Original Article

Diagnosis Of Skin Cancer By Using Fuzzy-Ann Expert System With Unification Of Improved Gini Index Random Forest-Based Feature

Rahul Boadh[1]*, Anil Yadav[2], Avneesh Kumar[3], Yogendra Kumar Rajoria[4]

[1*, 4]Department of Mathematics, School of Basic and Applied Sciences, K. R. Mangalam University, Gurgaon, India-122103
[2]Department of Chemical Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonipat, Haryana, India – 131027
[3]Department of mathematics, Kalindi College, University of Delhi

***Corresponding Author:** Rahul Boadh

*(Assistant Professor), Email ID: rboadh.iitkgp@gmail.com
DOI: 10.47750/pnr.2023.14.S02.175

Abstract

In addition to being one of the most widespread and lethal diseases in the world, skin cancer is also one of the most common types of cancer. However, due to its complexity and fuzzy nature, the clinical diagnosis process of any disease, including skin cancer, prostate cancer, coronary artery disorders, diabetes, and COVID-19, is frequently accompanied by doubt. In order to address the uncertainty and ambiguity surrounding the diagnosis of skin cancer as well as the heavier burden on the overlay of the network nodes of the fuzzy neural network system that frequently occurs due to insignificant features that are used to predict or diagnose the disease, a fuzzy neural network expert system with an improved Gini index random forest-based feature importance measure algorithm was proposed in this work. A Greater Gini Index Out of the 30 features in the dataset, the five most fitting features of the diagnostic Wisconsin breast cancer database were chosen using a random forest-based feature importance measure algorithm. Two sets of classification models were created using the logistic regression, support vector machine, k-nearest neighbour, random forest, and Gaussian naive Bayes learning algorithms. As a result, models for classification that included all features (30) and models that only used the top five features were used. The efficacy of the two sets of categorization models was assessed, and the results of the assessment were compared. The comparison's results show that the models with the fittest features outperformed those with the most complete features in terms of accuracy, sensitivity, and sensitivity. A fuzzy neural network-based expert system was therefore developed, utilising the five best features, and it achieved 99.83 percent accuracy, 99.86 percent sensitivity, and 99.64 percent specificity. The system built in this study also stands to be the best in terms of accuracy, sensitivity, and specificity when compared to prior research that used fuzzy neural networks or other applicable artificial intelligence techniques on the same dataset for the diagnosis of skin cancer. The z-test was also performed, and the test result demonstrates that the system has significantly improved accuracy for early skin cancer diagnosis.

Key words: Fuzzy-ANN Expert system, Skin Cancer, Gini Index, z-test, Forest-based Feature

1. INTRODUCTION

A category of disorders known as cancer is defined by the unchecked spread and expansion of aberrant cells [1, 2]. As a result, if the growth and spread of aberrant cells are not stopped, it could result in death. Skin disease for the most part happens on the external layer of the skin, which may initially show up as an enlarging, swell, or different part of the skin [3]. The skin blocks, heat, daylight, wounds, and contaminations; it additionally warms the body and stores fat water in the body and creates vitamin D. The skin possesses 2 chief layers: the shallow layer of the skin (epidermis) and the inward skin (dermis) [4]. The skin's surface layer (epidermis) is fundamentally made of level, flaky cells named Squamous cells. Round cells, called basal cells, are situated underneath the shallow cells of the squamous melanocyte in the surface layer [5]. Delayed openness to daylight upgrades melanoma skin disease's gamble over a lifetime. One of the riskiest sorts of disease is Melanoma, which is brought about by harm brought about by overexposure to the sun and a few different elements. Melanoma can be determined to have skin biopsies. The factual data of the lead tumors in amount and passing worth in 2019 [6]. If melanoma is analyzed and treated early, it very well may be dealt with, and late analysis can prompt the patient's passing.

Just experienced doctors can analyse melanoma on time utilizing suitable apparatuses and histological reports. One of the gadgets used to analyze melanoma is meatoscopic. With this device, changes in the pigmentation of skin sores in infections can be assessed. With the advancement of science as of late, computerized meatoscopic has been supplanted by ordinary dreamscapes with the capacity to catch and store skin pictures. With the improvement of computerized dreamscapes, the improvement of a calculation for the determination of melanoma was considered by specialists, and then, at that point, different techniques were proposed to upgrade the exactness of skin injury conclusion. Many researchers worked on different problems by using fuzzy expert system and Fuzzy-ANN modelling techniques [7-26]. This work plans to foster a fuzzy NN-based expert for early skin malignant growth conclusion in India that would address the equivocalness and vulnerability encompassing the sickness' demonstrative cycle as well as the bigger weight forced on the fuzzy-NN based framework's overlay of the organization hubs, which regularly emerges because of unimportant elements used to estimate or analyze the disease.

