SvedbergOpen
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

# Stratifying Success Of In-Vitro Fertilization Using Chi-Kruskal's Hybrid Feature Extractor And Machine Learning Classifiers

S. AlishMonica[1*], Dr. T. Kamalakannan[2]

[1*]Research Scholar, Department of Computer Science, Vels Institute of Science, Technology & Advanced Studies, Chennai, India. E-mail: alishmonica@gmail.com, https://orcid.org/0009-0004-4336-3490

[2]Professor, Department of Applied Computing and Emerging Technologies, Vels Institute of Science, Technology & Advanced Studies, Chennai, India. E-mail: kkannan.scs@vistas.ac.in, https://orcid.org/0000-0002-5982-8983

*Corresponding author: Email: alishmonica@gmail.com

**Abstract**

Fertility problems have become more common in recent times, with many challenged with quotidian routines that can contrive infertility amongst couples. The pivotal causes to these detriments can be the unhealthy living style, high stress levels and less to no exercise. The realm of medical sciences is burgeoned extensively to render progressive results, nonetheless simulative and swifter analysis of various parameters that trigger the success and failure of the In-Vitro Fertilization (IVF) process remains to defy the demands. The currently established empirical studies delineate various statistical and computational implementation effectuated toward the prediction of IVF, but fails to elaborate on the correlation of attributes in relevance to augmenting classifier accuracy of machine learning models. The proposed study thereby effectuates a stepwise analysis from pre-processing the dataset to efficiently unsheathing the features of vitality, and further to incorporate an assortment of machine learning classifiers to explicitly comprehend the performance of the networks. This paper proposes a fused feature extractor using the Chi-Square and the Kruskal Wallis feature extraction methods to form a hybrid feature extractor. Furthermore, the machine learning classifiers such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), Logit Boosting, Gradient Boosting Machine (GBM), Decision Tree (DT) and an ensemble classifier are entailed to cognize the stratification efficacy of the IVF data. The process of hyperparameter tuning, along with classification indagation is simulated in MATLAB, and the results procured in each phase of implementation aids to render better medical decisions for clinicians in IVF treatment formulation, and customized healthcare provisions for patients.

**Keywords:** In-Vitro fertilization (IVF), Chi-Kruskal's hybrid Feature Extractor, Machine learning classifiers, Support Vector Machine (SVM), Random Forest, K-Nearest Neighbor (KNN), Logit Boosting, Gradient Boosting Machine (GBM), Decision Tree, Ensemble Classifier.

## 1. Introduction

A realm of medical progression that has proliferated in a colossal manner is reproductive procedures such as the In-Vitro fertilization process, that significantly auspices a ray of hope for sterile couples, while aiding them to overcome the challenges with procreation. The initiation of IVF through guided reproduction methods has been successful since the beginning of this century. Nonetheless, ensuring the fertilization of an egg and sperm in an unnatural environment exterior to the human body necessitates meticulous milieus interms of clinical and biological routines. Thedistinct phases for the entirety of the process to be completely successful relies on several key steps, beginning with the stimulation of the ovaries to produce multiple eggs. These eggs are then retrieved and combined with the sperm in a constrained environment, leading to the formation of embryos [2]. After a period of cultivation, one or more healthy embryos are selected and transferred into the woman's uterus, with the hope that implantation will occur, resulting in a successful pregnancy. The journey through IVF is often characterized by emotional, physical, and financial challenges, as individuals navigate a series of medical procedures, fertility medications, and uncertainties about the treatment's outcome. However, the potential

outcome obtained from overcoming infertility, and achieving successful gestation results makes the IVF technique to be the most pursued beacon of hope for those struggling with reproductive complications. Whilethis domain hasdeveloped in neoteric times, the challenges with identifying the triggering causes of success and failure of IVF treatments is a complex bottleneck faced by medical practitioners. The intricate relations between corporeal parameters and medical solutions that can enhance the success rate of IVF treatments are indagated with evolving changes even in current times. This potential loophole entails the utilization of computational routines and simulative paradigms that handles multi-dimensional data, and deciphers patterns in data to further serve as a decisive aspect to achieving accurate results [5]. Besides the above constraints, the IVF treatments are expensive and many struggle to afford these procedures to achieve successful reproduction. The progression of advanced computational methods like artificial neural networks comprising of machine learning mechanisms have garnered attention to uncluttered data-driven approaches [10] to IVF success rate enhancement. The mounting need for machine learning classifiers [4] that outperform human inferences in agnizing complex correlations and preparing multi-dimensional data to render meaningful insights paves the way for their use in an escalated manner. The integration of machine learning paradigms [6], [7] in clinical practices is already underway due to their capability to augment the expertise of practitioners, and in establishing assisted planning of treatments for IVF necessitated patients.

The existing body of work concomitant to this area of study pivots on predictive modelling of IVF using statistical and algorithmic approaches, or on conventional classifiers [7]. However, the objective of this proposed study pivots to incorporate a correlative analysis, and hyperparameter tuning through the hybrid Chi-Kruskal's feature extractor prior to diving into the different machine learning classifier models. The correlation analysis and hyperparameter tuning proves crucial in establishing the vitality of attributes that lead to the success and failure of the IVF treatment. Furthermore, the classifier models incorporated for this study delve into identifying the stratification accuracy, while explicitly mitigating inaccuracies. The classifier models entail the standalone models such as the SVM, KNN, Random Forest [2], Logit Boost, Gradient Boosting Machine [4] and Decision tree along with the ensemble classification model to cognize the performance of the models.

This research is structured to comprehensively delineate the various steps involved in processing of the IVF data to analyzing the correlative hyperparameter tuning between attributes in the study. It is structured with section II explicating the empirical review relevant to the IVF approach, Section III elaborates the methodology, Section IV presenting the results procured in each phase of the study, with the final section concluding the indagationwith future work apposite to this vertical.

