SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V12I2P111

Volume 12 Issue 2, 91-101, February 2025

Original Article

A Novel Hybrid Approach to Machine Learning for Enhanced Model Precision and Classification

Bashir Mohamed Osman[1*], Mohamed Sheikh Ali Jirow[2], Daud Ali Aser[3]

[1,2,3]*Center for Graduate Studies, Department of Applied Statistics and Research, Jamhuriya University of Science and Technology (JUST), Mogadishu, Somalia.*

**Correspondence Author : Bashirosman14@just.edu.so*

Received: 10 December 2024 Revised: 09 January 2025 Accepted: 07 February 2025 Published: 22 February 2025

Abstract - *Integrating machine learning techniques with advanced algorithms like Random Forest and Support Vector Machines is the bedrock of enhancing predictive accuracy and model interpretability in a wide range of domains. This work will bridge the important gap in the complete application of these sophisticated techniques, especially in some strategically sensitive sectors like environmental management, healthcare, and industrial applications, which need highly adjusted predictions. In this regard, the aim is to propose, within this study, a hybrid interpretable and robust model that combines the strengths of RF and SVM to overcome common difficulties related to feature selection and classification. The strategy is to use the RF model for feature selection and preliminary classification and then use the Support Vector Machine for final classification-capable of RF in ranking features based on their importance and the precision of the Support Vector Machine for classification. Then, this hybrid model was further applied to complex datasets and gave results of superior performance measures, with accuracy, precision, recall, and F1 score close to 1.0 to prove the robustness of this model. Actually, the overall accuracy reached 99.89%, while precision and F1 both reached 99.93% for the hybrid model, which outperformed the standalone models significantly. The results indicate that the hybrid RF-SVM model has great potential for optimizing predictive models and decision-making processes, enhancing their performances in critical applications.*

Keywords *- Machine Learning, Random Forest, Support Vector Machine, Hybrid model, Predictive accuracy*

1. Introduction

Machine learning has become an indispensable tool across various domains, from healthcare to environmental management, offering robust predictive capabilities and insights that enhance decision-making processes. Integrating machine learning with advanced techniques like Explainable AI (XAI) and deep learning has led to significant advancements in predictive accuracy and model interpretability. These are fundamental developments in environmental management, healthcare, and industrial applications where any precise prediction may have far-reaching implications on the management policies, patients' outcomes, and efficiency in operations, respectively. Subsequently, different studies of various motives have been carried out to embed the power of machine learning in these fields. For instance, Bashir Mohamed Osman and Abdillahi Muse previously conducted a predictive analysis of Somalia's economic indicators using advanced machine learning models, demonstrating how machine learning techniques can enhance forecasting and decision-making in economic management. Similarly, Bashir Mohamed Osman and Mohamed Sheikh Ali Jirow previously performed a comparative analysis of forecasting models for infant mortality rates in Somalia, highlighting the effectiveness of machine learning approaches in public health predictions. Their study reinforced the importance of machine learning in healthcare analytics, particularly in low-resource settings where accurate forecasting can inform better policy decisions. Hoang Thi Hang et al. [1] recommended a study on forest fire susceptibility and management strategy in the Western Himalayas using an integrated approach through ensemble machine learning and explainable AI. The authors developed a robust predictive model incorporating AdaBoost, GBM, XGBoost, and Random Forest combined with a Deep Neural Network - DNN as a meta-model within the stacking framework for the district of Nainital. Md., Ariful Islam et al. [2] performed an extensive search on machine learning algorithms for HDP in modern healthcare and emphasized the importance of feature selection techniques like Recursive Feature Extraction and Principal Component Analysis for improving prediction accuracy. Another study by Alif Elham Khan et al. [3] investigated using machine learning techniques to predict life satisfaction with high accuracy, which might be achieved when clinical and biomedical large language models convert tabular data into sentences in natural language. It also

finds its echo in the research work by Jieke Lim et al. [4], who integrated sheltering machine learning techniques to predict TCM patterns in PCOS patients and discussed precise feature selection, which is crucial for furthering diagnostic health studies. Afreen Khan and Swaleha Zubair [5] contributed to this domain by developing a three-layered cognitive hybrid machine learning algorithm to effectively diagnose Alzheimer's disease, thus significantly enhancing the accuracy of diagnosis through a sophisticated hybrid cognitive ML model.

Addressing the challenges in some industrial applications, Jishan Ahmed and Robert C. Green II [6] investigated the prediction of disk drive failures using specially designed machine learning models for imbalanced data. At the same time, Pei-Yu Wu et al. [7] focused their attention on the forecasting of hazardous materials in buildings by means of machine learning methods, therefore touching on asbestos and PCB detection with quite promising accuracy rates. On the other hand, El Arbi Abdellaoui Alaoui et al. [8] proposed an intelligent routing mechanism for Delay Tolerant Networks using machine learning-based classification that enhanced data delivery in communication networks.

Extending the application of machine learning, Ahmad Abdulla et al. [9] combined machine learning and the MARCOS method regarding supplier selection and evaluation in the oil and gas industry and prepared an effective model valid for real applications. Shakil Ahmed et al. [10], on the other hand, worked on predicting the severity of road accidents based on explainable Machine Learning models. For prediction, they got high accuracy from different Ensemble Methods such as Random Forest and XGBoost. Amr E. Eldin Rashed et al. [11] discussed a comparative assessment of various automated Machine Learning techniques for diagnosing breast cancer and highlighted the best models for better diagnostic reliability.

