've clearly violated the normality assumption, the services of the health center number (15,14,1) are out of the line and need an increase in the type of services.

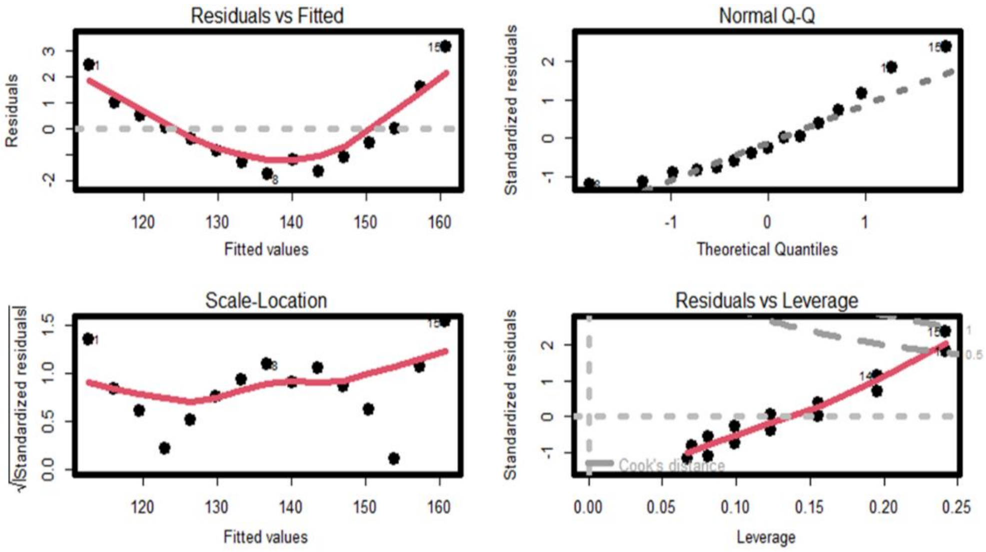


Figure 14. Diagnostic plots for the regression of services on density and center

5.3 The classicization of all samples

For training and testing, all samples were divided into two groups. At each step, dots reflect all the parameters that were inserted into the classification code to test five or three classes. In Figure 16. The distribution of all samples. which were split into two groups for training and testing overall 15 administrative divisions, is shown in Figure 15A&B lastly. Administrative units are first given a certain hue. The training sample color and the test sample color are then used to separate the samples. Figure 15C shows the three classes while figure 15D shows the classification into five categories within the limits of the classification, categorized as weak, strong, or medium. Figure (15C) shows the three classes.

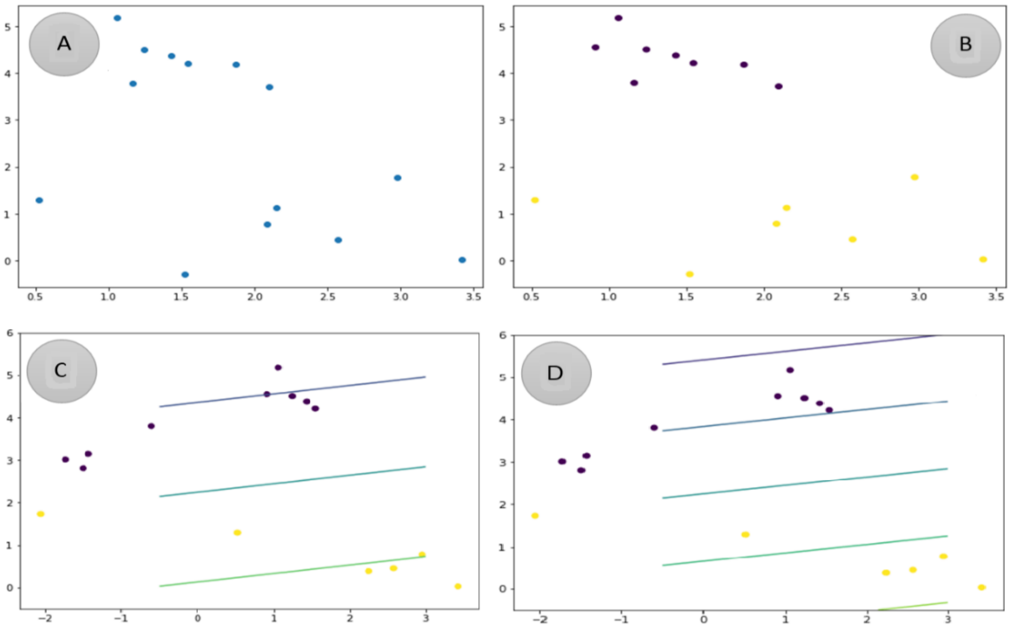
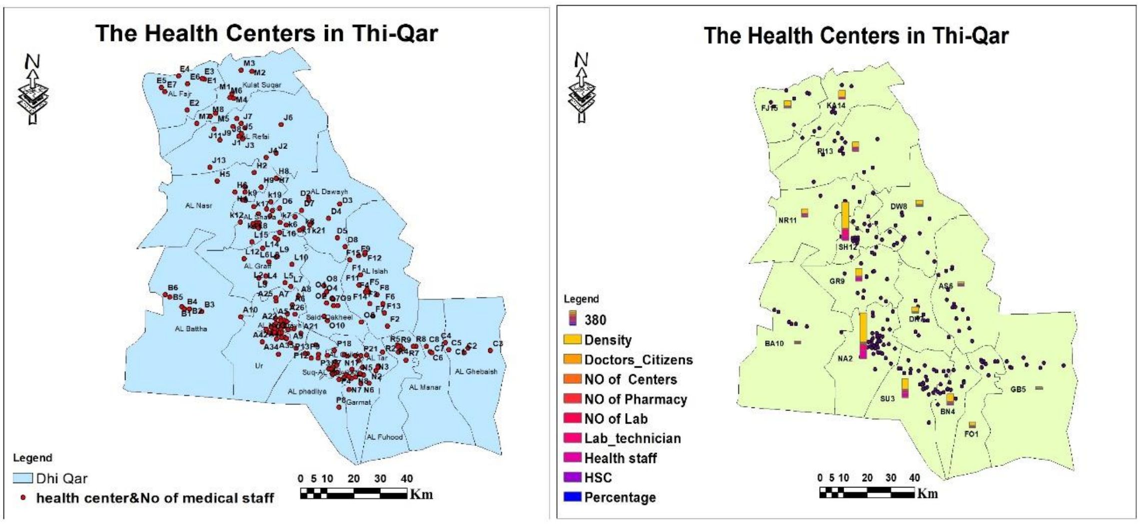