## 2. Literature Review

Pooja Bagane et al [1], in the paper titled "IVF Success Prediction using Machine Learning Techniques: A Comparative Study" elaborated about the classification of IVF dataset through a set of machine learning classifiers such as Support Vector Machines, Artificial Neural Networks, Gradient Boosting, Logistic regression and Random Forest. The classification detailed about the success and failure rate of the IVF treatment. The phases incorporated for the study entailed the collection of data, pre-processing to standardize the data, along with handling missing numbers is effectuated in this phase. In the feature selection phase, the study uses four different techniques such as univariate, correlative, mutual information, and Recursive feature elimination method to analyze the importance of different features in the study. The Principle Component Analysis (PCA) was used to extract the features. Following the extraction of features, the classifier methods were incorporated, and the results showed enhanced accuracy with Gradient Boosting evincing 87%, Random Forest with 86%, Logistic Regression showing 85% and SVM depicting an accuracy of 67%. The future work of the study pivots to implement hyperparameter tuning inorder to procure more precise results.

Zohar Barnett-Itzhaki et al [3], in their research evaluated the efficacy of machine learning algorithms over traditional statistical modelling practices for the prediction of IVF. For the study, the paper incorporated two machine learning algorithms such as the Support Vector Machine (SVM) and the Artificial Neural Networks in juxtapose with the statistical method such as the logistic regression to identify the efficiency of IVF prediction. The dataset entailed the common features such as the BMI, health of oocytes, positive beta-hCG levels, earlier pregnancies and live births to establish comparative analysis between the two groups of analysis. The results

evinced that the artificial neural networks rendered the highest accuracy of 90%, with SVM following a 77% highest range of accuracy, with the least accuracy rendered by logistic regression at a maximum of 74%, thereby effectively delineating that machine learning mechanisms surpass the traditional statistical logistical regression model. The study also elaborated the prominence of this results aiding medical practitioners to adopt these methods for further treatment counselling, assistance and targeted strategical planning of the treatment.

## 3. Methodology

The methodology for this study is implemented with a systematic, yet thorough process of sequential processing of the IVF dataset. The attributes for this research are pivotal in establishing the state of diagnosis entailed, while effectively fusing the phases to ensure precise classification of the target attribute. The work-flow diagram for this indagation is presented as shown in Figure 1.

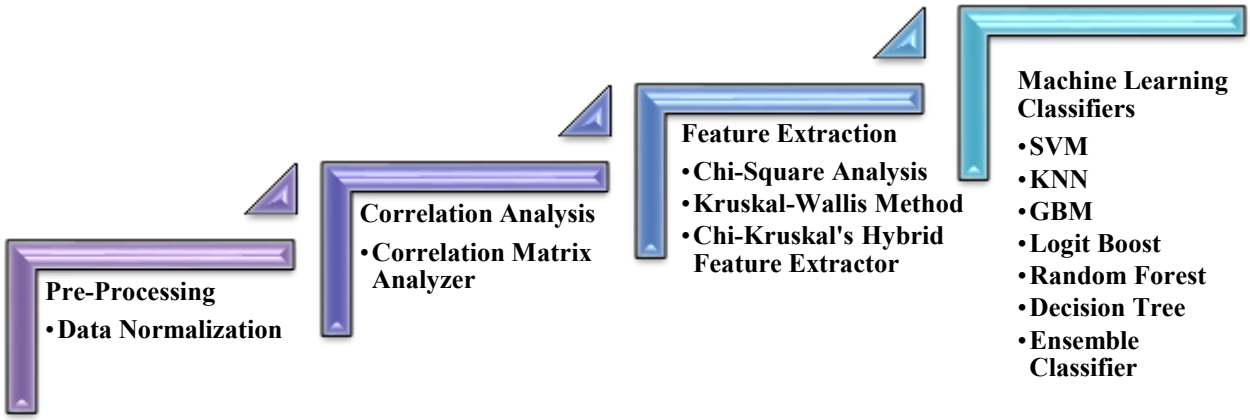


**Figure1.Implementation workflow**

Dataset procured for this study is garnered from Kaggle. The dataset comprises of a total of 1274 patient details, with the target attribute being the 'Diagnosis' that identifies the 'success' and 'failure' [8] of the IVF process implemented on the patient. The other attributes which hold a combination of categorical and numerical data are depicted in the diagram below:

Pre-processing is one of the crucial aspects of this study, entailing the process of normalization on the incorporated data [17], [20]. Since the data does not hold any missing values, the subsequent phase of data normalization is entailed. However, prior to normalization, the conversion of categorical to numerical data is effectuated inorder to standardize the data. Except the target attribute, the other attributes are normalized using the Interquartile range (IQR) and median-centered scaling [19]. This technique incorporates the median data points, which are centered around zero to be subtracted with other data points thereby normalizing the entire dataset to a specific range of distribution. The choice of IQR is primarily to control the outliers, and render consistent throughput. The formula for using IQR [20] for data normalization is as follows:

$$d' = \frac{d - \text{Median}}{\text{IQR}} \quad (1)$$

Where d' is the throughput obtained after application of the formula, or the final normalized data, d is the original attribute applied for normalization and Median infers to the middle data point and IQR is the interquartile range chosen for the study (Difference of Quartile 3 comprising of 75% of the data, and Quartile 1 comprising of 25% of the data [20]).