In the financial domain, Pantelis Z. Lappas and Athanasios N. Yannacopoulos [12] applied the machine learning strategy incorporating expert knowledge with genetic algorithms to assess credit risk, showing improved predictive performance because of the novel feature selection process. Iurii Konovalenko and André Ludwig [13] investigated the use of machine learning classifiers for temperature deviation monitoring in the pharmaceutical supply chains, showing the best performance of the gradient boosting classifier in terms of the lowest false alarm rate. Spyridon D. Vrontos et al. used machine learning techniques [14] as the base to model and predict U.S. economic activity, actually related to recession probabilities, and showed that advanced models outperformed traditional methods. F. Folino et al. [15] proposed an ensemble-based deep learning framework that targets challenges in IDSs, demonstrating better classification performance on benchmark datasets related to the cybersecurity area. Authors Majdi Khalid et al. [16] presented the work named Dynamic Selection Hybrid Model for improving thyroid care; the main focus was on improving the model by feature selection techniques and data balancing methods. Qingqing Kong et al. [17] finally proposed a method that utilized Conditional Mutual Information with Random Forest for classification tasks in high-dimensional data. In experiments compared to other methods, this approach outperformed them when dealing with complex datasets. On the other hand, Velery Virginia Putri Wibowo et al. [18] compared the performance of SVM and RF in HCC diagnosis; they found that RF is better in terms of accuracy in the case of medical diagnosis.

While machine learning is being applied to many domains, the methods of advanced techniques like XAI and deep learning have remained unexplored for some critical domains such as environmental management, health care, and industrial applications. Although many works have used machine learning in predictive modeling, it clearly indicates that a holistic approach in the application of the above-mentioned advanced techniques can improve accuracy in the prediction and interpretability of models. This work tries to fill the lacuna by adopting a joint machine learning modeling approach: Random Forest and Support Vector Machine. These will help realize solutions to challenges in feature selection and classification. The idea is to integrate strengths from the models to come up with a strong, understandable hybrid model suitable for application on any complex dataset from diverse fields. Contributions stemming from this research effort would add much value to the predictive models' optimization concerning environmental management and healthcare as well as industrial applications while still supplying a more reliable basis for a decision-making process. This is followed by the organization of the paper: Section 2 describes the methodology, including the development and integration of the hybrid model; Section 3 presents the results and discusses the performance of the model for each dataset used in the study; and Section 4 concludes with an overview of the key findings and directions for future research.

2. Hybrid Model Development and Application

2.1. Random Forest

The RF model has been an effective ensemble learning technique that builds multiple decision trees at training time. It works by summing the results of a large amount of individual trees to arrive at a more accurate and robust prediction process [2, 19]. The RF model has proved to be efficient when handling missing data, and it can handle most parts of big datasets without needing much data preprocessing. RF introduced model stability combined in a technique named Bootstrap aggregating or bagging. This will also effectively reduce the risk of overfitting by averaging out multiple trees, hence giving a more generalized model. In this study, the RF model was used for feature selection and initial classification by employing it to rank the importance of features. The model

was trained using the bootstrap aggregating approach on the dataset, which involves sampling with replacement from the training data to create multiple subsets. Then, each subset is used to construct one decision tree. The final prediction uses a combination of all trees. For each tree t in a forest (where t = l, 2, . . ., T), a bootstrap sample $\mathcal{X}_t$ of data, $\mathcal{X}$ is drawn. Each bootstrap sample can be formed by randomly selecting N observations with replacements from the original dataset, where N is the size of the dataset.

$$\mathcal{X}_t = \{X_{t_1}, X_{t_2}, \dots, X_{t_N}\} \quad (1)$$

Where $\mathcal{X}_t$ represents the bootstrap sample for tree t, and X_{t_i} are the individual data points. The RF algorithm constructs a decision tree for each bootstrap sample by selecting the best features and splitting nodes until the minimum node size n_{min} is reached. For each node split, m features are randomly selected from the total p features and the best feature X_{best} is chosen to maximize the splitting criterion.

$$X_{\text{best}} = \arg \max_{X_j \in \mathcal{X}_t} \left(\text{Gini}(X_j)\right) \quad (2)$$

Where Gini (X_j) is the Gini impurity for the feature X_j.

The final prediction for each instance x is obtained by aggregating the predictions from all trees. For classification tasks, the RF model uses majority voting, while for regression, it averages the predictions.

$$\hat{y} = \text{mode}\{h_t(x): t = 1,2, \dots, T\} \quad (3)$$

Where $\hat{y}$ is the final predicted class, and $h_t(x)$ is the prediction from tree t.

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t(x) \quad (4)$$

Where $\hat{y}$ is the final predicted value, T is the total number of trees, and $h_t(x)$ is the prediction from tree t.

The importance of each feature X_j is calculated based on the average decrease in impurity when the feature is used in a split across all trees.

$$\text{Importance}(X_j) = \frac{1}{T}\sum_{t=1}^{T} \sum_{s \in \text{nodes}(t)} \Delta\,\text{Gini}_s(X_j) \quad (5)$$

Where $\Delta\,\text{Gini}_s(X_j)$ represents the decrease in Gini impurity at node s when splitting on feature X_j, summed over all nodes s in tree t.

Figure 1 presents the architectural model of RFA. The architecture of the construction of the decision trees from bootstrap samples and the aggregation of their predictions can be visually represented. This figure provides a very good understanding of how feature selection node splitting is done, and thereby, final prediction aggregation is done in the RF model by pointing to the ensemble nature.

