Al Kut Journal of Economics and Administrative Sciences /ISSN: 1999 -558X /ISSN Online 2707-4560/ Vol (17) Issue: 57-2025 (June)

AL KUT JOURNAL OF ECONOMIC AND ADMINISTRATIVE SCIENCES
Publisher: College of Economics and Management - Wasit University

Comparative Analysis of Machine Learning Algorithms for Leukemia Disease Classification

Dara Hassan Ahmed

General director of Runaki Institute in Sulaimaniyah, Applied statistics[2]

Darahassan755@gmail.com

Abstract.

Leukemia is a cancer of the body's blood-forming tissues, such as the bone marrow and lymphatic system, resulting in the overproduction of abnormal white blood cells. These white blood cells multiply endlessly, disrupting the body’s ability to fight infections, manage bleeding and other basic tasks. The goal of this study is to compare three machine–learning algorithms Random Tree (RT), Extra Tree (ET) and Random Forest (RF) to classify leukemia disease. This study is based on data obtained from Hiwa Hospital located in Sulaimani City, consisting of records from 616 patients. This dataset assesses the efficiency of three machine learning classifiers in predicting 'Alive' and 'Death' occurrences in leukemia patients. Different metrics have been used to evaluate the performance of the models, including accuracy, sensitivity, specificity, ROC Area, F-measure or confusion matrix. Results with Random Forest (RF) yielding the highest accuracy (87.66%) and sensitivity (97.09%) and ROC Area (0.579) as compared to Random Tree (RT) and Extra Tree (ET). The second place is for Extra Tree with an accuracy of 87.01% and sensitivity of 96.17%, followed by Random Tree with accuracy (correctly classified) of 86.69% and worse performance in sensitivity and ROC Area. However, these three models show low specificity, highlighting the difficulty in predicting 'Death' cases accurately. our findings demonstrate that RF shows excellent performance in terms of accuracy in leukemia classification, while ET and RT also produce competitive predictions. The routing RF algorithm considers this piece of study as a potential choice for implementing the task whilst ET algorithm is highlighted as a strong competitive alternative and RT as a sufficient solution in situations where less computational consumption is a preferred requirement. Further work needs to be done on specificity to improve the identification of instances which predicted 'Death'.

Keywords: Machine learning algorithms, leukemia, Extra Tree, Random Forest, Random Tree

1. Introduction

Leukemia is a malignant disease that affects the blood and bone marrow, characterized by excessive production of white blood cells that compromise the body’s capacity to combat infections and generate normal blood cells [1]. The early and accurate identification of leukemia plays a critical role since it considerably impacts treatment planning and patient concurrent survival. Additionally, traditional diagnostic approaches are generally accurate but typically require significant time, resources, and can be subject to human interpretation, leading to differences in how results are interpreted. Machine learning (ML) has recently become popular in medical diagnostics, allowing for greater speed and accuracy in disease detection and classification[6] Data classification methods are one of the most used methods in statistical methods[10]. Machine learning classification algorithms learn to distinguish between classes of data from features presented to them [9]. These algorithms learn patterns from the training data and make predictions on new, unseen data [8]. The commonly used classification algorithms can be decision trees, support vector machines and neural networks that find applications in multiple fields like healthcare, finance and security domains. Decision tree algorithms are for example one of the well-known algorithms in here since they are simple, interpretable, and need a low computational time. They break decisions to a tree-like structure where data is split according to feature values until predictions are made. Decision trees are able to work on both numerical and categorical data and often require little pre-processing, which makes them suitable for many classification tasks[14].

In this study, we performed a comparative study of Random Tree (RT), Extra Tree (ET), and Random Forest (RF) algorithms to check their classification accuracy for leukemia disease classification. These algorithms are decision tree-based models which are interpretable, scalable, and optimal in performance, and are particularly performing well in biomedical data sets. When evaluating the performance of each model, standard evaluation metrics: Accuracy, Precision, Recall and F-measure are used [11]. Using the comparison of these algorithms, the study intends to determine the most appropriate method for leukemia classification. This can lead us to integrate certain novel ML approaches we might find useful into medical diagnostics, which in turn is very likely to benefit clinicians to offer a more precise diagnosis and so on, ultimately improving overall patient outcomes.