2. DATA & METHODOLOGY

This section discusses the development and implementation process for an enhanced Gini index random forest-based feature importance measure algorithm with fuzzy neural network expert system for the early detection of skin cancer in India.

2.1 Data

The involves gathering, pre-processing, and cleaning the pertinent dataset for the issue domain. As a result, with the aid and direction of human specialists in the problem domain to be modelled, only pertinent data would be gathered. In this stage, missing data are also handled, and redundant and derived dataset features are also removed. It is hardly ever the case when developing a system or model in the real world that not all the features in the dataset are helpful [27]. Reducing a model's capacity to generalise and potentially its overall accuracy are two effects of adding redundant variables [28-31]. Additionally, the system or models become more complex overall when additional variables are added to them [32, 33].

We employed an enhanced random forest Gini index algorithm developed by [34] as a feature selection method that scores input characteristics according to their value and significance for forecasting a target feature of the dataset. One of the well-liked machine learning techniques for processing large amounts of data and creating models is the random forest algorithm.

2.2 Methodology

The Gini index algorithm was based on Gini theory. The basic idea of Gini index theory is that considering X is the set of samples x and these samples have n classes (Y_i, i= 1…n). The equation can be obtained

$$Gini(X) = 1 - \sum_{i=1}^{m} P_i^2$$

Where P is the probability i.e. sample space divided by favourable events. In this study, for normalize the probability and produce the unbiased Gini values used the IGini-C equation from previous study done by different cancer over Saudi Arabia [33]. The equation is as follow:

$$IGini_C(W) = \left|\frac{1}{log_2 P(W/C_i)^2}\right| \sum_{i=1}^{m} P(C_i/W)^2.$$

2.3 Performance Evaluation of Classification Models with Comparative Analysis

Based on accuracy, specificity, and sensitivity, the classification models created with the five fittest features chosen using an upgraded random forest Gini index technique would be compared to the classification models created with all features (shown in table 1).

Table 1: Confusion matric of predicted values

	Positive 1	**Negative 0**
Positive 1	True Positive	False Positive
Negative 0	False Negative	True Negative

Where Accuracy calculated by $\frac{\text{True Positive + True Negative}}{\text{True Positive+ False Positive+ False Negative+ True Negative}}$

Specificity calculated by $\frac{\text{True Positive}}{\text{True Positive+ False Negative}}$ and

Sensitivity by $\frac{\text{True Positive}}{\text{True Positive+ False Positive}}$

2.4 Rule generation and Optimization

The rules are the core component of the fuzzy neural network-based expert system's knowledge base. A dynamic expert system that may be described by n characteristics, such as y_1, y_2,….y_n, are represented by a training set of n inputs and a single output. The neural network technique being used to redefine the fuzzy rules that were derived from the chosen training dataset. The fuzzy neural network-based expert system would therefore use five network neural layers, with each layer specifying the phases of the inference mechanism. The inference method of the fuzzy-NN based expert system is shown in Fig.1 with five network neural layers. The five nodes/layers such as Input, Fuzzification, Fuzzy Rules, Inference and Defuzzification are used in this study (Fig. 1).