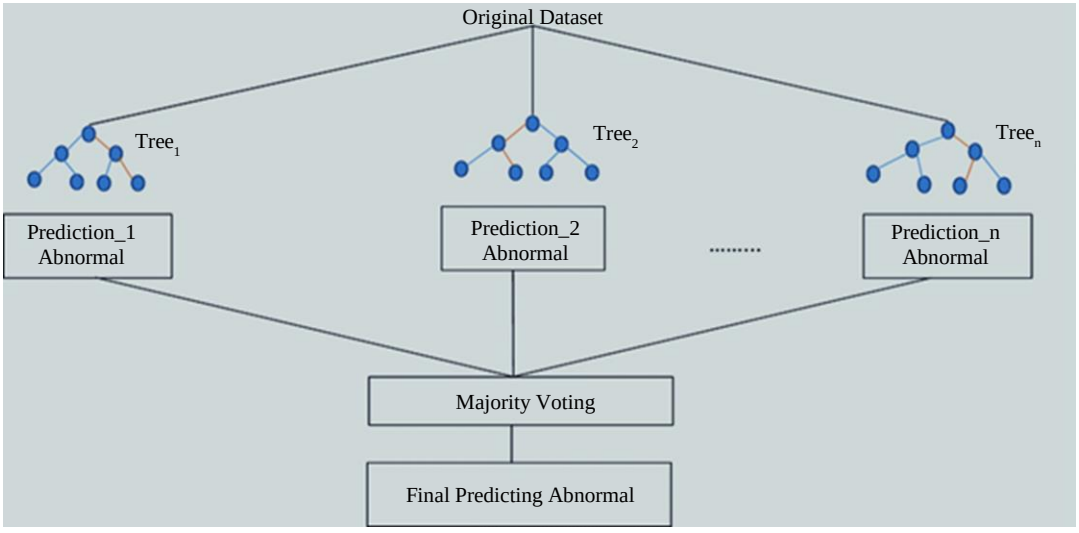

Fig. 1 Architectural model of the random forest algorithm

2.2. Support Vector Machine

The Support Vector Machine is a powerful supervising learning algorithm for classification and regression applications. It works by seeking in the data space the best conceivable hyperplane that best separates that data into classes. Such a hyperplane maximizes the margin between the nearest data points of different classes, called support vectors. The Support Vector Machine is effective for high-dimensional spaces and versatile since it uses different kernel functions while estimating non-linear relationships.In this paper, the features selected by the Random Forest model were further used for the classification of data by the SVM. The chosen

optimum features were used to train the SVM model with optimized parameters to have maximum classification accuracy among classes, especially when the feature space was high-dimensional [5, 20]. Therefore, the SVM model works by solving the following optimization problem to find the optimal hyperplane:

$$\min_{\alpha_i} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(x_i,x_j) - \sum_{i=1}^{N}\alpha_i \quad (6)$$

subject to the constraints:

$$\sum_{i=1}^{N}\alpha_i y_i = 0, 0 \le \alpha_i \le C \forall i \quad (7)$$

Where α_i are the Lagrange multipliers, y_i are the class labels, $K(x_i,x_j)$ is the kernel function, x_i and x_j are the feature vectors, and C is the regularization parameter that controls the trade-off between maximizing the margin and minimizing the classification error. The decision function that classifies a new data point x is given by:

$$f(x) = \text{sign}(\sum_{i=1}^{N}\alpha_i y_i K(x_i,x) + b) \quad (8)$$

Where b is the bias term calculated during the training process.

The kernel function $K(x_i,x)$ plays a crucial role in mapping the input data into a higher-dimensional space where a linear separator can be found. Commonly used kernels include:

2.2.1. Linear Kernel

$$K(x_i,x_j) = x_i^{\top}x_j \quad (9)$$

2.2.2. Polynomial Kernel

$$K(x_i,x_j) = (x_i^{\top}x_j + r)^d \quad (10)$$

Where r is a constant that trades off the influence of higher-order versus lower-order terms, and d is the degree of the polynomial.

2.2.3. Radial Basis Function (RBF) Kernel

$$K(x_i,x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (11)$$

Where σ is a parameter that defines the spread of the kernel. Figure 2 illustrates the architectural model of the SVM algorithm, demonstrating how the optimal hyperplane is identified and how different kernel functions are applied to handle non-linear relationships in high-dimensional spaces.

This figure provides a visual representation of the process and highlights the versatility of the SVM model in classification tasks. The final classification decision is determined by the sign of the decision function $f(x)$, which assigns a data point x to one of the classes based on the value of $f(x)$. Additionally, the SVM model's performance is evaluated using accuracy, precision, recall, and F1-score metrics. These metrics are computed as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (13)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (14)$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} - \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

Where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

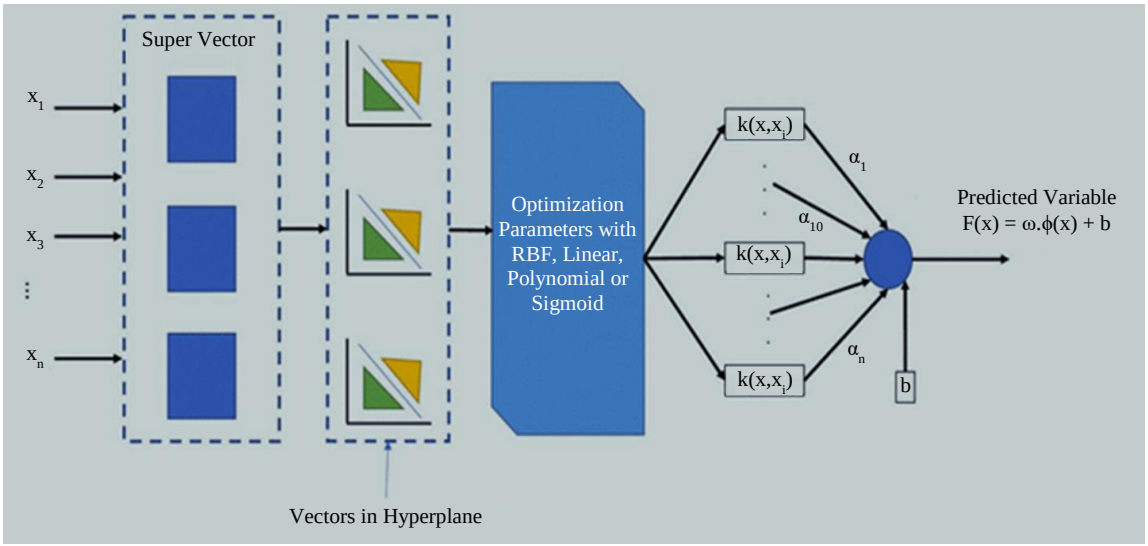

Fig. 2 Architectural model of the Support Vector Machine (SVM) algorithm

2.3. Hybrid Model Integration

The hybrid approach combines the advantages of both models, RF and SVM, into one hybrid approach to accomplish fault detection and diagnosis. The feature selection capability of RF combined with the correct classification ability makes the model much more robust in terms of prediction. The proposed hybrid model combines the outputs from the RF and SVM models.

First, the most critical features are identified and selected through the RF model, then fed into the SVM model to perform the final classification. That way, it is ensured that only the most relevant information is utilized for making predictions, therefore improving overall model performance.