2. Objective of the Research

This study's main goal is to assess the accuracy of leukaemia illness prediction by comparing several machine learning methods. Additionally, the information, which included 616 patient records, came from Hiwa Hospital in Sulaimani City. Three machine learning approaches are used in this study: Random Forest (RF), Extra Tree (ET), and Random Tree (RT).

3. Materials and Methods

Leukemia disease has several classification algorithms which has classification purposes RT (Random Tree), ET (Extra Tree) and RF (Random Forest) as well as is studied. A demonstration of the working machine is shown in Figure 1. Once the suitable classification algorithm become used to in finding the appropriate one for predicting leukemia disease.

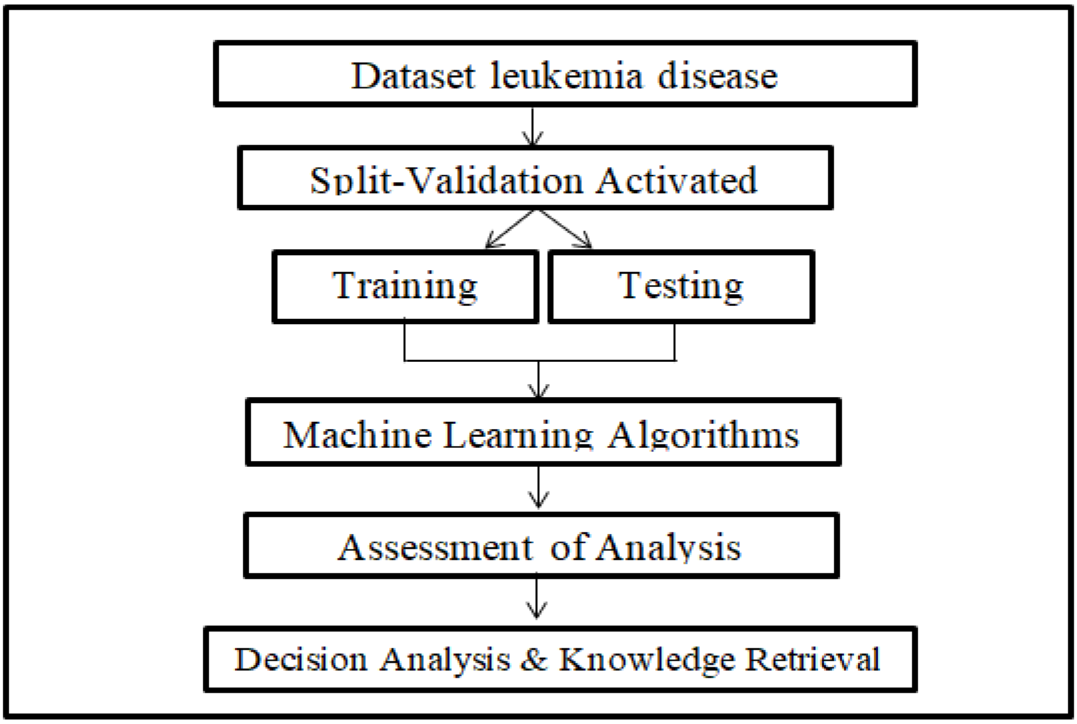

Figure (1): working process

In this lookup work, the leukemia disease is to be classified using often used three-machine mastering algorithms are given below:

Random Tree (RT) [4][7]

Random tree (RT) is a decision tree algorithm that employs a recursive process that splits the dataset based in each iteration on one of the randomly selected features. RT introduces a randomization element in selecting the features at each node, unlike classical decision trees. This randomization helps avoid over fitting and increases generalization. A fast, simple algorithm suitable for large datasets with a high degree of efficiency at low computational cost is offered. But because RT only uses one tree, it might be sensitive to data variations and less accurate than ensemble methods. RT is frequently employed to find a baseline comparison in exploratory data classification tasks..

Extra Tree (ET) [3][11][12]

Extremely Randomized Trees (ET) is a machine learning algorithm used to build an ensemble of unpruned decision trees based on random subsets of features and random split thresholds. Unlike Random Trees, ETs alone does not depend on best split, but rather relies on random splits, but doing this usually results in better variance and lower training times. We trained and combined them to form a random ensemble. This randomness allows ET avoids over fitting and is robust against noisy datasets or complex datasets. Due to the fact that ET is trained faster and generalizes better, it is often used in a group of models (known as an ensemble), and it is also suitable for both classification and regression tasks.