Figure 15. The Classification of the services of health centers a, b, c, d.

In the Geographic information system following what was stated above, we can represent the distribution of the 217 health centers over the governorate's administrative units using geographic mapping Figure 16(a), as well as the vertical ratios of each medical service that is offered in each health center Figure 16(b). Figure 16 (a), (b). In this study, we suggested a system for categorizing lab and medical services using GIS maps to data spatial.



(a) (b)

Figure 16(a, b). The distribution of the centers and services of health centers.

According to the maps' findings, creating categorization maps by combining population density measurements with a medical services ratio is an efficient method. Much research has employed strategies to handle services, however, these methods may not be ideal. With the aid of GIS and machine learning, this study provides the results are astounding and exactly the same, giving our study a confidence rating of 95%—a proportion that is exactly the same as the categorization of a test sample using training samples in the Python and R languages with GIS maps.Finally, the classification outcomes that were previously mentioned through classification algorithms can be summarized in a drawing that enables the reader to understand what was previously mentioned. as shown in Figure 17.

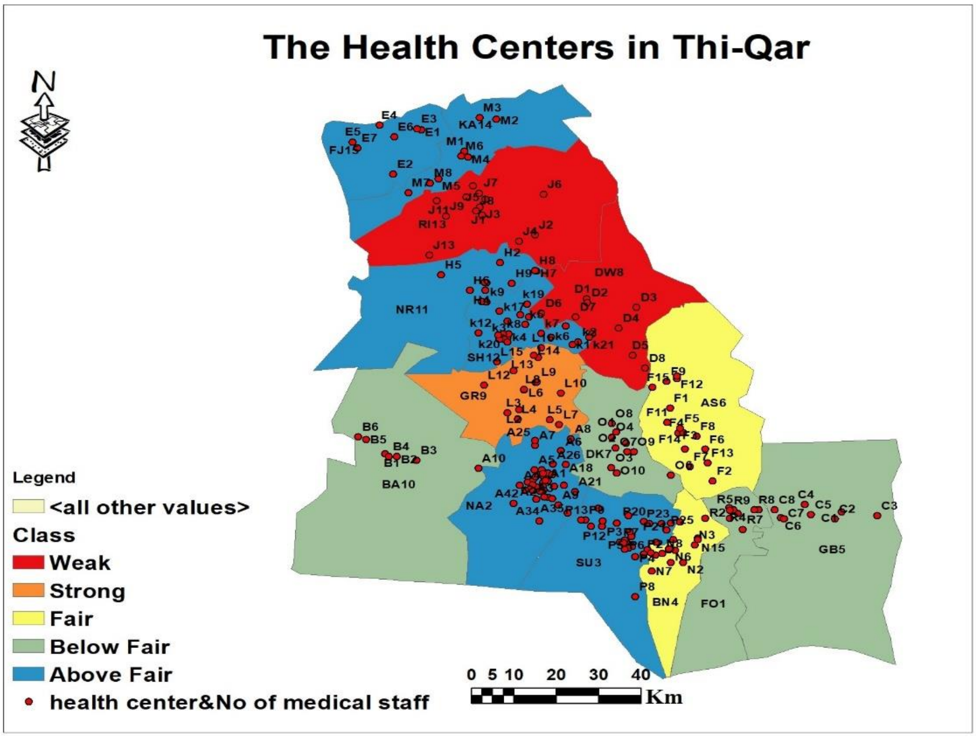


Figure 17. Represents the classification of the services of health centers

6. Conclusions

The categorization values were determined based on the population and the health centers' services. The results demonstrated that the major health centers were evenly distributed throughout the study region. There were a variety of concentrations because the classification of medical services affects how big the center's services are. Using the interpolation technique, the spatial analysis found a 2-9% ratio in health centers (5–26) as a robust medical service and 25–39% in centers (55–85) as weak and very weak. The classification of machine learning and some applied algorithms reveals that there are three abnormal units, either positive or negative districts, according to the data and statistics examined in this research; consequently, the ratios are close in terms of the classification of medical services provided to the population in primary health centers. The most important recommendations of the research are the need for health and medical centers to be improved in the area.

After, what concerns us first is the value p = 0.8808, which is greater than 0.05. Therefore, we conclude that there is no significant relationship between X and Y. R gives us a %95 confidence interval for the correlation coefficient, which is [0.258, 0.298] that is, with %95 confidence, the R of the population will be between - %25.8 and %29.8, where the R of the sample was 0.022, which is very small, which is the reason why it is not significant according to Pearson's product-moment correlation analysis. On the other hand, A way to plot many labeled values on a straightforward horizontal scale is to use dot plots. Using the syntax dot chart (x, labels=), where x is a numeric vector and the labels specify a vector that labels each point, you may generate them using the dot chart function as shown in Figures (7) & (8).