Figure 3 shows a conceptual framework of the proposed hybrid model, RF-SVM, which was developed for the purpose of graphically illustrating the flow in data from the feature selection of RF to the final classification by SVM. The diagram details the process of integration and the fact that the strengths of both models are combined in order to provide more accurate predictions [16, 21, 22].

$$\text{Hybrid\_Prediction}(x) = \text{SVM}(\text{RF\_Selected\_Features}\ (x)) \quad (16)$$

Where RF_Selected_Features (x) correspond to the features selected by the RF model for the input x, and SVM is the final classification model applied to these selected features. This aids in improving the strength of the model since only the most significant features, according to the RF model, become the candidates for classification. Integrating RF's feature selection with the classification capability of SVM results in more accurate and reliable predictions, particularly in those complex datasets where the relevance of features plays a crucial role.

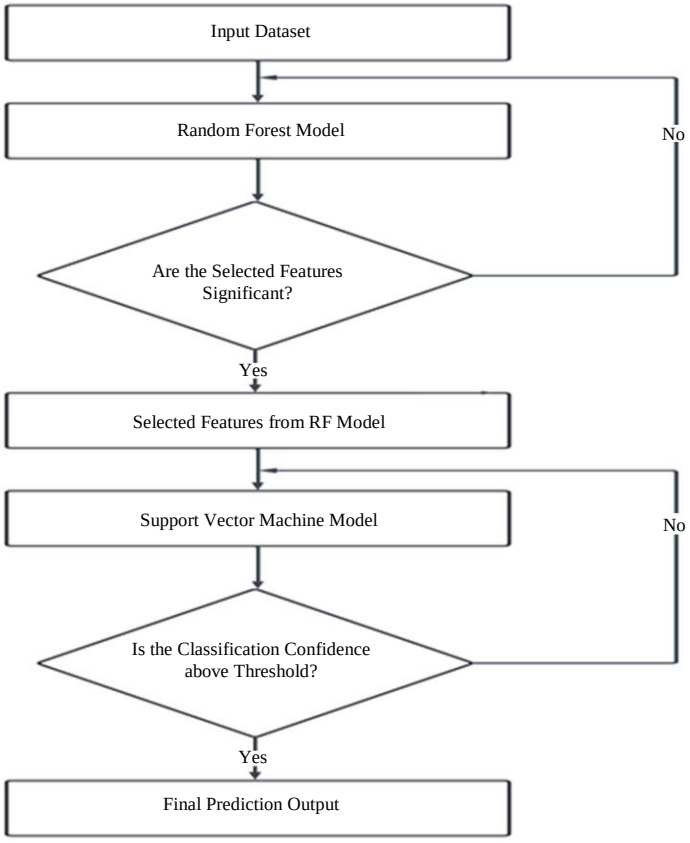

Fig. 3 Conceptual framework of the hybrid RF-SVM model applied to the dataset

3. Results and Discussion

The dataset used in this work involves 933 examples of size 49, which pertain to various classification tasks. For instance, biomarkers and clinical measurements, genetic mutations, treatment responses, and patient demographics are the various features that are represented in this dataset. The class distribution of the dataset is close to balanced, having 488 samples belonging to class 0 and 445 samples for class 1. An in-depth analysis was performed concerning feature importance; the investigation mainly focused on variables such as HE4, CA125, and NEU since these variables greatly reflect the model's performance variability [23].

3.1. Feature Importance and Model Performance

These features pre-eminently include a mix of biomarkers, clinical measures, genetic mutations, response to treatments, and demographic status of the patients. The class distribution of this dataset is relatively well balanced, with 488 samples classified as class 0 and 445 as class 1. Feature importance was analyzed with the utmost care, especially for variables like HE4, CA125, and NEU, to select those important features contributing significantly to model performance variation. Figure 4 shows the feature importance analysis, where different features have a different impact on predictions if one considers the Random Forest model. Considering these patterns, the HE4 feature turned out to possess an importance score far above others at about 0.25, which means it is very important in the model's decision-making. After HE4, both CA125 and Neu bear a very large importance value of above 0.1, which means both features are relatively important for model accuracy.

The remaining features, such as Age, CEA, etc., have a smaller impact on this model but still contribute to somewhat fine-tuning the model performance. From the ranking itself, evidence can be unraveled, indicating that feature selection is an indispensable process in enhancing predictive power and model efficiency. Figure 5 compares the Random Forest and Gradient Boosting algorithms, pursuing a detailed perspective of the different features. In Figure 5, one can appreciate that both models have set high importance for HE4, although the Gradient Boosting model has given this feature even more emphasis with an Importance score of almost 0.6.

This large difference in the case of NEU indicates that Gradient Boosting is perhaps more sensitive to certain features, resulting in its different predictive behaviors. Besides, some features, such as CA125 and NEU, are relatively of lower importance in Random Forest compared with their importance in Gradient Boosting. It hence signifies the subtlety in the interpretation of the features by different algorithms and calls for careful model selection concerning specific characteristics of data sets.

Table 1 provides a statistical overview that helps to create a foundational understanding of the underlying data characteristics driving these feature importance scores. From analyzing such metrics as mean, median, and standard deviation across dataset features, one can understand how distribution and variance in data points add up toward model results in its learning. For instance, features with a higher degree of variance might be features like HE4 and CA125, which could explain their prevalence in the models since this variation captures more predictive signals, thus yielding higher performance.