Random Forest (RF) [2][4]

Random forest (RF) is a robust ensemble learning technique that combines various decision trees to increase classification performance and minimize over fitting. Random forests work by creating many trees based on random subsets of the training data and features, averaging prediction from those, or majority voting for classification tasks. This collection strategy creates a more robust and stable model with higher levels of generalization. RF has been shown to work well when the dataset is high-dimensional, imbalanced in nature with presence of noise which is often encountered in medical applications like disease classification. The built-in feature importance metrics also provide some insights on the relevance of data, thus improving the interpretability of the model and predictive performance

3. Performance Evaluation [5][14][7]

A confusion matrix, indicated by Table 1, assesses the performance of a classifier. True positives (TP) refers to the true positive cases and true negatives (TN) refers to the true negative cases. FP: Negative instances are incorrectly classified as positive, FN: Positive instances are incorrectly classified as negative. This confusion matrix is crucial t determine the accuracy of the classifier and distribution of errors to capture the required improvements and improve the model. Understanding these metrics helps you to know where the classifier excels and where it needs improvement, to make modifications that can improve its predictive power..

Table (1) Confusion Matrix

Actual	Predicted	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Table 2 presents the classification metrics of the classifier (Classification Accuracy (CA), Sensitivity (SE), Specificity (SP), TP Rate (TP Rate), FP Rate (FP Rate), Precision (PR), Recall (RE), F-Measure (FM), MCC, and Area Under the Receiver Operating Characteristic Curve (ROC Area)). These metrics are derived from the contingency table and give a complete picture of classifier performance. Taking Sensitivity and Specificity measures into account we get a grasp of how well a classifier can detect positive and then negative cases, while Precision, Recall and F-Measure indicate the classifier’s performance on false positives and false negatives. ROC Area (AUC) and Matthews Correlation Coefficient (MCC) are advanced performance metrics used to evaluate classification models. AUC measures a model’s ability to discriminate between classes, while MCC assesses overall prediction quality by considering true and false positives and negatives. Together, they provide a balanced and comprehensive evaluation of how well a model distinguishes data in relation to actual outcomes

Table (2) Detailed Accuracy by Classes

Tools	Statistic
Accuracy	$\frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \times 100$
Sensitivity (Recall or True Positive Rate)	$\frac{\mathrm{TP}}{\mathrm{TP+FN}} \times 100$
Specificity (True Negative Rate)	$\frac{\mathrm{TN}}{\mathrm{TN+FP}} \times 100$
TP Rate (True Positive Rate)	$\frac{\mathrm{TP}}{\mathrm{TP+FN}} \times 100$
FP Rate (False Positive Rate)	$\frac{\mathrm{FP}}{\mathrm{FP+TN}} \times 100$
Precision (Positive Predictive Value)	$\frac{\mathrm{TP}}{\mathrm{TP+FP}} \times 100$
Recall (Sensitivity or True Positive Rate)	$\frac{\mathrm{TP}}{\mathrm{TP+FN}} \times 100$
F-Measure (F1 Score)	$\frac{2\,.(\mathrm{Precision}\,.\,\mathrm{Recall})}{\mathrm{Precision+\,Recall}} \times 100$
Matthews Correlation Coefficient (MCC)	$\frac{\mathrm{TP\,.TN-FP\,.FN}}{\sqrt{(TP+FP)(TP+FN)(TN+Fp)+(TN+FN)}} \times 100$
Receiver Operating Characteristic (ROC)	$\frac{\mathrm{TP+TN}}{\mathrm{TP+FN+TN+FP}} \times 100$

4. Data Analysis

This research aims to analyze various machine learning algorithms for leukemia disease classification. The Data includes 616 patients selected from Hiwa Hospitals in Sulaimani City to be investigated for leukemia (a cancer of the body's blood-forming tissues). The dataset has two common groups of variables and the first part group of explanatory variables includes age, gender, blood group, place of treatment and disease stage. The second group is the response variable, which is a determined leukemia of the patient represented by the HGB codes "alive" and "death". For analysis, Weka was applied to the use of machine learning algorithms, including Random Tree (RT), Extra Tree (ET), and Random Forest (RF).