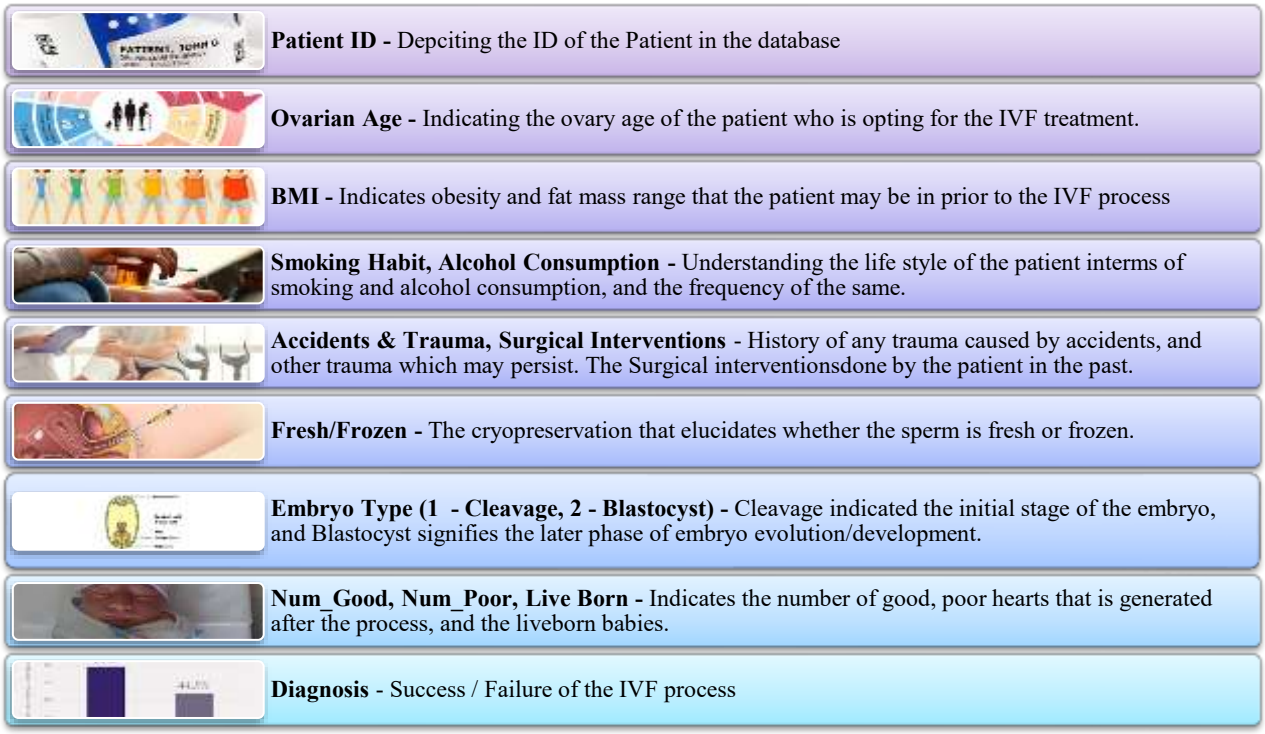


**Figure 2. Attributes for the study**

The correlation analysis between the different attributes of the study is implemented through the correlation analyzer using the Pearson method. The Pearson correlation coefficient is used to analyze the cohesiveness and proximity of attributes, while incorporating linear relationshipbetween the variables. The range of correlation lies between -1 to +1, with the negative value indicating negative linear relationship, and the positive tangent reflecting the positive relationship between attributes [9]. The value 0 will signify no potential relationship between the variables. The formula for Pearson Correlation Coefficient is as given below:

$$P = \frac{\sum(m_i - \bar{m})(n_i - \bar{n})}{\sqrt{\sum(m_i - \bar{m})^2 \sum(n_i - \bar{n})^2}} \quad (2)$$

Where $\mathbf{m_i}$and$\mathbf{n_i}$are individual data points, and $\bar{\mathbf{m}}$and$\bar{\mathbf{n}}$ are mean values of the individual mean points. The covariance of the data points m and n are achieved by the product of the difference from their mean values, while the denominator is established with the product difference between the standard deviation of the two points.

The next process implemented in this study is the feature extraction phase. The features [12] in this indagation after pre-processing and correlation analysis is effectuated into a contingency table that correlates with the target attribute to compute the feature frequencies. The chi-square formula [1] is a difference established between the observed and the expected frequencies of the data points, while relating them to a specific p-value based on which the features are ranked. The formula for chi-square feature extractor is as given below:

$$X^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (3)$$

Where$O_i$, $E_i$ are the observed and expected frequency of the data.

The Kruskal-Wallis [21] method on the other hand delineates the efficacy of the features through a non-parametric approach that analyzes the distributions of the feature group using the chi-square value of the individual attributes. The ranks are then displayed post summation of the feature groups. The Kruskal-Wallis feature extractor for this study is effectuated using the below equation:

$$KW = \frac{12}{m\,(m+1)} \sum \frac{S_r^2}{m_k} - 3\,(m+1) \quad (4)$$

Where KW is the throughput obtained after processing the above formula, m is the number of records used in the database, $m_k$ is the number of records in the $K^{th}$ group, and $S_r$ is the sum of ranks in the $K^{th}$ group.

This study also incorporates a hybrid feature extractor by combining the Chi-Square and the Kruskal-Wallis methods through the following steps:

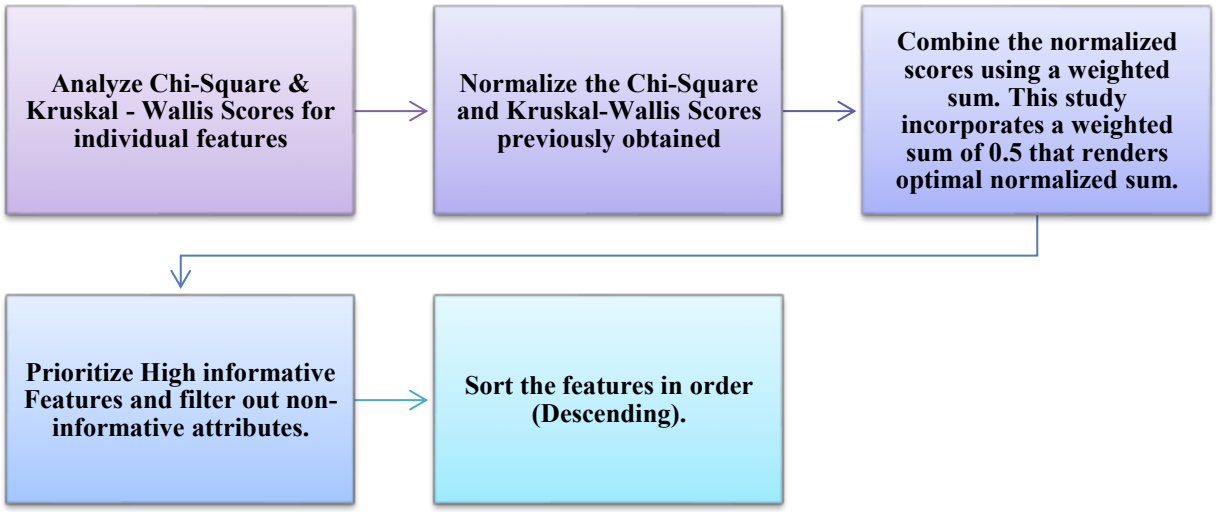


**Figure 3. Step-wise implementation of chi-kruskal hybrid feature extractor**

The hybrid feature extractor is specifically incorporated due to the combinational data comprising of categorical and continuous values, thereby ensuring optimal feature selection, extraction and ranking.

