Feature Selection Strategies in Machine Learning for Acute Respiratory Distress Syndrome: Enhancing Diagnosis, Risk Prediction, and Management

Pawan Kumar Mall[1], Prof Dr Divya Midhun[2], Dr. Rupali Atul Mahajan[3]

[1] Lincoln University ; Lincoln University[2] ; [3] Vishwakarma Institute of Technology, Pune

Email ID pawankumar.mall@gmail.com[1];divya@lincoln.edu.my; rupali.mahajan@viit.ac.in

---

Abstract: Acute Respiratory Distress Syndrome (ARDS) is an acute condition characterized by rapid lung inflammation and compromised oxygen exchange, frequently occurring in critically ill patients. It is vital that early diagnosis and risk stratification are made to enhance outcomes. Feature selection techniques in machine learning (ML) are investigated in this study to improve the diagnosis, risk estimation, and management of ARDS. A set of 1,000 ICU patients with 22 clinical and demographic characteristics was evaluated through a hybrid Genetic Algorithm (GA) and Permutation Importance feature selection approach. Six machine learning models—Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine, Neural Network, and Naive Bayes—were compared before and after feature selection. The hybrid methodology enhanced model performance for all metrics significantly. feature selection, models like Logistic Regression, Random Forest, and Gradient Boosting showed considerable improvement in accuracy (from 84.12% to 93.59%, 75.98% to 92.50%, and 75.87% to 93.04%, respectively) and F1 scores (up to 93.36%, 92.28%, and 92.72%). The results underscore the imperative value of smart feature selection in enhancing ML model performance and clinical usefulness for ARDS care.

Keywords: Machine Learning; Acute Respiratory Distress Syndrome; Risk Prediction; Diagnosis; Random Forest; Support Vector Machine

---

Introduction:

Acute Respiratory Distress Syndrome (ARDS) is a serious lung condition involving rapid development of diffuse inflammation in the lungs, resulting in deranged oxygen exchange [1]. It usually happens in critically ill patients, frequently as a complication of other critical illness such as sepsis, pneumonia, or trauma. According to the Cleveland Clinic, ARDS leads to fluid leaking into the lungs' air sacs (alveoli), causing difficulty breathing and lowered oxygen in the blood. Severe shortness of breath, fast breathing, and low oxygen levels characterize the symptoms [2]. Treatment aims at reversing the underlying cause, oxygen therapy, and lung support, which usually necessitates mechanical ventilation in the ICU. There is a need for early identification and management to enhance outcomes [3].

ARDS is a serious lung disorder with the rapid development of diffuse inflammation in the lungs, which decreases the process of oxygen exchange. It most commonly affects critically ill patients, usually as a complication of severe diseases such as sepsis, pneumonia, or trauma. The Cleveland Clinic explains that ARDS makes fluid leak into the air sacs of the lungs (alveoli), causing breathing to become hard and lowering the levels of oxygen in the blood [4]. It has symptoms such as extreme shortness of breath, fast breathing rate, and decreased oxygen saturation. Treatment involves treating the underlying condition,

oxygen therapy, and the support of lung function, which is often done with mechanical ventilation in an ICU environment. Early detection and treatment are crucial for a better outcome [5].

Definition and Diagnostic Criteria

ARDS is diagnosed using the Berlin Definition (2012), which includes:

- Timing: Acute onset or worsening of respiratory symptoms within 1 week of a known clinical insult (e.g., sepsis, trauma, pneumonia).
- Chest Imaging: Bilateral opacities on chest X-ray or CT scan, consistent with pulmonary edema, not fully explained by cardiac failure or fluid overload.
- Oxygenation Impairment:
- Mild: PaO_2/FiO_2 ratio of 200–300 mmHg with PEEP ≥ 5 cmH_2O.
- Moderate: PaO_2/FiO_2 ratio of 100–200 mmHg with PEEP ≥ 5 cmH_2O.
- Severe: PaO_2/FiO_2 ratio ≤ 100 mmHg with PEEP ≥ 5 cmH_2O.
- Exclusion of Cardiac Cause: No evidence of left atrial hypertension or congestive heart failure as the primary cause (e.g., confirmed by echocardiography or pulmonary artery catheter).