4.1 Confusion Matrix for Machine learning algorithms

Table (3) Confusion matrix for machine learning algorithms

Random Tree (RT)				Extra Tree (ET)				Random Forest (RF)			
Actual Class	Predicted Class			Actual Class	Predicted Class			Actual Class	Predicted Class		
	Alive	Death	Total		Alive	Death	Total		Alive	Death	Total
Alive	526	23	549	Alive	528	21	549	Alive	533	16	549
Death	59	8	67	Death	59	8	67	Death	60	7	67
Total	585	31	616	Total	587	29	616	Total	593	23	616

Refer to Table 3 the confusion matrix for the Random Tree (RT), Extra Tree (ET), and Random Forest (RF) classifiers presents an interesting comparison among their classification performance. For the classifiers, all three models successfully predict the Alive class with high accuracy, with RF, ET, and RT correctly predicting 533, 528, and 526 respectively. Misclassification of Death class is same for all models, RF having minimum misclassification of 7, ET 8, and RT 59. Overall, these results show a high degree of accuracy, especially for the 'Alive' class, but also reveal some difficulties in the correct classification of 'Death' instances. As can be seen from the confusion matrix, Random Forest (RF) is the best-performing classifier out of the three algorithms. 469 instances, followed by 'Alive' (533 instances), the lowest misclassification of 'Death' (7 instances), and the highest overall accuracy of (87.66%). Hence the right prediction lies with RF, next is Extra Tree (ET) having 87.01% accuracy and lowest among rest Random Tree (RT) with

86.69% accuracy. This ranking highlights the RF’s greatest ability to reduce misclassification and attain the best overall performance.

4.2 Classification Accuracy, Sensitivity and Specificity of Proposed

Table (4) : The Classification Accuracy, Sensitivity and Specificity of Classifier

Classifier	Random Tree (RT)	Extra Tree (ET)	Random Forest (RF)
Sensitivity %	95.81	96.17	97.09
Specificity %	11.94	11.94	10.45
Accuracy %	86.69	87.01	87.66
Correctly classified %	534 (86.69%)	536 (87.01%)	540 (87.66%)
Incorrectly classified %	82 (13.31%)	80 (12.99%)	76 (12.34%)

The predicted values were calculated for the Random Tree (RT), Extra Tree (ET), and Random Forest (RF) classifiers, and the sensitivity, specificity, and accuracy are shown for each such metric in Table 4. The sensitivity, defined as the number of true positives divided by the number of true positives and false negatives, indicates that RF (97.09%) performs slightly better than ET (96.17%) and RT (95.81%) at correctly identifying 'Alive' records. This indicates that RF are better in identifying 'Alive' instances. But for the 'Death' instances, the specificity - the ability to identify the actual 'Death' instances is fairly low for all classifiers with a value between 10.45% and 11.94% that indicates there is a difficulty in making a prediction for class 'Death'. For all the models, the classification generalization accuracy is quite high, which is 87.66% for RF, followed by 87.01% for ET and 86.69% for RT. These findings explain a well general performance in classify input data with accuracy, struggling with 'Death' instances.