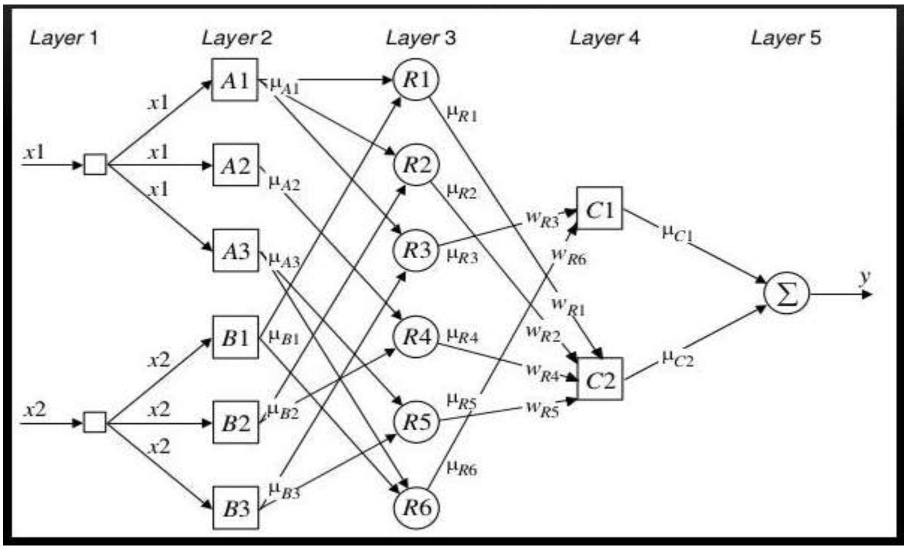

Fig. 1: Five layers of Fuzzy-ANN based expert inference system (Taken from Google)

Rule Generation from Dataset

In order to prevent over- and under-flitting of the system, the dataset was divided into 80% training and 20% testing sets. The rules that would be used for the diagnosis of breast cancer were created using the dataset. The rules (Fig. 2) in this work were produced using the MALTAB14b neuro-fuzzy designer toolbox.

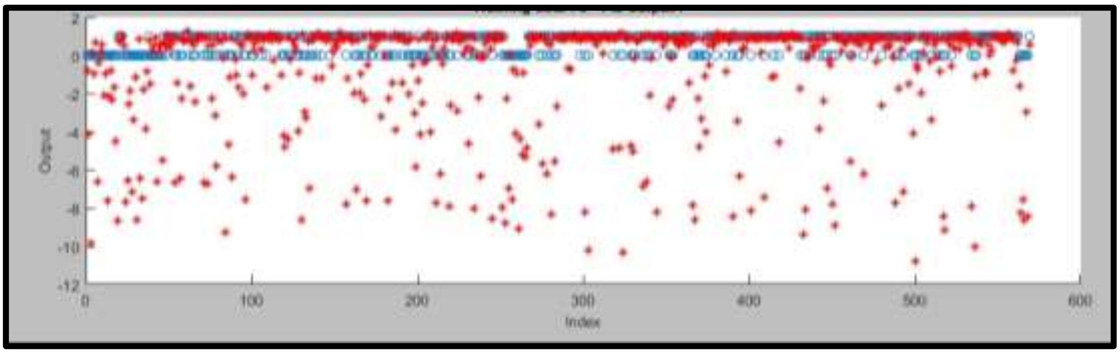

Fig.2: Movement of Training data set
*(Training data represent by 0 and Fuzzy inference system output represent by *)*

3. RESULTS AND DISCUSSION

In this section the input parameters of fuzzy-NN expert system and risk of skin cancer in the form of output is discussed. The input and output feature ranges, together with the system's membership function and linguistic terms, are displayed in Table 2.

Table 2: The Range of Input and Output membership functions

Input (Features)	Range	Linguistic Term	Membership Function
RadiusMean	<10 7–28 >21	Smaller Normal Larger	Trapezoid
TextureMean	<15 10–30 >25	Smaller Normal Larger	Trapezoid
PerimeterMean	<40 30–180 >160	Smaller Normal Larger	Trapezoid
AreaMean	<150 130–2200 >2300	Smaller Normal Larger	Trapezoid
SmoothnessMean	<0.04 0.06–0.14 >0.12	Smaller Normal Larger	Trapezoid
Diagnosis	<1 1–2 >2	Mild Healthy Severe	Triangle