The machine learning classification process is subsequent to feature extraction, and is implemented through a set of classifiers as delineated below:

After the feature map ranking and extraction, and prior to delving into classification [11], the dataset is split into training and testing sets comprising of 80% for the former, and 20% for the latter. After splitting the dataset, the classifiers are applied to comprehend the overall accuracy and the performance of the entailed machine learning classifiers.

The Support Vector Machine is a machine learning classifier [13], that aids in establishing stratification of the incorporated dataset. To effectively classify the IVF diagnosis based on the features extracted, the 'fitcsvm' function is used. The kernel function is used to efficiently establish decision boundaries, while significantly embedding the Radial Basis Function (RBF) kernel. The 'boxconstraint' of 0.1 and kernel scale fixed to auto mode is set. The RBF kernel in the SVM is activated by using the below formula:

$$R = \exp(-h\,|m - n|^2) \quad (5)$$

Where h represents the gamma value, and m, n are the datapoints used for evaluation.

The accuracy procured for the above classifier is delineated in the subsequent section of this study.

The next classifier used for this indagation is the K-Nearest Neighbor (KNN) algorithm [14], that classifies the data based on the number of neighbors, and the ration of neighborhood stratification executed through Euclidean distance. This algorithm is also referred to as lazy learner due to its capacity to learn from the neighbors, and assigns those neighbors that lie in close proximity with each other through the concept of majority voting adopted by the 'fitcknn' function. The below formula is indicative of the Euclidean distance, and is as follows:

$$D = \sqrt{(M_2 - M_1)^2 + (N_2 - N_1)^2} \quad (6)$$

Where M and N are coordinate points for any two data in a two-dimensional dataspace.

The decision tree [15], [16]is a classifier model that utilizes decision nodes to subsequently form the leaf nodes from the procured features and a certain threshold value using the splitting mechanism. The slitting criteria isimplemented by using the Gini Index, thereby enabling optimality of classification, and further to reduce the

mean square error that induces higher loss in the model. This algorithm is implemented by using the 'fitctree' function, with the maximum split set to 20, and the number of cycles fixed to be 150. The accuracy results thus obtained from this tree classifier is presented in the subsequent section.

The subsequent set of classifiers such as the Random Forest, Logit boosting and Gradient boosting machine are a class of ensemble classifiers [19] that use the cluster of 'TemplateTree' as a learner base, and are implemented using the 'fitcensemble' function.

The Random Forest (RF) classifier [18]uses a group of clustered decision trees to render the throughput of accuracy. This algorithm adopts the bootstrap aggregation method encompassed with the 'fitcensemble' function. The number of cycles for this study is set to 150, and as delineated previously, uses the 'TemplateTree' learners as clustered learners. The maximum split number is set to 30 for optimal decision boundary. The RF classifier is defined by the below mathematical equation:

$$F_r = \frac{1}{G}\sum_{g=1}^{G} P_g(x) \quad (7)$$

Where G is the number of trees, and $P_g(x)$is the classification throughput obtained after processing.

Logit boosting adopts the working of additive logistic regression that effectuates optimal stratification. As elucidated above, the logit boosting algorithm is effectuated using the 'fitcensemble' function and the 'TemplateTree' learner, with the number of cycles fixed to 150, and the maximum number of tree splits set to 50. This model adaptively adjusts the weights to the base learner to sequentially grow with every iteration, thereby significantly mitigating the bias and variance in the network. The network works based on the following equation:

$$LG = \frac{C_i - P_i}{P_i\,(1 - P_i)} \quad (8)$$

Where $C_i$ represents the diagnosis class labels, while $P_i$ signifies the predicted probability entailed in the algorithm.

The Gradient Boosting Machine (GBM) [4], [20] is trained to mitigate residual errors in the network by using gradient descent. This network pivots of sequential model building to effectively reduce differentiable loss in the network. The 'gentle boost' mechanism established with 'TemplateTree' leaners comprising of maximum number of splits to be 80, and the number of cycles at 150 is used to build this model. The mathematical representation of this model is given below:

$$g_r = -\left[\frac{\partial L(x_i,\ F(y_i))}{\partial F(y_i)}\right]_{F_y = F_{m-1}(y)} \quad (9)$$

Where $g_r$ is the residual gradient for the network, $\frac{\partial L(x_i,\ F(y_i))}{\partial F(y_i)}$is the differential loss function computed for the network after the updation of weights, and $F_{m-1}(y)$ is the updated phase after residual error correction.

The final classifier used for this study is the hybrid ensemble classifier that is built using KNN as the base classifier, through the subspace method. This method optimizes the base classifier during the process of training by randomizing the feature subsets to enhance generalization and incorporate diversity of KNN subclassification to consecutively reduce overfitting. The 'fitcensemble' method with the number of neighbors for the KNN base learner is set to 5, and the number of cycles is set to 150 is used for this study.The resultsin relevance to the different algorithm used interms of their accuracy are visually presented for enhanced comprehension in Section IV.

## 4. Results

This section presents the results procured from the simulations effectuated in MATLAB. The data view not only presents the structural view of the attributes and the data it comprises, but also renders the statistics pertaining the attribute of study. A total of 13 attributes excluding the target attribute is loaded for this research. The target attribute is the diagnosis implemented for the patient in identifying if the success and failure of the treatment process.