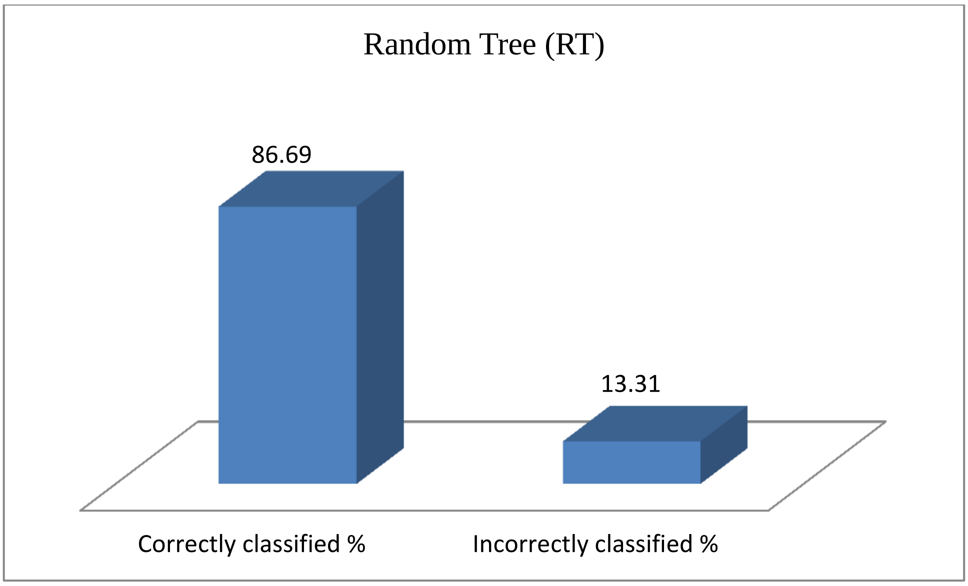

Figure (2) illustrate the rate of the correctly and incorrectly classified of the Random Tree (RT).

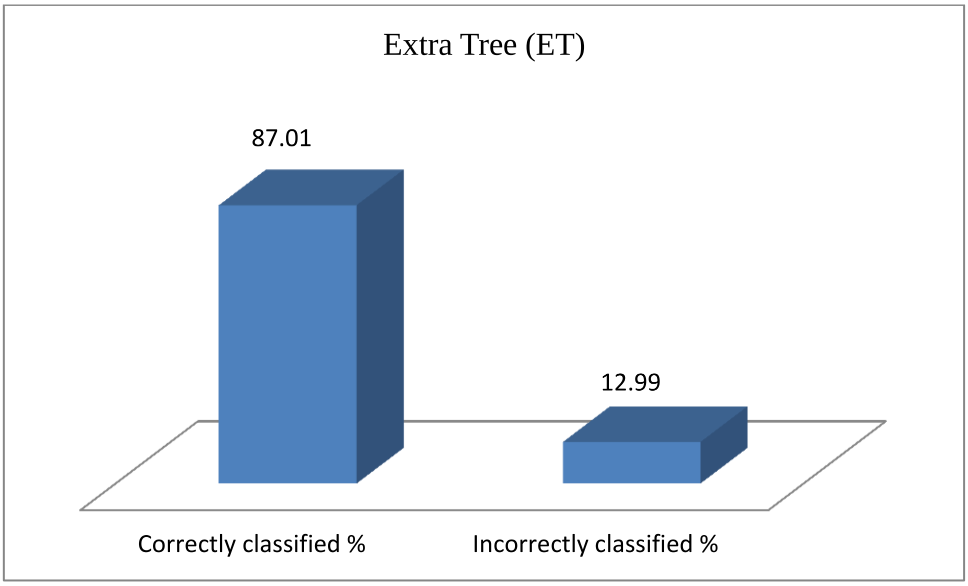

Figure (3) illustrate the rate of the correctly and incorrectly classified of the Extra Tree (ET)