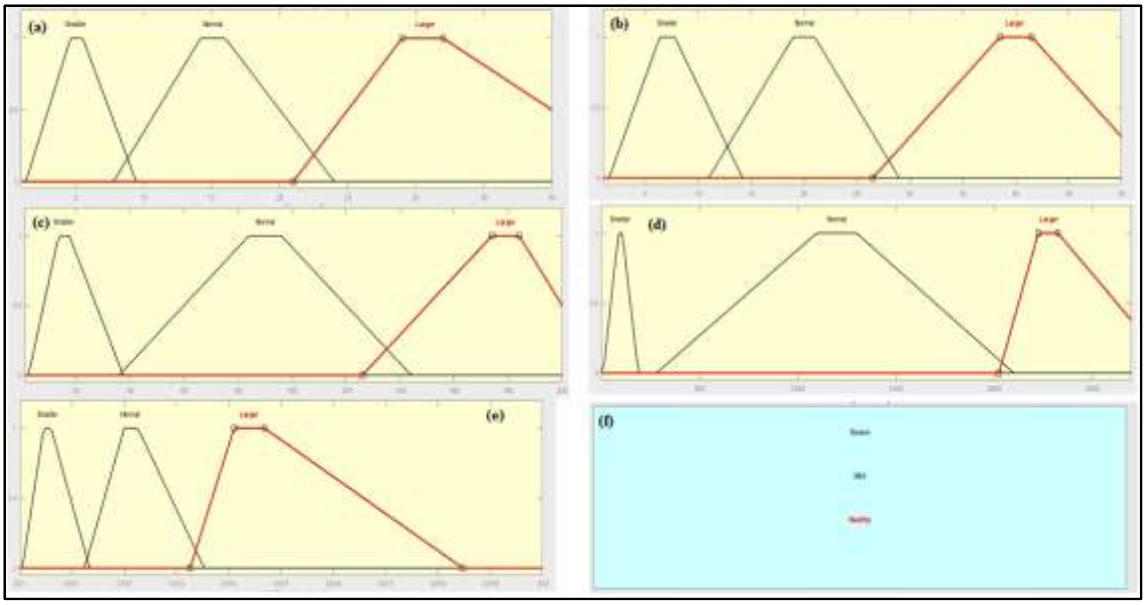

Fig. 3: Membership function of input parameters (a) Radios mean (b) Texture mean (c) Perimeter mean input (d) Area mean (e) Smoothness mean and (f) Diagnosis Skin Cancer.

The MATLAB14b Fuzzy Logic Toolbox, the distribution of the trapezoid membership function, which is used for the system's inputs and outputs, was determined. The system's input features for the linguistic variables of radius, texture, perimeter, area, and smoothness mean are each represented visually by a membership function (Figs. 3a to 3e), while the output features are represented visually by Fig. 3f. The total seventeen rules are generated for output.

The defuzzification nodes shown in the Fig. 4, which takes input from the output fuzzy rules and merges them into a single fuzzy rule. As a result, all the outputs that were Fuzzified and displayed in the figure are defuzzied into their corresponding crisp values. To optimise the tuning of the membership function parameters, back propagation learning is performed in rule optimization. In order to change the membership functions of the neurons in layer 2, the training data of input and output are compared with the system output and the mistakes are propagated backwards through the network from output layer to input nodes.