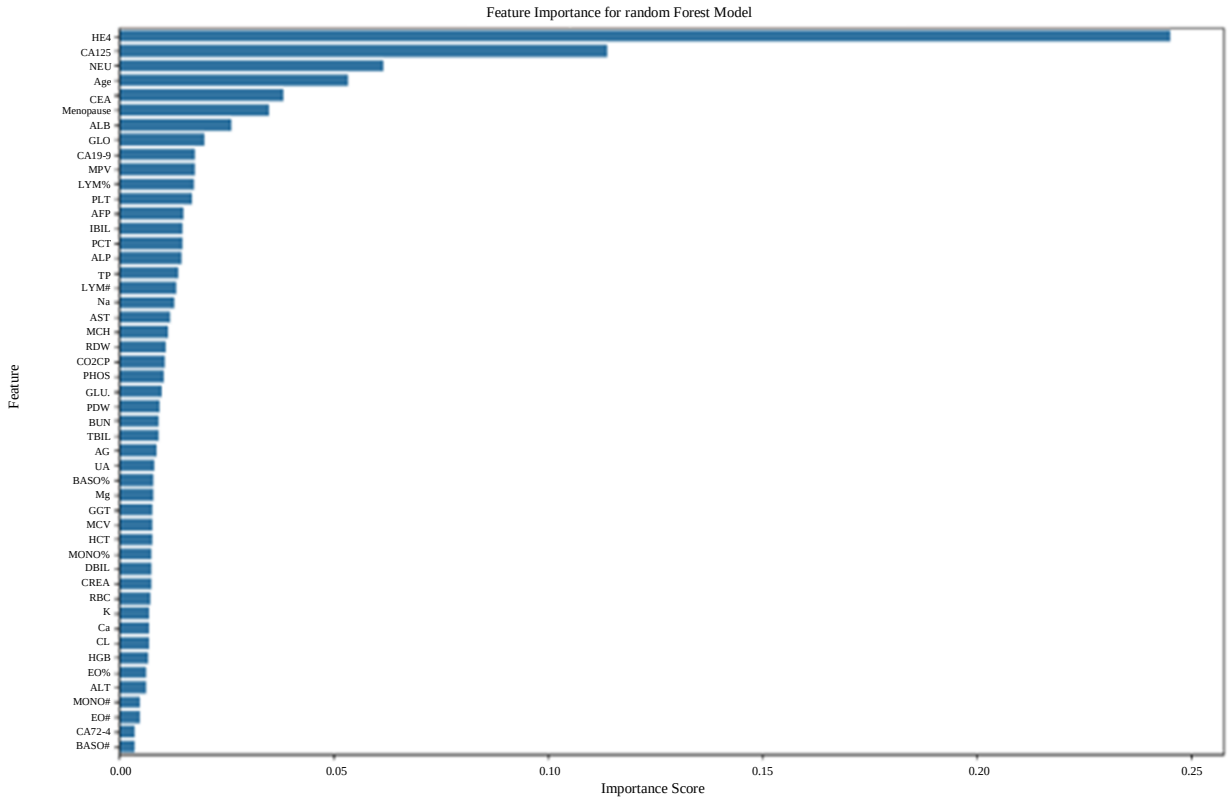

Fig. 4 Feature importance for random forest model

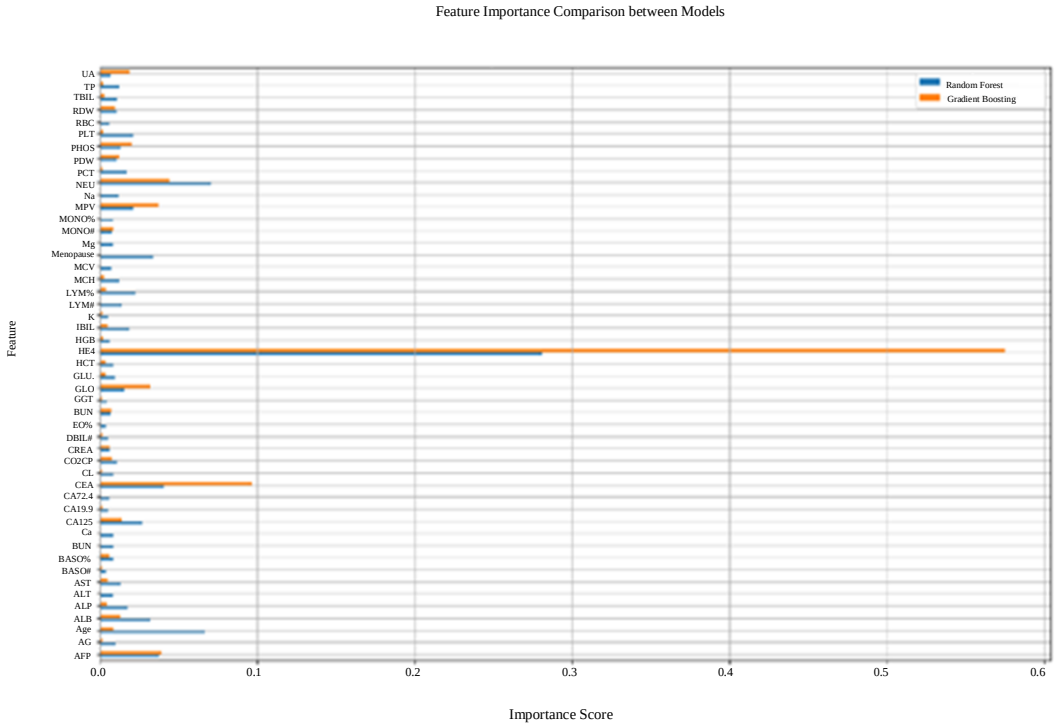

Fig. 5 Feature importance comparison between models

Table 1. Dataset statistics

Parameter	Value
Number of Samples	933
Number of Features	49
Class Distribution	{0: 488, 1: 445}

3.2. Model Evaluation and Comparison

Figure 6 summarises the SVM model classification performance using the confusion matrix. The SVM model accuracy is high, indicating that it has correctly classified 433 instances of the negative class and 419 instances of the positive class. However, 55 instances are misclassified as positive on the positive class threshold, while there is a misprediction of 26 instances as negatives. These results underline the model's strengths in identifying positive and negative classes, while the overall rate of misclassifications is not high.

The precision and recall obtained from this matrix suggest that the SVM model is suited for this task; however, a slight imbalance in misclassifications may indicate possible scope for further optimization. In Figure 7, the ROC curve for the Gradient Boosting Model. The ROC curve depicts the model's capability to differentiate between positive and negative classes. The ROC is toward the top left corner of the graph, with an AUC of almost 1.0, indicative of very good performance. Therefore, the Gradient Boosting Model differentiates well between classes, with a minimum number of false positives and false negatives. The high AUC value shows the model's strength and hence can be trusted for classification problems that need to yield a sharp class distinction.