The scatterplot matrix (medical services) function provides scatter plots of the variables against each other in the off-diagonals and superimposes smoothed (loess) and linear fit lines on these plots. The principal diagonal contains the density and ratio plots for each variable as shown in Figure 9.

The function scatterplot matrix in these packages provides univariate plots for each variable linear regression line and loess smoothed curves for each pair, automatically labeling noteworthy observations. Visual summaries are frequently more informative than direct representations of the raw data for bigger data sets. as shown in Figure 11.

In figure 14.The closest affinity ratio or features were used to pick a test sample, training samples, and service comparison. It was then categorized using these criteria and used by many algorithms. The end result was the same as the Training samples.

This assumption appears. Finally, the Residual versus Leverage graph (bottom right) provides information on specific observations you may wish to pay attention to. Outliers, high-leverage points, and other anomalies are identified in the graph. An outlier is an observation that does not fit the fitted regression model well (that is, it has a large positive or negative residual). This includes health centers in (1,15,13). A high-leverage observation has an unusual combination of predictor values. In other words, it is an outlier in the predictor space. The value of the dependent variable is not used to calculate the leverage of an observation.

Linearity —There should be no systematic relationship between the residuals and the predicted (that is, fitted) values if the dependent variable is linearly related to the independent variables. In other words, the model should capture all systematic variance in the data while leaving only random noise. The Residuals versus Fitted graph (upper left) clearly shows no evidence of a curved relationship, implying that you should avoid including a quadratic term in the regression. There are issues with service distribution. If the constant variance assumption is met, the points in the Scale-Location graph (bottom left) should be a random band around a horizontal line. As show Figure (15).

In this study, we suggested a system for categorizing lab and medical services using GIS maps, such as in Figure 16 (a) and (b).According to the maps' findings, creating categorization maps by combining population density measurements with a medical services ratio is an efficient method. Much research has employed strategies to handle services. However, these methods may not be ideal. With the aid of GIS and machine learning, this study provides astounding and exactly the same results, giving our study a confidence rating of 95%. This proportion is exactly the same as categorizing a test sample using training samples in the Python and R languages with GIS maps. as shown in Figure (17).

Acknowledgement

Special thanks to the Department of Public Health Thi-Qar and the Head of Department Dr. Raheem Nayef and Mr. Jaffar Mohamed, health visit /GIS Department, Thi-Qar , for their unlimited help and support during our work.

References

[1] E. U. OCDE, *Health at a Glance: Europe 2020:State of in the EU Cycle*. 2020.