Related work: In this section we will be discuss about the previous work done in ARDS domain. [6] Jiang et al. (2024) sought to create a machine learning model for the early prediction of Acute Respiratory Distress Syndrome (ARDS) in septic ICU patients. The research employed a broad spectrum of clinical information, including demographics, diagnoses, complications, and laboratory results. Eight machine learning algorithms were compared, and ten features were identified as most important using the LASSO method. The Gaussian Naive Bayes (GaussianNB) model was found to have the best performance and was assessed using measures like AUC, accuracy, and SHAP values for explainability. The model had an AUC of 0.781 and accuracy of 78.6%, better than existing approaches in terms of both performance and explainability.

[7] Rubulotta et al. (2024) performed a narrative review that examined the use of machine learning tools in ARDS detection and prediction among ICU patients. The review emphasized the utilization of clinical and imaging information, including vital signs and laboratory results, to identify risk patterns and facilitate early clinical interventions. It presented an overview of how ML can improve precision medicine strategies and early diagnosis and prognosis in ARDS conditions. Although informative, the review was missing original model formulation and validation that would have enhanced its empirical rigor.]

[8] He et al. (2025) conducted a systematic review and meta-analysis comparing artificial intelligence (AI) models with logistic regression for predicting mortality in ARDS. The article summarized findings from eight published studies and applied a bivariate mixed-effects model to assess performance measures like sensitivity, specificity, and the summary receiver operating characteristic (SROC) curve. AI models were found to outperform logistic regression (SROC: 0.84 vs. 0.81), with high sensitivity (0.89), especially for moderate to severe ARDS cases. Nonetheless, the small number of studies included and heterogeneity of data limited the generalizability and real-time applicability of the results.

[9] Mu et al. (2025) sought to build and validate a machine learning-based mortality prediction model for sepsis-associated ARDS patients using the MIMIC-III database. The retrospective study involved 2466

patients and employed the Boruta algorithm to choose 24 informative features. Seven ML algorithms were compared, with the highest AUC of 0.8015 being achieved by the random forest model. The model recognized important clinical predictors like blood urea nitrogen and age and promises to assist clinical decision-making. The retrospective nature of the study and use of a single dataset without external validation limit its wider generalizability.

[10] Zhou et al. (2024) developed a deep learning model for early ARDS prediction and lung CT segmentation on a multicenter dataset of 928 ICU patients. A UNETR model, augmented with MONAI-based data augmentation, was implemented and compared to a Densenet-based model. The performance metrics were Dice coefficient for segmentation and AUC for prediction, with the results demonstrating better performance (AUC: 0.916 internally and 0.876 in a prospective cohort). The interpretability of the model was improved using Shapley plots. Although it was a strong performer and generalizable, its dependence on CT scans and high computational requirements can be limiting in a real-world application.

[11] Ding et al. (2024) aimed to evaluate the prediction capability of dynamic clinical indices for ARDS mortality by applying machine learning. Based on the ARDSNet FACTT trial database (n=1000), the study used a random forest model to contrast baseline and day 3 clinical measurements. Results indicated that day 3 data inclusion improved the predictive performance dramatically (AUC: 0.84 vs. 0.72 at baseline), underlining the utility of dynamic monitoring in the ICU. However, the study considered only nine clinical parameters and was based on retrospective trial data, potentially impacting its generalizability and precluding causal interpretation.

Table 1. Highlighting advancements, methodologies, advantages, and limitations of AI-driven approaches in ARDS.