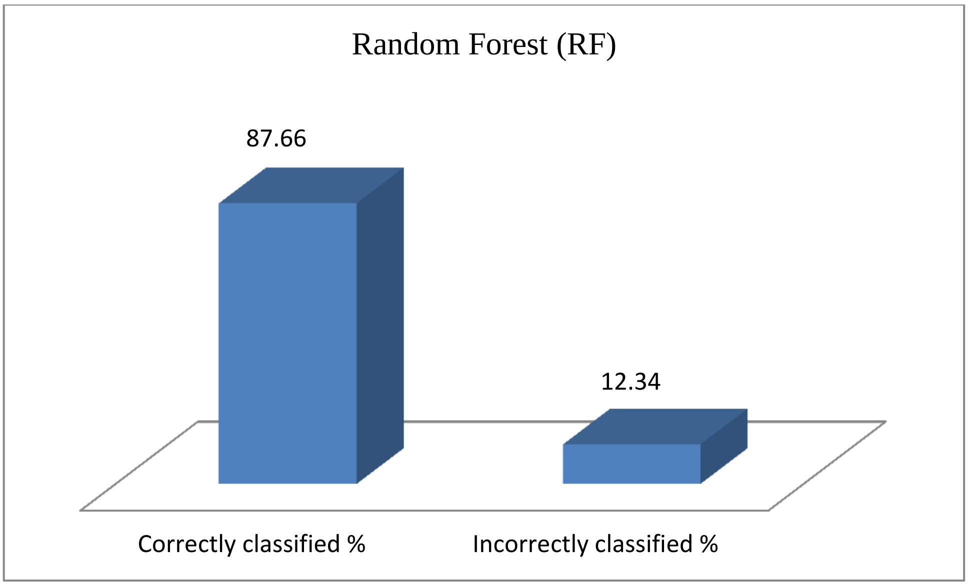

Figure (4) illustrate the rate of the correctly and incorrectly classified of the Random Forest (RF)

4.3 Calculation Detailed Performance Metrics

Table (5) Detailed Performance Metrics by Class for Classifiers

Detailed Accuracy By Classes		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area
Random Tree (RT)	Alive	0.958	0.881	0.899	0.958	0.928	0.11	0.539
	Death	0.119	0.042	0.258	0.119	0.163	0.11	0.539
	Weighted Avg.	0.867	0.789	0.829	0.867	0.845	0.11	0.539
Extra Tree (ET)	Alive	0.962	0.881	0.899	0.962	0.93	0.119	0.573
	Death	0.119	0.038	0.276	0.119	0.167	0.119	0.573
	Weighted Avg.	0.87	0.789	0.832	0.87	0.847	0.119	0.573
Random	Alive	0.971	0.896	0.899	0.971	0.933	0.124	0.579

Forest (RF)	Death	0.104	0.029	0.304	0.104	0.156	0.124	0.579
	Weighted Avg.	0.877	0.801	0.834	0.877	0.849	0.124	0.579

TP-rate, FP-rate, accuracy, recall, F-measure, MCC, and ROC area are associated with certain measures. The performance analysis of classifiers is displayed in Table 5. Random Forest (RF) performs better than Random Tree (RT) and Extra Tree (ET) classifiers, according to the performance measures. RF's strong predictive power is demonstrated by its greatest true positive rate (0.877), F-measure (0.849), and superior Matthews Correlation Coefficient (0.124) and ROC Area (0.579). With a true positive rate of 0.87 and an F-measure of 0.847, ET performs moderately, but RT performs worse with a true positive rate of 0.867 and an F-measure of 0.845. These findings imply that RF offers more accurate and balanced categorisation.

5. Discussion

Analysis of the comparative performance of Random Tree (RT) Extra Tree (ET) and Random Forest (RF) classifiers based on the result of performance metrics reveals significant differences in classification ability. From that, the best overall performing algorithm is the Random Forest (RF) classifier. It also shows the most accurate (87.66%) and sensitivity (97.09%) of 540 'Alive' & only 7 'Death' (not 'Death') were classified correctly by (RF). RF has the highest True Positive Rate (TP rate) value for 'Alive' (0.971), indicating that it is better at detecting positive cases than RT and ET. Furthermore, RF has the highest ROC Area (0.579), showing the highest ability to separate the classes. The Extra Tree (ET) has a similar but lesser accuracy of 87.01 than RF, but is almost as strong. On the other hand, it demonstrates high TP (True Positive) rate on 'Alive' (0.962), and low misclassification rates on 'Death'. ET also showed a good ROC Area of 0.573, further confirming good class discrimination but was behind that of RF. The Random Tree (RT), only a little bit better than the baseline with an accuracy of 86.69%, still has a lower sensitivity (95.81%) compared to RF and ET TP (0.958 for 'Alive' class). It also reveals a lower ROC Area (0.539), which demonstrates that the model has a less powerful ability (compared to some lower values) in class discrimination. In summary, although all the classifiers demonstrate good predictive accuracy, RF emerges as the most suitable, followed by ET, and RT categorically shows the lowest performance.