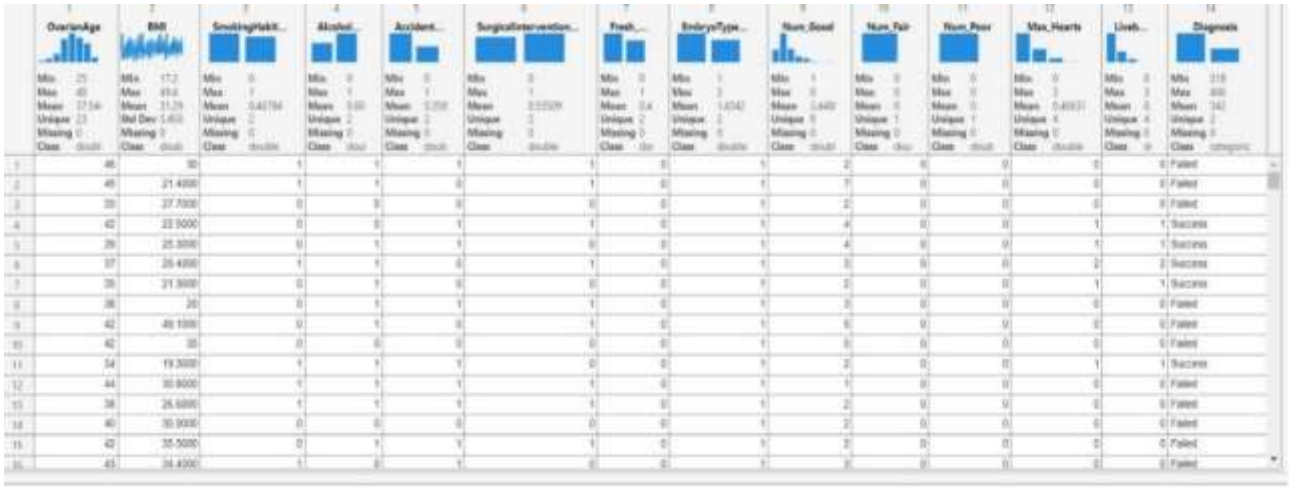

**Figure4. Data view with data statistics**

The Min, Max, Mean, Unique, Missing and Class are some of the statistical and analytical data pertaining each attribute in the dataset. This analytical components aid in significanctly pre-processing the data in an efficient manner.

Since the data is cleaned, and the noise level is trivial, the next phase of this study entails the process of normalization, which aids in standardizing the data. The process of normalization is performed through the center and scale normalization using the interquartile range (IQR) and median-centered scaling that is employed on each attributed of the study inorder to compress them within a set of values. The result thus obtained for normalization is presented in Figure5.

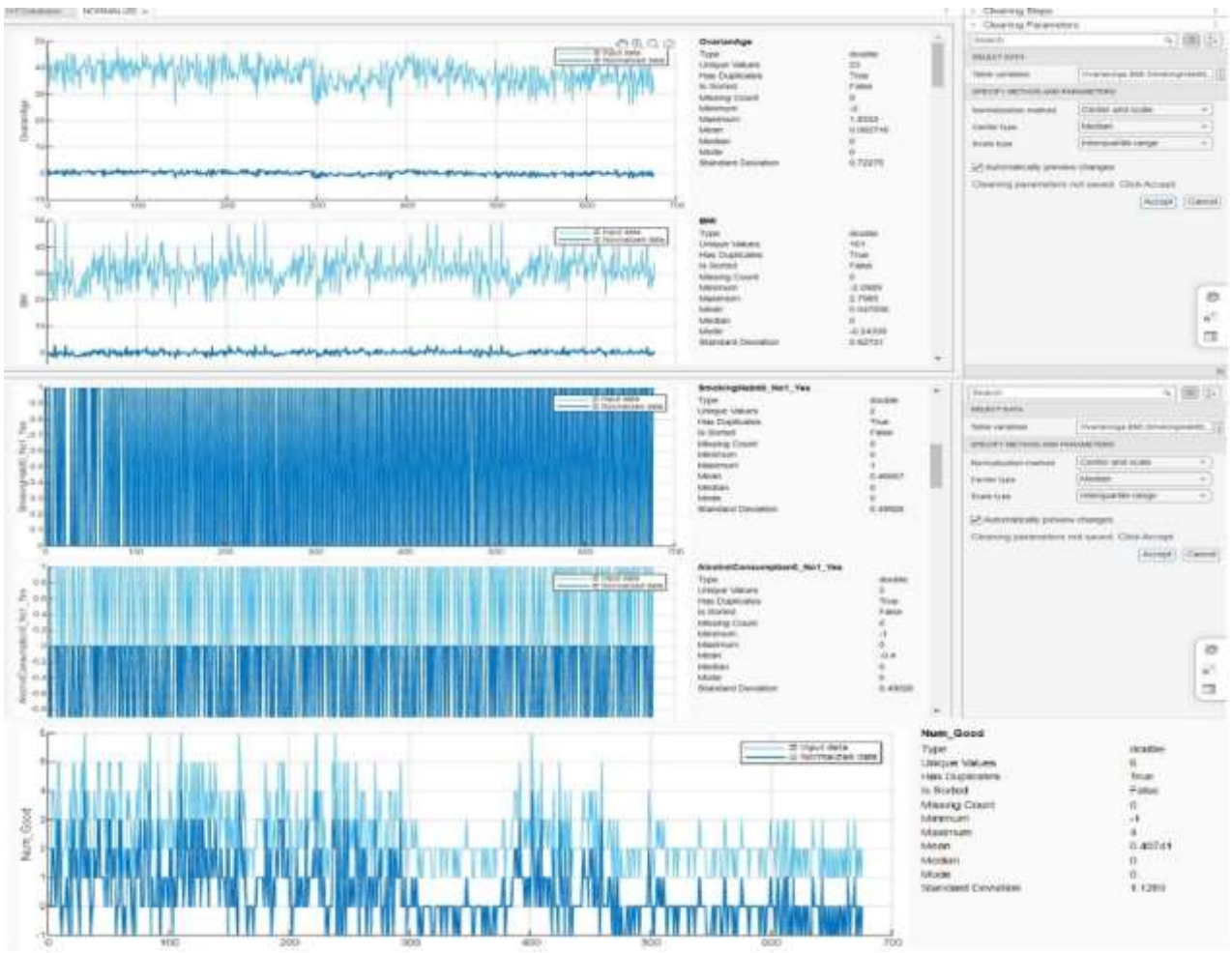

**Figure5. Normalization through interquartile range-median centered Scaling**

Next, the correlative feature matrix calculator using Pearson correlation is entailed to analyze the attribute relevance between the features, and the same is presented in Figure6 through a heatmap. The correlation matrix is generated for those features that hold high cohesive value, and ignores those parameters that holdminimal cohesive value in this study.