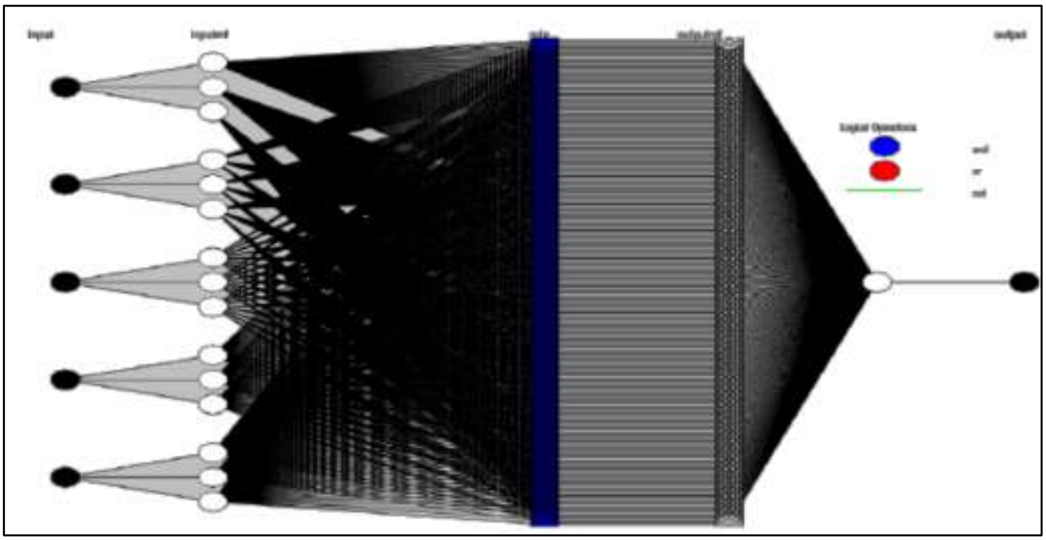

Fig. 4: Early Diagnosis of Skin Cancer by using Fuzzy-NN Expert System

Only the five features that are the best fits and have the highest random forest Gini based importance measure scores were taken into consideration when the system was used to predict skin cancer on 450 randomly chosen instances from the diverse laboratories skin cancer dataset. The radius mean, texture mean, perimeter mean, area mean, and smoothness mean are some of the characteristics. Figure 5 illustrates how the algorithm uses the healthy, mild, and severe linguistic factors to predict and diagnose skin cancer.

For the ability to accurately diagnose individuals with healthy skin, mild cases of skin cancer, and severe cases of skin cancer, the system got a 99.46% accuracy rating, while it achieved a 99.32% sensitivity rating. Additionally, the system scored 99.78% specificity for its capacity to accurately diagnose breast cancer patients in healthy situations.

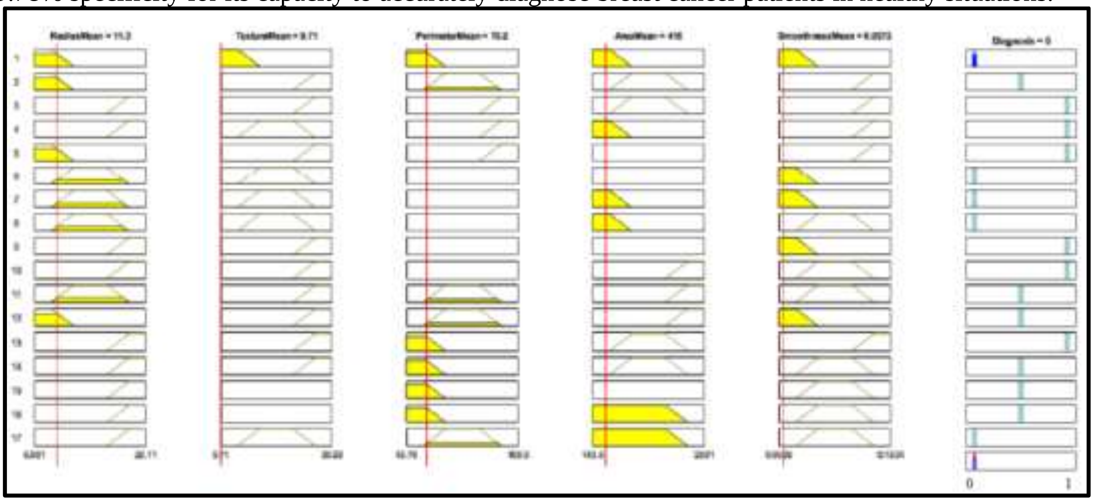

Fig. 5: Output of the Skin cancer at different level of input parameters

Performance Evaluation

The outcome of the performance evaluation obtained from the fuzzy neural network expert system in this work, with an improved Gini index random forest-based feature importance measure algorithm, and has been compared with previous works that used fuzzy neural network and other applied artificial intelligence techniques on the same dataset for the diagnosis of breast cancer using the same dataset instances. The suggested technique of this work stands to be the best based on comparisons with other works that have used the dataset [33] and the proposed method, which achieved 99.46%, 99.78%, and 99.32% for accuracy, specificity, and sensitivity, respectively.