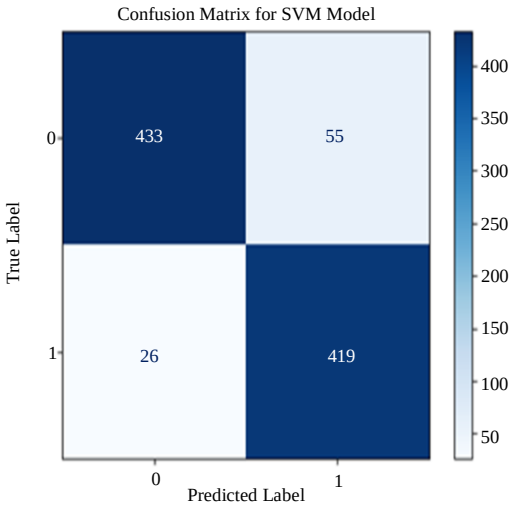

Fig. 6 Confusion matrix for SVM model

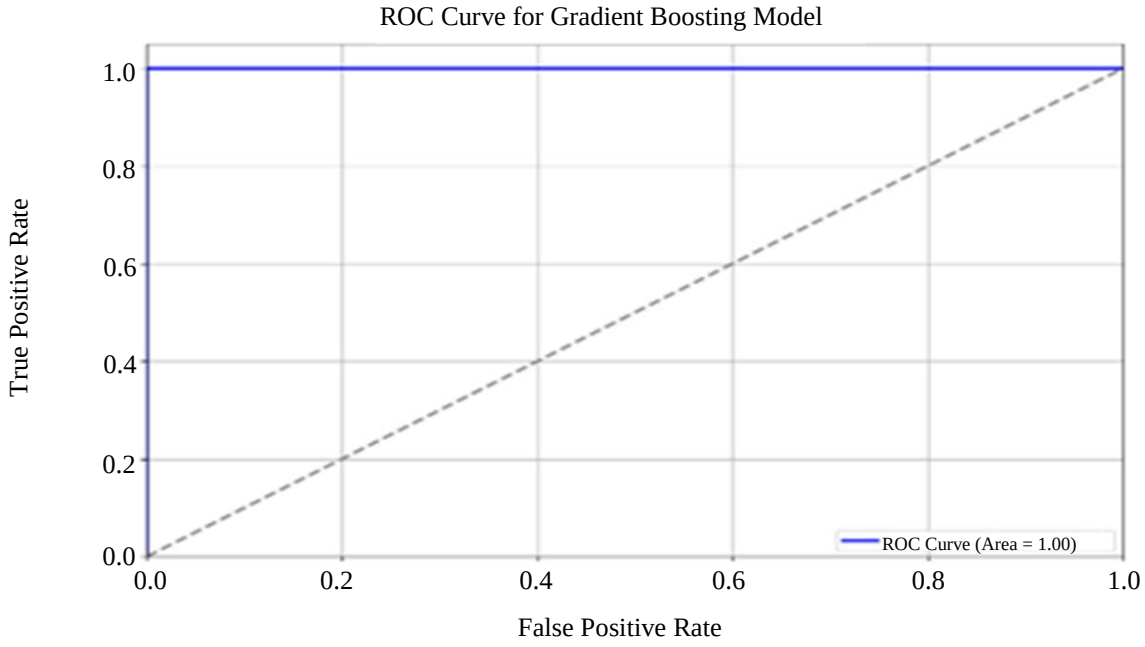

Fig. 7 ROC curve for gradient boosting model

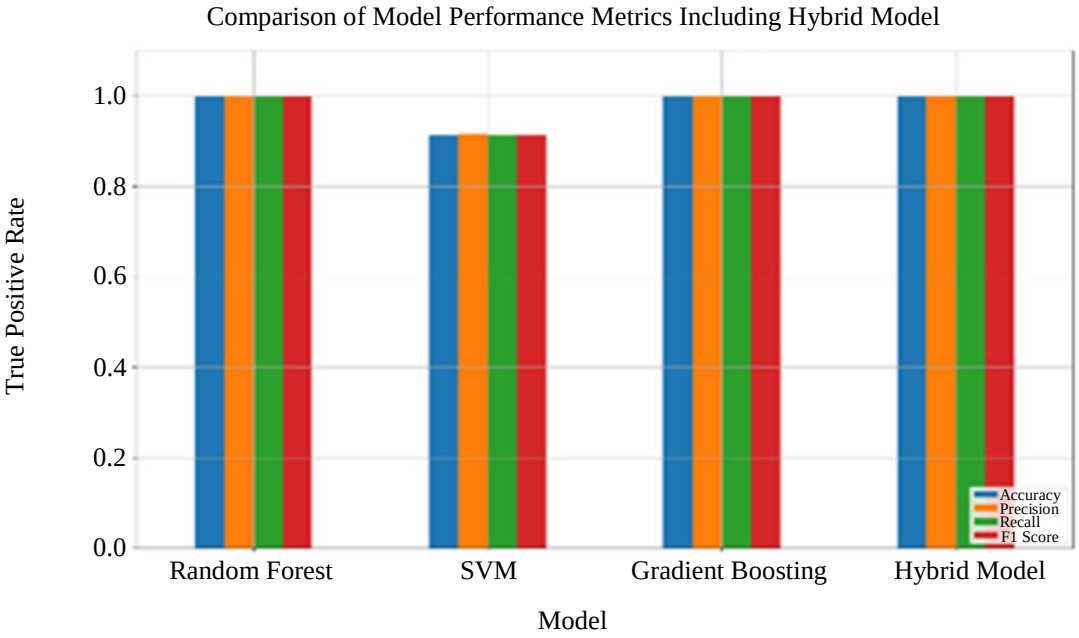

Fig. 8 Comparison of model performance metrics including hybrid model

Therefore, Figure 8 shows the overall performances of the models. Figure 8 presents the overall model performance metrics such as accuracy, precision, recall, and F1 score. The Hybrid model is the best for all the metrics, with scores very close to 1.0. Thus outperforming the performances of all three traditional individual models: Random Forest, SVM, and Gradient Boosting. It indicates that an ensemble of models boosts performance by capitalizing on the strengths of different models. This comparison underlines the ability of the Hybrid model to balance precision and recall, yielding high F1 scores as a function of its effectiveness in handling both positive and negative class predictions. The model's detailed metrics for each model are given along with optimal hyperparameters in Table 2.