[2] K. V. Babu, S. Das, G. N. J. Sree, S. K. Patel, M. P. Saradhi, and M. R. N. Tagore, "Design and development of miniaturized MIMO antenna using parasitic elements and Machine learning (ML) technique for lower sub 6 GHz 5G applications," *AEU-International J. Electron. Commun.*, vol. 153, p. 154281, 2022.

[3] R. GeethaRamani and L. Balasubramanian, "Retinal blood vessel segmentation employing image processing and data mining techniques for computerized retinal image analysis," *Biocybern. Biomed. Eng.*, vol. 36, no. 1, pp. 102–118, 2016.

[4] P. Martin *et al.*, "Developing a case mix classification for child and adolescent mental health services: the influence of presenting problems, complexity factors and service providers on number of appointments*," *J. Ment. Heal.*, vol. 29, no. 4, pp. 431–438, 2020, doi: 10.1080/09638237.2017.1370631.

[5] V. P. Kanawade *et al.*, "Observations of new particle formation at two distinct Indian subcontinental urban locations," *Atmos. Environ.*, vol. 96, pp. 370–379, 2014.

[6] G. Piras, "sphet: Spatial models with heteroskedastic innovations in R," *J. Stat. Softw.*, vol. 35, no. 1, pp. 1–21, 2010, doi: 10.18637/jss.v035.i01.

[7] C. Perlich, "Tree Induction vs. Logistic Regression:," no. April 2002, 2014.

[8] C. P. Robert, G. Casella, and G. Casella, *Introducing monte carlo methods with r*, vol. 18. Springer, 2010.

[9] S. Chauhan, M. Sharma, and M. K. Arora, "Landslide susceptibility zonation of the Chamoli region, Garhwal Himalayas, using logistic regression model," *Landslides*, vol. 7, no. 4, pp. 411–423, 2010.

[10] J. M. Landwehr, D. Pregibon, and A. C. Shoemaker, "Graphical methods for assessing logistic regression models," *J. Am. Stat. Assoc.*, vol. 79, no. 385, pp. 61–71, 1984.

[11] M. C. Belavagi and B. Muniyal, "Performance evaluation of supervised machine learning algorithms for intrusion detection," *Procedia Comput. Sci.*, vol. 89, pp. 117–123, 2016.

[12] L. Salvador-Carulla, M. Poole, J. L. Gonzalez-Caballero, C. Romero, J. A. Salinas, and C. M. Lagares-Franco, "Development and usefulness of an instrument for the standard description and comparison of services for disabilities (DESDE)," *Acta Psychiatr. Scand.*, vol. 114, no. SUPPL. 432,

pp. 19–28, 2006, doi: 10.1111/j.1600-0447.2006.00916.x.

[13] F. Moya Rueda, R. Grzeszick, G. A. Fink, S. Feldhorst, and M. Ten Hompel, "Convolutional neural networks for human activity recognition using body-worn sensors," in *Informatics*, 2018, vol. 5, no. 2, p. 26.

[14] P. M. Spacks, *Desire and Truth: Functions of Plot in Eighteenth-Century English Novels*. University of Chicago Press, 1990.

[15] R. I. Kabacoff, *Data analysis and graphics with R - bonus chapter 23*. 2011.

[16] S. M. Lesch and D. L. Corwin, "Prediction of spatial soil property information from ancillary sensor data using ordinary linear regression: Model derivations, residual assumptions and model validation tests," *Geoderma*, vol. 148, no. 2, pp. 130–140, 2008.

[17] S. Al-Sultan, M. M. Al-Doori, A. H. Al-Bayatti, and H. Zedan, "A comprehensive survey on vehicular ad hoc network," *J. Netw. Comput. Appl.*, vol. 37, pp. 380–392, 2014.

[18] A. D. Albasi and M. Wells, "Introduction to Mazeinda," 2022.

[19] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*, Springer, 2009, pp. 1–4.

[20] Arseni, Voiculescu, Georgescu, Iticescu, and Rosu, "Testing Different Interpolation Methods Based on Single Beam Echosounder River Surveying. Case Study: Siret River," *ISPRS Int. J. Geo-Information*, vol. 8, no. 11, p. 507, 2019, doi: 10.3390/ijgi8110507.

[21] R. Xu *et al.*, "Altered gut microbiota and mucosal immunity in patients with schizophrenia," *Brain. Behav. Immun.*, vol. 85, pp. 120–127, 2020.

[22] M. Learning and R. Cookbook, "Ministry of Health, Iraqi Health Centers Guide 2019."

[23] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic regression*. Springer, 2002.

[24] V. Kanade, "What Is Logistic Regression? Equation, Assumptions, Types, and Best Practices," 2022.

[25] C. F. Thomas, *An Introduction to Geographic Information Systems*. 2020.

[26] S. Menard, *Applied logistic regression analysis*, no. 106. Sage, 2002.