Reference	Objective	Methodology	Advantages	Limitations
[6] Jiang et al. (2024)	Develop ML model for early prediction of ARDS in sepsis ICU patients	Used clinical data (demographics, labs, etc.), tested 8 ML models, selected 10 features via LASSO, final model built with GaussianNB, evaluated via AUC, accuracy, etc., used SHAP for interpretability	GaussianNB outperformed others (AUC: 0.781); identified key predictors; interpretable; better than past models	Focused only on sepsis-related ARDS; 10% test set is small; single-center data limits generalizability
[7] Rubulotta et al. (2024)	Review ML tools for ARDS detection/prediction in ICU	Narrative review of ML use in ARDS based on clinical and imaging data, highlighted risk	Broad clinical insights; underlines importance of ML in precision	No model development/validation; lacks quantitative rigor; limited discussion on model bias

		patterns and intervention timing	medicine and early ARDS detection	
[8] He et al. (2025)	Compare AI vs logistic regression for ARDS mortality prediction	Systematic review/meta-analysis of 8 studies; applied QUADAS-2; used bivariate mixed-effects model with SROC, sensitivity/specificity	AI showed superior performance (SROC: 0.84, Sensitivity: 0.89); robust across varying ARDS severity	Only 8 studies included; heterogeneity impacts accuracy; lacks real-time validation
[9] Mu et al. (2025)	Build ML model for mortality prediction in sepsis-associated ARDS using MIMIC-III	Retrospective analysis (n=2466); used Boruta for feature selection; tested 7 ML models, Random Forest best; evaluated via AUC, sensitivity, etc.	High AUC (0.8015); identified key clinical predictors; useful for decision-making	Retrospective; MIMIC-III data limits generalizability; no external validation
[10] Zhou et al. (2024)	Develop DL framework for lung CT segmentation and early ARDS prediction	Multicenter (n=928); developed UNETR model with MONAI augmentation; compared with Densenet; used Dice coefficient and AUC; Shapley plots for interpretation	High segmentation (DC: 0.734) and prediction accuracy (AUC: 0.916); better than Densenet; generalizes across cohorts	Computationally intensive; CT-based prediction may reduce accessibility; data may not reflect latest trends
[11] Ding et al. (2024)	Assess dynamic clinical indices for ARDS mortality via ML	Retrospective (n=1000, ARDSNet FACTT Trial); used Random Forest on baseline vs Day 3 data; evaluated via AUC; feature importance analysis	Day 3 data improved prediction (AUC: 0.84); highlighted dynamic indicators for early risk stratification	Single trial data; only 9 parameters used; retrospective design limits causality

Method, Experiments and Results

Dataset: The ARDS (Acute Respiratory Distress Syndrome) dataset contains 1,000 patient records, each characterized by 22 clinical and demographic features. The data comprises simple patient information such as age, sex, and BMI, in addition to risk factors and comorbidities such as smoking status, hypertension, diabetes, COPD, and cardiovascular diseases. Important vital signs and lab values are captured, such as oxygen saturation, PaO2/FiO2 ratio, blood pressure measurement, respiratory and heart rates, CRP level, D-dimer level, and lactate level—all of which are pertinent to the diagnosis and severity of ARDS. The data also include type of ventilation support received (none, non-invasive, or invasive), length of stay in the ICU, and mortality status, whether the patient survived or not.

Table 2: Dataset description

Features	Description	Data Type	Example Value
Age	Patient age	Integer	68
Sex	0 = Female, 1 = Male	Integer	1
BMI	Body Mass Index	Float	27.4
Smoking_Status	0 = Never, 1 = Former, 2 = Current	Integer	1
Hypertension	1 = Yes, 0 = No	Integer	1
Diabetes	1 = Yes, 0 = No	Integer	0
COPD	1 = Yes, 0 = No	Integer	0
Chronic_Kidney_Disease	1 = Yes, 0 = No	Integer	0
Cardiovascular_Disease	1 = Yes, 0 = No	Integer	0
Liver_Disease	1 = Yes, 0 = No	Integer	0
Oxygen_Saturation	Measured in percentage	Float	96.5
PaO2_FiO2_Ratio	Measure of respiratory efficiency	Float	105.7
Blood_Pressure_Systolic	mmHg	Integer	93
Blood_Pressure_Diastolic	mmHg	Integer	94
Heart_Rate	Beats per minute	Integer	108
Respiratory_Rate	Breaths per minute	Integer	22
CRP_Level	C-reactive protein level (mg/L)	Float	9.1
D_Dimer	Blood clot indicator (mg/L)	Float	1.02
Lactate_Level	mmol/L	Float	3.66
Ventilation_Type	0 = None, 1 = Non-invasive, 2 = Invasive	Integer	0
ICU_Length_of_Stay	In days	Integer	11
Mortality	1 = Deceased, 0 = Survived	Integer	1