6. Conclusions and recommendations

6.1 Conclusion

Key insights into classification algorithms performance for leukemia classification:

1. Random Forest (RF) has the maximum overall accuracy (87.66%), the maximum sensitivity (97.09%), and the maximum ROC Area (0.579). Hence, we can conclude that best model for predicting 'Alive' cases is the RF model, although it struggles with specificity for 'Death' instances.
2. Extra Tree (ET) is following on the heels with 87.01% accuracy and 96.17% sensitivity. It works well, but performance is still slightly behind RF, with RF providing better sensitivity and the ROC Area (0.573).
3. Random Tree (RT) while still effective ranks lowest with an accuracy of 86.69%. RT shows the lowest sensitivity (95.81%) and weakest ROC Area (0.539), which indicates RT struggles a bit more at classification compared to than RF and ET in classification, particularly predicting 'Alive' class.
4. The specificity is low for all models and ranges from 10.45% to 11.94%, which suggests that all the classifiers face a challenge in differentiating 'Death' instances.

So, these classifiers have achieved the most performance in classifying leukemia, but overall, Random Forest (RF) classifier is superior with the most valuable performance in terms of accuracy and sensitivity. Extra Tree (ET) has the second best accuracy whereas Random Tree (RT) achieves better results than ET in some cases, but gives worse results in most of the tests

6.2 Recommendations

Based on this conclusion, a few of recommendations are as follows for this study.

- Random Forest (RF) is the most reliable algorithm in leukemia classification since it has the best overall accuracy and sensitivity
- Extra tree (ET) can be a good option for some applications when simplicity or efficiency of a model is important, as it achieves relatively high performance at high sensitivity.

- Random Tree (RT) can be applied where resource efficiency is not a concern but may need more tuning for its sensitivity, and overall performance.
- Future work must ensure that all models are specific enough to get better classifying performance for 'Death' instances and also may look around for other algorithms or hybrid models

7. References

1. American Cancer Society. (2023). Leukemia - What is leukemia? https://www.cancer.org/cancer/leukemia.html
2. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
3. El Bilali, A., Abdeslam, T., Ayoub, N., Lamane, H., Ezzaouini, M. A., & Elbeltagi, A. (2023). An interpretable machine learning approach based on DNN, SVR, Extra Tree, and XGBoost models for predicting daily pan evaporation. Journal of Environmental Management, 327, 116890.
4. Elbeltagi, A., Pande, C. B., Kumar, M., Tolche, A. D., Singh, S. K., Kumar, A., & Vishwakarma, D. K. (2023). Prediction of meteorological drought and standardized precipitation index based on the random forest (RF), random tree (RT), and Gaussian process regression (GPR) models. Environmental Science and Pollution Research, 30(15), 43183-43202.
5. Faraj, S. M. (2024). Performance Evaluation of Machine Learning Algorithms for Predictive Classification of Anemia Data. University of Kirkuk Journal For Administrative and Economic Science, 14(3).
6. Hussein, M. M. F., Mohamad, S. H., & Qadir, B. H. (2024). Machine Learning Approaches to Classify and Predict Congenital Jaundice. Khazayin of Economic and Administrative Sciences, 52-64.
7. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8–17. https://doi.org/10.1016/j.csbj.2014.11.005
8. Lavanya, D., & Rani, K. U. (2011). Performance evaluation of decision tree classifiers on medical datasets. International Journal of Computer Applications, 26(4), 1-4.
9. Mohamad, S. H. (2025). Machine Learning Classification Methods Performance Comparison in Liver Cancer Cohort. Al-Furat Journal of Innovations in Management Sciences, 1(1), 1-11.

10. Mohamad, S. H., & Hussein, M. M. F. (2024). Diagnosing Anemia Disease using Partial Least Square (PLS) Models and Support Vector Machines (SVM). Journal of Kurdistani for Strategic Studies, 2(9).
11. Omer, A., Faraj, S. M., & Mohamad, S. H. (2023). An application of two classification methods: hierarchical clustering and factor analysis to the plays PUBG. Iraqi Journal of Statistical Sciences, 20(1), 25-42.
12. Pagliaro, A. (2023). Forecasting significant stock market price changes using machine learning: Extra trees classifier leads. Electronics, 12(21), 4551.
13. Salih, A. K., & Hussein, H. A. A. (2022). Lost circulation prediction using decision tree, random forest, and extra trees algorithms for an Iraqi oil field. The Iraqi Geological Journal, 111-127.
14. Yousefi, S., & Poornajaf, M. (2023). Comparison of result of machine learning algorithms in predicting heart disease. Frontiers in Health Informatics, 12.