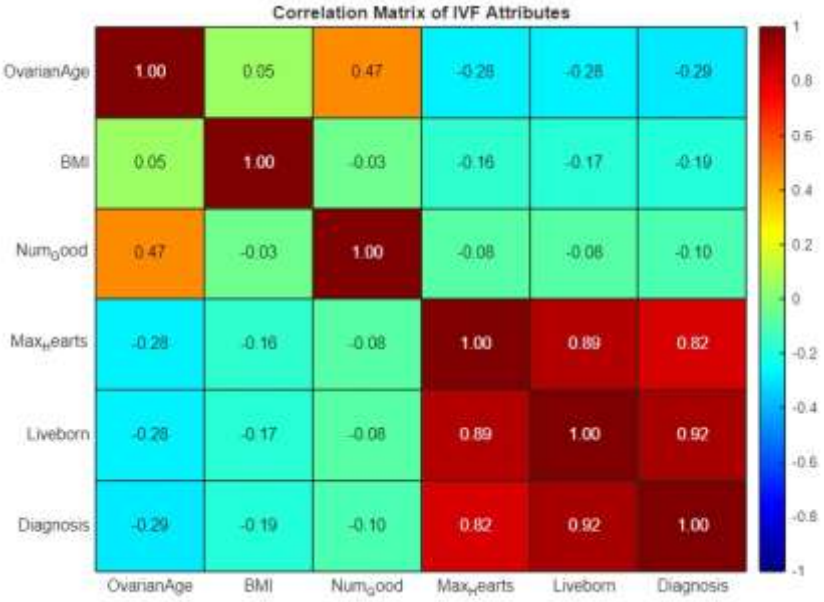


**Figure6. Feature matrix using pearson correlation**

The feature analysis through hyperparameter tuning is implemented by three methods such as the Chi-Square method, the Kruskal-Walli's method, and the hybrid approach of Chi-Kruskal's feature extractor that unsheathes the best features by combining the normalized scores using a weighted sum value for all attributes in the database. The results thus procured for each feature analyzer is as given below through a series of diagram from Figure7. Through to Figure9.