Pre-processing:

The visualisation gives an overall figure 1 of three prominent ICU vital signs: systolic arterial blood pressure (SysABP), temperature, and white blood cell count (WBC). Time series plots indicate different patient-level trends and the histograms and box plots indicate central points and outliers. SysABP values are generally between 100–140 mmHg with a normal distribution of 120 mmHg and sporadic hypertensive outliers. Temperature values cluster around 37°C, although outliers indicate periods of fever or hypothermia. WBC counts are more variable, with a right-skewed distribution and a number of high outliers above 20 K/uL, which may indicate infections or inflammatory reactions. In general, although most values lie within normal clinical ranges, significant deviations reflect the physiological stress of some ICU patients.

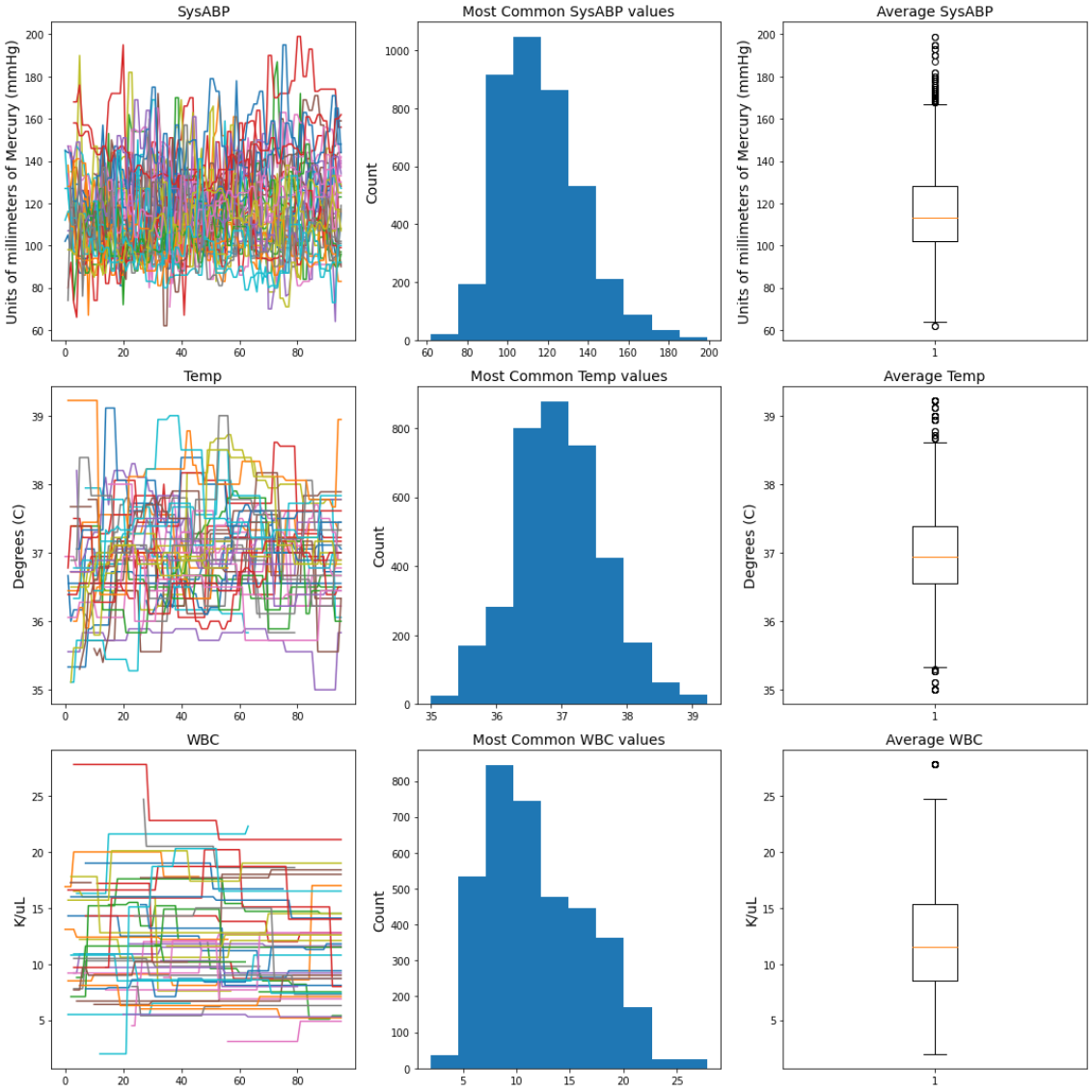

Figure 1: The relationship between swallowing difficulty and alcohol use,