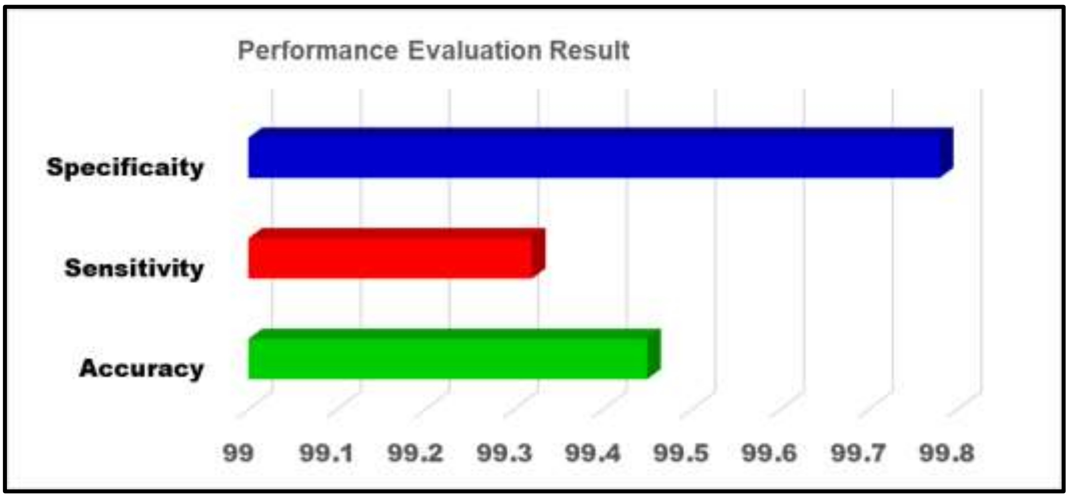

Fig.6: Performance of Fuzzy-NN modelling result

4. CONCLUSION

The current review, a fuzzy-NN expert system with a better Gini index random forest-based highlight significance measure calculation has been created to address the difficulties which remember the heavier weight for the overlay of the organization hubs of fluffy brain network-based master framework because of numerous immaterial elements that are utilized to anticipate or analysed the sickness and the vulnerability and vagueness frequently connected with symptomatic decision making of skin malignant growth. The fuzzy-NN expert system with a better Gini index irregular backwoods based highlight significance measure calculation accomplished 99.46% exactness for the capacity to accurately analysed solid, gentle, and extreme instances of bosom disease patients, while for the capacity to analysed gentle and serious instances of bosom malignant growth patients accurately, the framework accomplished 99.32% awareness and, similarly for the capacity to analysed sound cases accurately of bosom malignant growth patients, the framework accomplished 99.78% explicitness.

Besides, in view of the examination of the framework with past works that utilized fuzzy-NN organization or other applied computerized reasoning procedures on the equivalent dataset Fundamental Twist Progressed Rework for anticipating skin disease, the framework stands to be awesome with the accomplishment of 99.46%, 99.32%, and 99.78% for exactness, awareness, and particularity, separately. Z-test measurable testing was too. Directed, and the testing result shows that there is huge precision accomplished by the framework for analysis of bosom disease. This study advocates that, the technique utilized in this work would be utilized in a few different spaces to additionally learn its exactness.