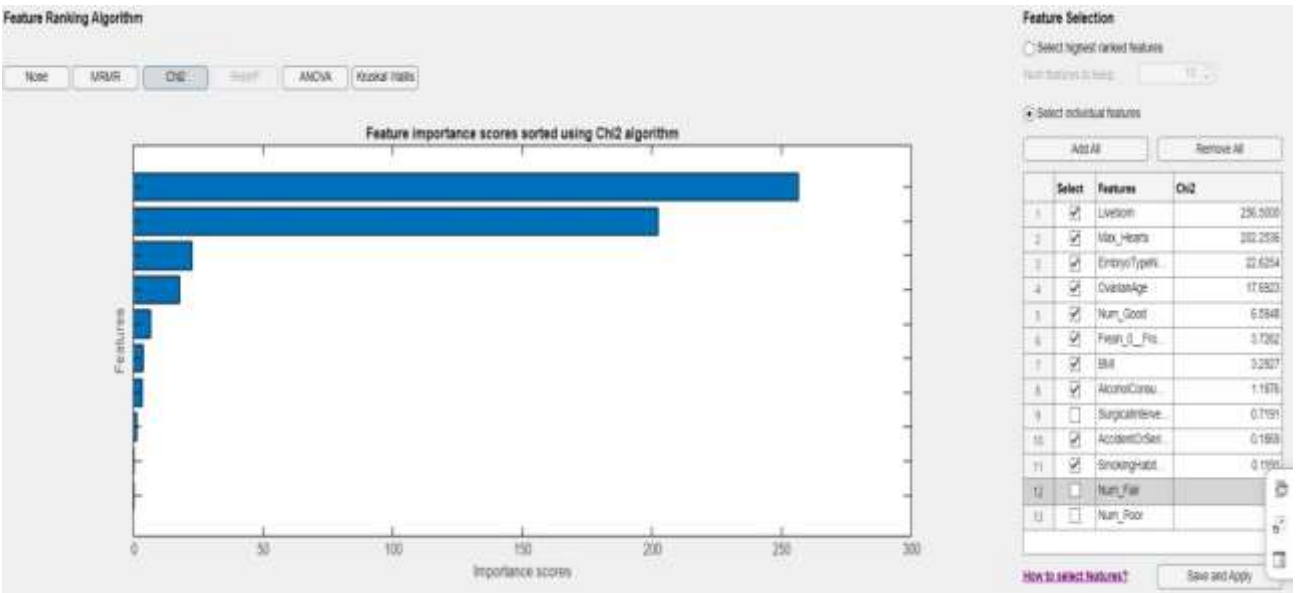


**Figure7. Chi-square feature extraction**

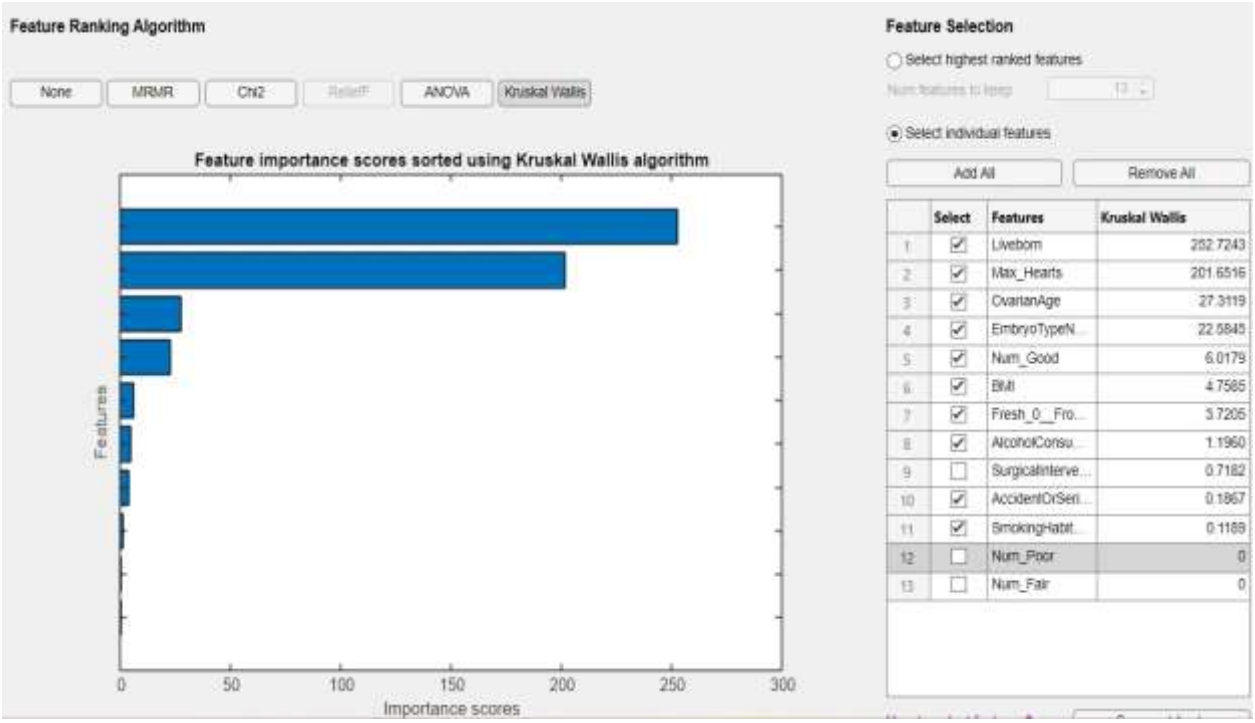


**Figure8. Kruskal-walli's feature extraction**

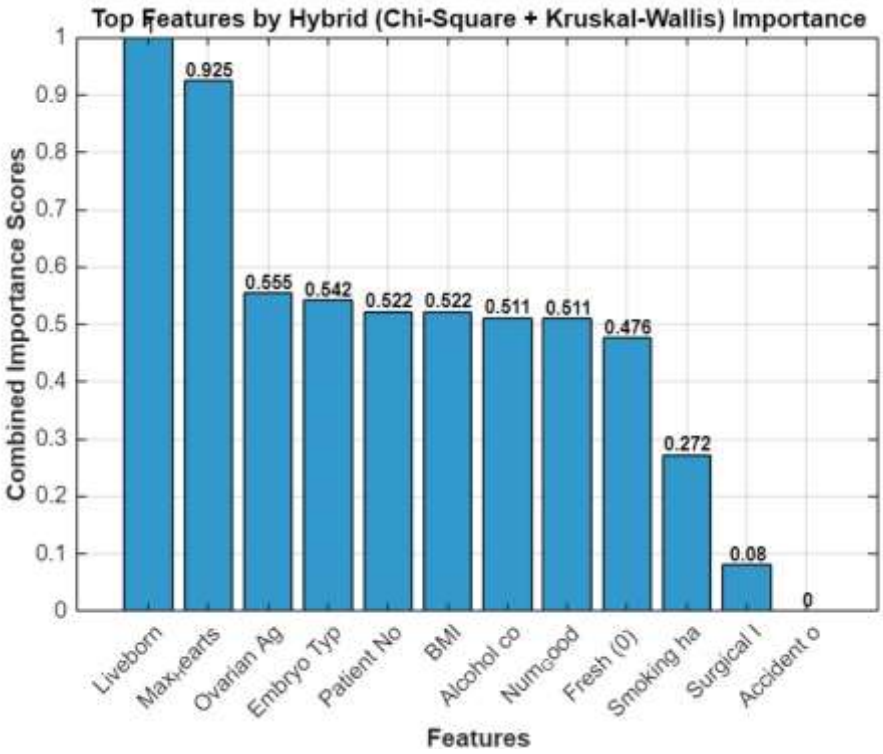


**Figure9. Chi-kruskal feature extractor**

The final process of this study is the incorporation of machine learning classifiers that are utilized to analyse the specific performance of the models. The comparative performance chart for the models, along with the respective accuracy analysis is illustrated in Figure10.

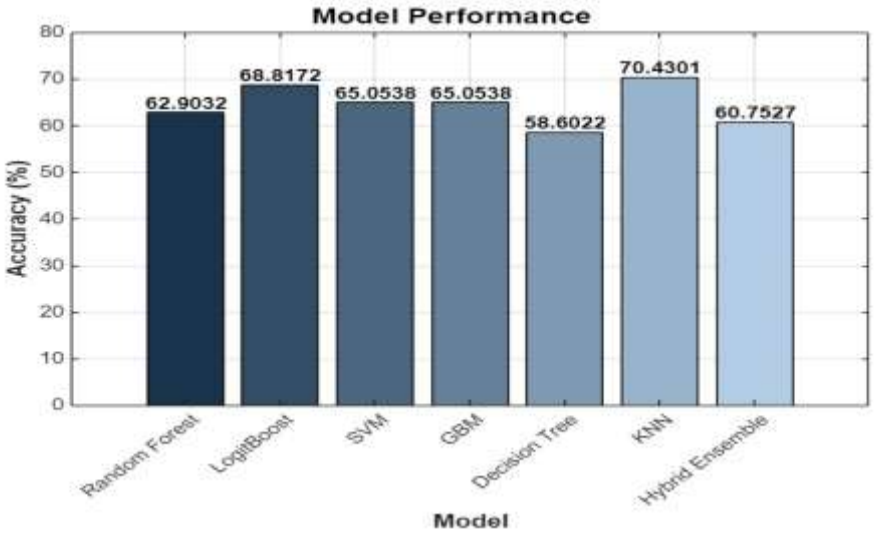


**Figure10. Comparative classifier performance chart**

## 5. Conclusion

The process of In-Vitro Fertilization in recent times has evinced progressive changes, and the need for computational inclusion for accurate results is indispensable. Concluding the indagation presented, the hyperparameter tuning and correlation analysis prove to be imperative for precise classifier evaluation. The utilization of the hybrid Chi-Kruskal's feature extractor expediates ascertaining of crucial features that can explicitly determine the success and failure rate of the IVF process. Furthermore, while the pre-processing and feature extractors accelerate the accuracy of the classifier models, the performance of the classifiers rendered mediocre throughput. The K-nearest neighbors (KNN) yielded an accuracy rate of 70.4%, while the decision tree funneled down to 58.6%, with other classifiers positioned within the range of 60.7% to 68.8%. The mediocrity of the model's performances is attributed to the study's establishment of pragmatic evaluation that is induced with gaussian noise, thereby achieving real-time environmental results for each of the machine learning classifier models. However, further studies can be incorporated with novel optimization methods that augment the classifier performance, while significantly establishing a potent mechanism to achieve impeccable accuracy in reproductive medicine.