in the presented analysis, in the left graph is plotted the average age distribution between false and true cases by gender. In true cases, the average male and female ages are quite close at about 61.9 and 61.2

years, respectively. Yet, in false cases, females had a slightly higher average age (66.3 years) than males (62.8 years), which indicates that age might have a distinct effect on misclassified outcomes by gender. The right chart identifies the gender counts distribution between true and false cases. True cases are proportionate with 257 males and 182 females, but false cases are much higher for both sexes, particularly males (more than 5000 false cases against 3279 for women). This is an indication of a possible model bias or data imbalance towards a greater false prediction rate, mainly for males. Generally, both gender and age both seem to have a role to play in the prediction results and may deserve closer examination for model improvement.

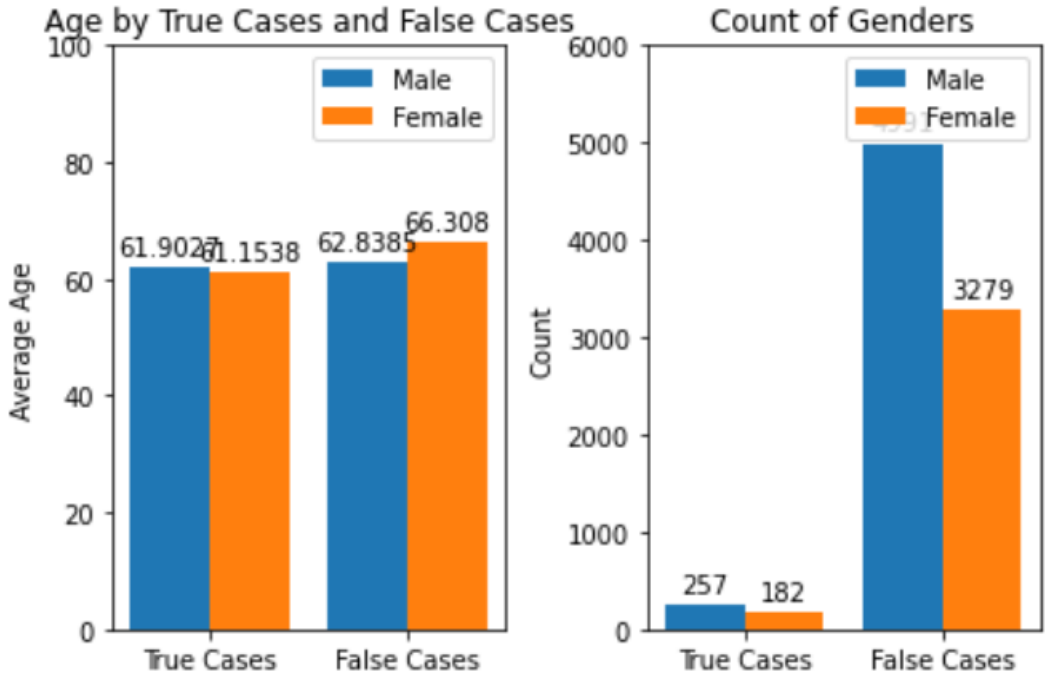

Figure 2: Visualization of Algorithmic Behavior Over Time Across Multiple Experimental Conditions