REFERENCES

1. Song, H. Detectability of Breast Tumors in Excised Breast Tissues of Total Mastectomy by IR-UWB-Radar-Based Breast Cancer Detector. IEEE Trans. Biomed. Eng. 2019, 66, 2296–2305.
2. Alharthi, H. Healthcare predictive analytics: An overview with a focus on Saudi Arabia. J. Infect. Public Health 2018, 11, 749–756.
3. R. Navid, "Computer-aided diagnosis of skin cancer: a review," Current Medical Imaging, vol. 16, no. 7, pp. 781–793, 2020.
4. S. Ahuja, B. K. Panigrahi, N. Dey, V. Rajinikanth, and T. K. Gandhi, "Deep transfer learning-based automated detection of COVID-19 from lung CT scan slices," Applied Intelligence, vol. 51, no. 1, pp. 571–585, 2021.
5. Q. Liu, Z. Liu, S. Yong, K. Jia, and N. Razmjooy, "Computer aided breast cancer diagnosis based on image segmentation and interval analysis," Automatika, vol. 61, no. 3, pp. 496–506, 2020.
6. A. Costa, Y. Kieffer, A. Scholer-Dahirel et al., "Fibroblast heterogeneity and immunosuppressive environment in human Breast cancer," Cancer Cell, vol. 33, no. 3, pp. 463–479, 2018.
7. Gautam, A., Umesh Kumar, Anil Yadav, Rahul Boadh, Manjeet Aggarwal, R. K. Khandal. (2022). Nutritional Aspects of Ready-to-Eat and Homemade Food Products with Emphasis on Fatty Acid Profiling of Ready-to-Eat Food using Gc-Fid Technique- A Comparative Study. Journal of Pharmaceutical Negative Results, Vol. 13(08), pp. 2825–2846. https://doi.org/10.47750/pnr.2022.13.S08.352
8. Kalra, G., Kishor,K., Kumar, A., Rajoria, Y. K., Yadav, A., and Boadh, R. (2022a). FIS-Based Prediction and Estimation of Health Insurance for Workers in the Manufacturing Sector, Journal of Pharmaceutical Negative Results. Vol. 13 (S8), pp. 2236-2245. DOI: 10.47750/pnr.2022.13.S08.33118.
9. Bhagat, R. K., Yadav, A., Rajoria, Y. K., Raj, S. and Boadh, (2022). Study of Fuzzy and Artificial Neural Network (ANN) BasedTechniques to Diagnose Heart Disease, Journal of Pharmaceutical Negative Results. Vol. 13 (S5), pp. 1023-1029. Doi: 10.47750/pnr.2022.13. S05.161.19.
10. Poswal, P., Chauhan, A., Aarya, D. D., Boadh, R., Rajoria, Y. K. and Gaiola, S. U. (2022a). Optimal Strategy for Remanufacturing System of Sustainable Products with Trade credit under Uncertain Scenario, Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2022.08.303.20.
11. Boadh, R., Yadav, S. N., Tiwari, A., Rajoria, Y. K. and Singh, J. (2022a). Application of fuzzy inference system (FIS) for assessment and predication of compressive asset of concrete containing fly ash, Materials Today: Proceedings. https://doi.org/10.1016/j.matpr.2022.08.160.21.
12. Poswal P., Chauhan A, Boadh R and Rajoria Y.K (2022b). A review on fuzzy economic order quantity model under shortage, AIP Proceeding, 2481, 040023(1-13) 10.1063/5.0103757.22.
13. Karla, G., Boadh, R., Rajoria, Y. K. and Rajendra, P. (2022b). Prediction and Analysis of Health Risk by using Fuzzy Control System, International Journal of Health Sciences (IJHS). Vol. 6(S2), pp. 7511–7524. https://doi.org/10.53730/ijhs.v6nS2.6808.23.
14. Poswal, P., Chauhan, A., Rajoria, Y. K., Boadh, R. and Goel, A. (2022d). Fuzzy optimization model of two parameter weibull deteriorating rate with quadratic demand and variable holding cost under allowable shortages, Yugoslav Journal of Operations Research, Vol.32 (4).25.
15. Boadh, R. Aarya, D. D., Dahiya, M., Rathee, S., Kumar, A., Jain, S., Rajoria, Y. R., and Rani, S. (2022b). Study and prediction of prostate cancer using fuzzy inference system, Materials Today: Proceedings, vol. 56 (1), pp 157-164. doi:https://doi.org/10.1016/j.matpr.2022.01.040.26.