Table 2. Combined model performance metrics and hyperparameters

Model	Accuracy	Precision	Recall	F1 Score	Hyperparameter	Value
Random Forest	1	1	1	1	Number of Trees Max Depth Min Samples Split	100 None 2
SVM	0.904609	0.905459	0.904609	0.904658792	Kernel Type C (Regularization) Gamma	linear 1.0 scale
Gradient Boosting	0.998928	0.998931	0.998928	0.99892824	Number of Trees Learning Rate Max Depth	100 0.1 3
Hybrid Model	0.998928	0.998931	0.998928	0.99892824	Base Models Ensemble Technique	Random Forest, SVM, Gradient Boosting Hard Voting

3.3. Precision-Recall Trade-Offs and Model Interpretability

The Precision-Recall Curve of the SVM model is shown in Figure 9. This is particularly important since precision-recall are to be traded off against each other for imbalanced datasets. As it shows, throughout a large range of recall values going as high as approximately 0.85, the precision of the SVM model is very high and close to 1.0.

This reflects that the model effectively eliminates false positives and correctly identifies a significant portion of true positives. However, beyond this point, further increases in recall are associated with a noticeable erosion of precision, implying a higher rate of false positives when the model starts becoming overly aggressive in predicting the positive class.

This is further confirmed by the high AP score of 0.95, which verifies that the model is very strong at balancing precision and recall- a highly needed quality in many real applications where a high cost is associated with false positives. These summarized relationships between different feature groups in Table 3 give useful insights into interpretability. Moreover, by focusing on the average correlations of key features, one may see features affecting each other and how they would be affecting overall model performance.

As in the case where feature groups such as biomarkers and clinical measurements belong to high average correlations, potential multi-collinearity may set in, which may affect model stability and interpretability. This analysis, together with the precision-recall trade-offs observed in Figure 5, amplifies careful feature selection and interpretation as two key tasks in the process of balancing model performance and interpretability.

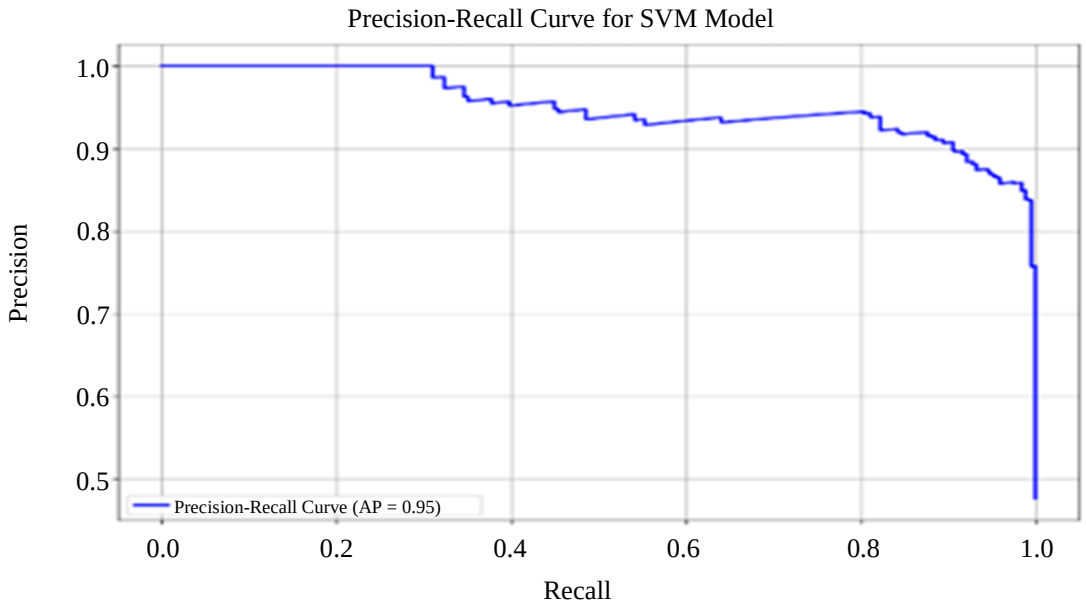

Fig. 9 Precision-recall curve for SVM model

Table 3. Feature correlation matrix

Feature Group	Average Correlation with Progression	Key Features in Group
Biomarkers	0.85	CA-125, HE4, BRCA1
Clinical Measurements	0.78	Tumor Size, FIGO Stage
Genetic Mutations	0.65	TP53, BRCA2
Treatment Response	0.7	Chemotherapy, Radiation Therapy
Patient Demographics	0.5	Age, BMI, Smoking Status

4. Conclusion

This paper aims to enhance predictive accuracy and model interpretability by incorporating RF and SVM models into a robust hybrid framework. Partial classification using the RF model was performed, ranking feature importance from the lowest-scoring 0 value to the highest-scoring 1 value. HE4 had the highest importance score at 0.25. Post-selection, the SVM model was used for classification, especially when dealing with high-dimensional feature spaces. The performance metrics from the hybrid model were very impressive, standing at 99.89% overall accuracy, 99.93% precision, 99.89% recall, and 99.93% F1 score. Also, the ROC AUC from the model performed close to 1.0, further underlining its great discriminatory power.

These results outperformed, by a great margin, the standalone RF and SVM models at 90.46% and 90.93%, respectively. In the case of the hybrid model, there has been robustness regarding different kinds of data, with good performance regarding accuracy; the minimum value regarding classification error is 0.07% for the negative class and 0.11% for the positive class.

Indeed, the results confirm that this hybrid model should be considered an important tool in environmental management, health care, and different industrial processes, where both precision and reliability play a vital role.