## References

1. Bagane, P., Kulshretha, S., Mishra, M., Asrani, Y., & Maheswari, V. (2024). IVF success prediction using machine learning techniques: A comparative study. International Journal of Intelligent Systems and Applications in Engineering, 819–829.
2. Raef, B., Maleki, M., & Ferdousi, R. (2020). Computational prediction of implantation outcome after embryo transfer. Health Informatics Journal, 26(3), 1810–1826.https://doi.org/10.1177/146045821989213
3. Barnett-Itzhaki, Z., Elbaz, M., Butterman, R., Amar, D., Amitay, M., Racowsky, C., Orvieto, R., Hauser, R., Baccarelli, A. A., &Machtinger, R. (2020). Machine learning vs. classic statistics for the prediction of IVF outcomes. Journal of Assisted Reproduction and Genetics, 37(10), 2405–2412. https://doi.org/10.1007/s10815-020-01908-1
4. Louis, C. M., Handayani, N., Aprilliana, T., Polim, A. A., Boediono, A., & Sini, I. (2023). Genetic algorithm assisted machine learning for clinical pregnancy prediction in IVF. AJOG Global Reports, 3(1), 100133. https://doi.org/10.1016/j.xagr.2022.100133
5. Septiandri, A. A., Jamal, A., Iffanolida, P. A., Riayati, O., &Wiweko, B. (2020). Human blastocyst classification after in vitro fertilization using deep learning. IEEE. https://doi.org/10.48550/arXiv.2008.12480
6. Gowda, G., Gowramma, G. S., & Mahesh, T. R. (2023). An automatic system for IVF data classification by utilizing multilayer perceptron algorithm. Proceedings of the International Conference on Current Trends in Engineering, Science and Technology (ICCTES).

7. Tadepalli, S. K., & Lakshmi, P. V. (2019). Application of machine learning and artificial intelligence techniques for IVF analysis and prediction. International Journal of Big Data and Analytics in Healthcare, 4(2), 21–33. https://doi.org/10.4018/IJBDAH.2019070102
8. A.Surendar. (2026). Geometry-Aware Dual-Timescale RIS Configuration for Indoor 6G IoT Links. VSF Journal of Electronics and Communication Engineering, 1(1), 1–7.
9. Centers for Disease Control and Prevention. IVF success estimator. https://www.cdc.gov/art/ivf-success-estimator/
10. Chen, T.-J., Zheng, W.-L., Liu, C.-H., Huang, I., Lai, H.-H., & Liu, M. (2019). Using deep learning with large dataset of microscope images to develop an automated embryo grading system. Fertility & Reproduction, 1(1), 51–56.
11. Zaninovic, N., Elemento, O., &Rosenwaks, Z. (2019). Artificial intelligence: Its applications in reproductive medicine and the assisted reproductive technologies. Fertility and Sterility, 112(1), 28–30.
12. Khosravi, P., Kazemi, E., Zhan, Q., Malmsten, J. E., Toschi, M., Zisimopoulos, P., Sigaras, A., Lavery, S., Cooper, L. A., Hickman, C., et al. (2019). Deep learning enables robust assessment and selection of human blastocysts after in vitro fertilization. npj Digital Medicine, 2(1), 21.
13. Robbi Rahim. (2026). Crispr-Assisted Metabolic Engineering Of Microbial Cell Factories For Enhanced Industrial Enzyme Biosynthesis. Sirashmi Archives Of Applied Biotechnology And Microbial Research, 1-9.
14. Li, L., Cui, X., Yang, J., Wu, X., & Zhao, G. (2023). Using feature optimization and LightGBM algorithm to predict the clinical pregnancy outcomes after in vitro fertilization. Frontiers in Endocrinology. https://doi.org/10.3389/fendo.2023.1305473
15. Blank, C., Wildeboer, R. R., DeCroo, I., Tilleman, K., Weyers, B., & de Sutter, P. (2019). Prediction of implantation after blastocyst transfer in in vitro fertilization: A machine-learning perspective. Fertility and Sterility, 111(2), 318–326. https://doi.org/10.1016/j.fertnstert.2018.10.030
16. Vogiatzi, P., Pouliakis, A., &Siristatidis, C. (2019). An artificial neural network for the prediction of assisted reproduction outcome. Journal of Assisted Reproduction and Genetics, 36(7), 1441–1448. https://doi.org/10.1007/s10815-019-01498-7
17. Handayani, N., Louis, C. M., Erwin, A., Aprilliana, T., Polim, A. A., Sirait, B., Boediono, A., & Sini, I. (2022). Machine learning approach to predict clinical pregnancy potential in women undergoing IVF program. Journal of Fertility & Reproduction, 4(2), 77–87. https://doi.org/10.1142/S2661318222500098
18. Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. Journal of Applied Science and Technology Trends, 2(1), 20–28. https://doi.org/10.38094/jastt20165
19. Louis, C. M., Erwin, A., Handayani, N., Polim, A. A., Boediono, A., & Sini, I. (2021). Review of computer vision application in in vitro fertilization: The application of deep learning-based computer vision technology in the world of IVF. Journal of Assisted Reproduction and Genetics. https://doi.org/10.1007/s10815-021-02123-2
20. P. Joshua Reginald. (2026). Performance Evaluation of Geopolymer Green Concrete Incorporating Industrial By-Products for Sustainable Structural Construction. Sirashmi Sustainable Civil Engineering and Smart Architecture Research, 1-9.
21. Raef, B., & Ferdousi, R. (2019). A review of machine learning approaches in assisted reproductive technologies. Acta Informatica Medica, 27(3), 205–211. https://doi.org/10.5455/aim.2019.27.205-211
22. de Amorim, L. B. V., Cavalcanti, G. D. C., & Cruz, R. M. O. (2023). The choice of scaling technique matters for classification performance. Applied Soft Computing, 133, 109924. https://doi.org/10.48550/arXiv.2212.12343
23. Izonin, I., Tkachenko, R., Shakhovska, N., Ilchyshyn, B., & Singh, K. K. (2022). A two-step data normalization approach for improving classification accuracy in the medical diagnosis domain. Mathematics, 10(11), 1942. https://doi.org/10.3390/math10111942
24. Kijkarncharoensin, A., &Innet, S. (2023). The Kruskal-Wallis test to examine performance inconsistency in the biomass higher heating value models for proximate analysis. Journal of Engineering Science and Technology, 18(1), 463–480.