16. Boadh, R. Grover,R., Dahiya, M., Kumar, A., Rathee, R., Rajoria, Y. R., Rawat, M., and Rani, S. (2022c) Study of fuzzy expert system for the diagnosis of various types of cancer, Materials Today: Proceedings, vol. 56 (1), pp 298-307. doi: https://doi.org/10.1016/j.matpr.2022.01.16127.
17. Boadh, R., Chaudhary, K., Dahiya, M., Dogra, N., Rathee, S., Kumar, A., Rajoria, Y. K. (2022d). Analysis and investigation of fuzzy expert system for predicting the child anaemia, Materials Today: Proceedings, vol. 56 (1), pp 231-236. doi:https://doi.org/10.1016/jmatpr. 2022.01.094.28.
18. Karla, G., Rajoria, Y. K.. Boadh, R., Rajendra, P., Khatak, N., and Kumar, A. (2022c). Study of fuzzy expert systems towards prediction and detection of fraud case in health care insurance, Materials Today: Proceedings, vol. 56(1), pp477-480. doi: https://doi.org/10.1016/j.matpr. 2022.02.157.29.
19. Poswal, P., Chouhan, A., Boadh, R., Rajoria, Y. K., Kumar, A. and Khatak, N. (2022e). Investigation and analysis of fuzzy EOQ model for price sensitive and stock dependent demand under shortages, Materials Today: Proceedings, vol. 56 (1), pp 542-548 doi:https://doi.org/10.1016/j.matpr.2022.02.273.30.
20. Aarya, D., Rajoria, Y. K., Gupta, N., Singh.Y, & Rathee, R, Boadh, R. and Kumar, A. (2022). Selling price, time dependent demand and variable holding cost inventory model with two storage facilities. Materials Today: Proceedings. 56(1), pp. 245-251, doi: 10.1016/j.matpr.2022.01.11138.
21. Bhagat, R. K., Yadav, A., Rajoria, Y. K., Raj, S. and R. Boadh, (2022). Study of Fuzzy and Artificial Neural Network (ANN) Based Techniques to Diagnose Heart Disease, Journal of Pharmaceutical Negative Results. Vol. 13 (S5), pp. 1023-1029. Doi: 10.47750/pnr.2022.13. S05.161.40.
22. Kalra, G., Kishore, K., Kumar, A., Rajoria, Y. K., Yadav, A., Boadh, R. (2022). FIS-Based Prediction and Estimation Of Health Insurance For Workers In The Manufacturing Sector. Journal of Pharmaceutical Negative Results, Vol. 13(S08), pp. 2636–2645. https://doi.org/10.47750/pnr.2022.13.S08.33143.
23. Tiwari, A., Rajoria, Y. K,, Boadh, R. (2022). Comparative Study of Maximization Assignment Model by Existing Method and NewlyProposed Methods. International journal of Operational Research, (in press). DOI: 10.1504/IJOR.2022.1004664445.
24. Kaur, K., Bhardwaj, H. K., Rajoria, Y.K., Yadav, A., Maurya, D. K., Boadh, R. (2022). Fuzzy Expert System (FES) Integration for Oral Cancer Risk (OCR) Diagnosis. Journal of Pharmaceutical Negative Results, Vol. 13 (08), pp. 3433–3437. https://doi.org/10.47750/pnr. 2022.13.S08.42146.
25. Kumar, S., Kalra, G., Bhardwaj, H. K., Rajoria, Y.K., Kumar, D., Boadh, R. (2022). Internet of Medical Thing and FIS Evaluation for Selecting and delivering the best Health Insurance Coverage. Journal of Pharmaceutical Negative Results, Vol.13(08), pp. 3438–3446. https://doi.org/10.47750/pnr.2022.13.S08.42247.
26. Kaur, K., Jai Shankar Prasad, Deo Datta Aarya, Hare Krishna, Vishal B. Tambe, Yogendra Kumar Rajoria, Rahul Boadh. (2022). Implementation of an Adaptive Artificial Neural Network with Fuzzy Expert System for Diagnoses the Breast and Prostate Cancer: A Hybrid Technique. Journal of Pharmaceutical Negative Results, Vol. 13 (08), pp. 3806–3812. https://doi.org/10.47750/pnr.2022.13.S08.475
27. Nadia, G.; Bilal, R.; Ebrahem, A.A. The dynamics of fractional order Hepatitis B virus model with asymptomatic carriers. Alex. Eng. J. 2021, 60, 3945–3955.
28. Han, H.; Guo, X.; Yu, H. Variable selection using Mean Decrease Accuracy and Mean Decrease Gini based on Random Forest. In Proceedings of the 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 26–28 August 2016; pp. 219–224.
29. Aldrich, C. Process Variable Importance Analysis by Use of Random Forests in a Shapley Regression Framework. Minerals 2020, 10, 420.
30. Ebrahem, A.A.; Din, R. On global dynamics of COVID-19 by using SQIR type model under non-linear saturated incidence rate. Alex. Eng. J. 2021, 60, 393–399.
31. Kanagarathinam, K.; Algehyne, E.A.; Sekar, K. Analysis of 'earlyR'epidemic model and time series model for prediction of COVID-19 registered cases. Mater. Today Proc. 2020, 10, 2214–7853.
32. Sitaula, C.; Aryal, S. Fusion of whole and part features for the classification of histopathological image of breast tissue. Health Inf. Sci. Syst. 2020, 8, 38.
33. Ebrahem A. A., Muhammad, L. J., Naseh, A. A., Osama, A. A., and Abdullah, K. A. (2022). Fuzzy Neural Network Expert System with an Improved Gini Index Random Forest-Based Feature Importance Measure Algorithm for Early Diagnosis of Breast Cancer in Saudi Arabia. Big Data Cogn. Comput. Vol 6 (13), pp. 1-29. https://doi.org/10.3390/bdcc6010013.
34. Park, H.; Kwon, H. Improved Gini-Index Algorithm to Correct Feature-Selection Bias in Text Classification. IEICE Trans. Inf. Syst. 2011, 94, 855–865.