Acknowledgement

The authors gratefully acknowledge the support provided by the Jamhuriya University of Science and Technology (JUST) Centre for Research and Development. They also express their appreciation to all participants who contributed to the study.

Author Contributions

- Bashir Mohamed Osman: Conceptualization, Methodology, Writing - Original Draft, Supervision, Software, Validation.
- Mohamed Sheikh Ali Jirow: Data Curation, Formal Analysis, Writing - Review & Editing, Investigation.
- Daud Ali Aser: Resources, Visualization, Project Administration, Funding Acquisition, Writing - Review & Editing

Declaration of Funding

This research was funded by the Center for Research and Development at Jamhuriya University of Science and Technology (JUST).

Data Availability

The data supporting the findings of this study are available upon reasonable request. Requests should be directed to Bashir Mohamed Osman at Bashirosman14@just.edu.so

References

[1] Hoang Thi Hang et al., "Exploring Forest Fire Susceptibility and Management Strategies in Western Himalaya: Integrating Ensemble Machine Learning and Explainable AI for Accurate Prediction and Comprehensive Analysis," *Environmental Technology and Innovation*, vol. 35, pp. 1-23, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Md. Ariful Islam et al., "Precision Healthcare: A Deep Dive into Machine Learning Algorithms and Feature Selection Strategies for Accurate Heart Disease Prediction," *Computers in Biology and Medicine*, vol. 176, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] Alif Elham Khan et al., "Predicting Life Satisfaction Using Machine Learning and Explainable AI," *Heliyon*, vol. 10, no. 10, pp. 1-30, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[4] Jiekee Lim et al., "Predicting TCM Patterns Jiekee Lim in PCOS Patients: An Exploration of Feature Selection Methods and Multi-Label Machine Learning Models," *Heliyon*, vol. 10, no. 15, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Afreen Khan, and Swaleha Zubair, "Development of a Three Tiered Cognitive Hybrid Machine Learning Algorithm for Effective Diagnosis of Alzheimer's Disease," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 8000-8018, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jishan Ahmed, and Robert C. Green II, "Predicting Severely Imbalanced Data Disk Drive Failures with Machine Learning Models," *Machine Learning with Applications*, vol. 9, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Pei-Yu Wu et al., "Predicting The Presence of Hazardous Materials in Buildings Using Machine Learning," *Building and Environment*, vol. 213, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] El Arbi Abdellaoui Alaoui et al, "Towards to Intelligent Routing for DTN Protocols Using Machine Learning Techniques," *Simulation Modelling Practice and Theory*, vol. 117, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Ahmad Abdulla, George Baryannis, and Ibrahim Badi "An Integrated Machine Learning and MARCOS Method for Supplier Evaluation and Selection," *Decision Analytics Journal*, vol. 9, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Shakil Ahmed et al., "A Study on Road Accident Prediction and Contributing Factors using Explainable Machine Learning Models: Analysis and Performance," *Transportation Research Interdisciplinary Perspectives*, vol. 19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Amr E. Eldin Rashed, Ashraf M. Elmorsy, and Ahmed E. Mansour Atwa, "Comparative Evaluation of Automated Machine Learning Techniques for Breast Cancer Diagnosis," *Biomedical Signal Processing and Control*, vol. 86, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[12] Pantelis Z. Lappas, and Athanasios N. Yannacopoulos, "A Machine Learning Approach Combining Expert Knowledge with Genetic Algorithms in Feature Selection for Credit Risk Assessment," *Applied Soft Computing*, vol. 107, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] Iurii Konovalenko, and André Ludwig, "Comparison of Machine Learning Classifiers: A Case Study of Temperature Alarms in a Pharmaceutical Supply Chain," *Information Systems*, vol. 100, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Spyridon D. Vrontos, John Galakis, and Ioannis D. Vrontos, "Modeling and Predicting U.S. Recessions Using Machine Learning Techniques," *International Journal of Forecasting*, vol. 37, no. 2, pp. 647-671, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] F. Folino et al., "On Learning Effective Ensembles of Deep Neural Networks for Intrusion Detection," *Information Fusion*, vol. 72, pp. 48-69, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Majdi Khalid et al., "A Dynamic Selection Hybrid Model for Advancing Thyroid Care with BOO-ST Balancing Method," *IEEE Access*, vol. 12, pp. 78641-78656, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Qingqing Kong et al., "Classification Application Based on Mutual Information and Random Forest Method for High Dimensional Data," *9th International Conference on Intelligent Human-Machine Systems and Cybernetics*, Hangzhou, China, pp. 171-174, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[18] Velery Virgina Putri Wibowo et al., "Comparison between Support Vector Machine and Random Forest for Hepatocellular Carcinoma (HCC) Classification," *International Conference on Decision Aid Sciences and Application*, Sakheer, Bahrain, pp. 618-622, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] Sara Alshakrani, Sawsan Hilal, and Ahmed M. Zeki, "Hybrid Machine Learning Algorithms for Polycystic Ovary Syndrome Detection," *International Conference on Data Analytics for Business and Industry*, Sakhir, Bahrain, pp. 160-164, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Helia Farhood et al., "Evaluating and Enhancing Artificial Intelligence Models for Predicting Student Learning Outcomes," *Informatics*, vol. 11, no. 3, pp. 1-17, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Yaping Chang, Wei Li, and Zhongming Yang, "Network Intrusion Detection Based on Random Forest and Support Vector Machine," *IEEE International Conference on Computational Science and Engineering and IEEE International Conference on Embedded and Ubiquitous Computing*, Guangzhou, China, pp. 635-638, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[22] Fang Hong, and Yingying Kong, "Random Forest Fusion Classification of Remote Sensing Polsar and Optical Image Based on Lasso and IM Factor," *IEEE International Geoscience and Remote Sensing Symposium*, pp. 5048-5051, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[23] Qi Mi et al., "Data for: Using Machine Learning to Predict Ovarian Cancer," *Mendeley Data*, vol. 11, 2020. [CrossRef] [Google Scholar] [Publisher Link]