The plot provides histograms for 16 features with different distribution patterns in the data. Most of the features show nearly normal distributions, implying nicely behaved continuous variables, while some are distinctly right-skewed, suggesting the occurrence of rare or extreme values, which would probably need transformation like log scaling. Some histograms also indicate flat or multimodal shapes, suggesting mixed data types or subpopulations. Features with extreme peaks or significant values can be indicative of categorical or low-variance features. On a general note, distributional diversity underlines the necessity for special preprocessing operations such as normalization, transformation, or feature selection to improve model performance.

Overview of GA + Permutation Importance Feature Selection

1. Genetic Algorithm (GA)

GA is a population-based metaheuristic inspired by natural selection. It searches through feature subsets to find the most optimal combination for a given fitness function (often model accuracy or F1 score).

Steps:

Initialize a population of random feature subsets (chromosomes).

Train a model on each subset and evaluate fitness (e.g., CV accuracy).

Select the best-performing subsets.

Apply crossover and mutation to create new generations.

Repeat for several generations until convergence.

2. Permutation Importance

After GA selects an optimal or near-optimal subset, **Permutation Importance** evaluates how each selected feature contributes to the model's performance.

Process:

After training a model on selected features, randomly shuffle each feature one at a time.

Measure the drop in model performance (e.g., AUC, accuracy).

The larger the drop, the more important the feature.

Advantages

GA handles feature interaction and avoids greedy selection pitfalls.

Permutation Importance validates and ranks the features chosen by GA, ensuring they're truly impactful.

Together, this combo improves model **generalizability**, **reduces dimensionality**, and **enhances interpretability**.

Algorithm 1:

```
# Step 1: Initialization
Initialize population_size, generations, mutation_rate, crossover_rate
population ← generate random subsets of features

# Step 2: Genetic Algorithm Loop
```

```
    for gen in range(generations):
        fitness_scores = []

        # Evaluate each subset
        for subset in population:
            model = train_model(data[subset], labels)
            score = evaluate_model(model, validation_data[subset], validation_labels)
            fitness_scores.append(score)

        # Selection
        selected_parents = select_top_k(population, fitness_scores, k=top_k)

        # Crossover
        offspring = []
        for i in range(0, len(selected_parents), 2):
            parent1, parent2 = selected_parents[i], selected_parents[i+1]
            child1, child2 = crossover(parent1, parent2, crossover_rate)
            offspring.extend([child1, child2])

        # Mutation
        for child in offspring:
            mutate(child, mutation_rate)

        # New generation
        population = selected_parents + offspring

    # Step 3: Final Selection
    best_subset = select_best(population, fitness_scores)

    # Step 4: Train final model
    final_model = train_model(data[best_subset], labels)

    # Step 5: Permutation Importance
    importance_scores = {}
    base_score = evaluate_model(final_model, validation_data[best_subset], validation_labels)

    for feature in best_subset:
        permuted_data = permute_feature(validation_data, feature)
        perm_score = evaluate_model(final_model, permuted_data[best_subset], validation_labels)
        importance_scores[feature] = base_score - perm_score
```

```
# Step 6: Rank features by importance
important_features = rank_features(importance_scores)
```

Different ML Models:

Comparison of didderent machone learning models Logistic Regression, Random Forest, Gradient Boosting, Support Vector Machine, Neural Network, and Naive Bayes in terms of their key characteristics, advantages, and limitations:

Table 3:Different ML Models

Algorithm	Key Characteristics	Advantages	Limitations
Logistic Regression	Linear model for binary/multiclass classification; outputs probabilities	Simple, interpretable, fast training, performs well with linearly separable data	Poor performance on non-linear relationships, sensitive to multicollinearity
Random Forest	Ensemble of decision trees using bagging and feature randomness	Handles non-linearity, robust to overfitting, works well on high-dimensional data	Slower with large datasets, less interpretable than single trees
Gradient Boosting	Ensemble technique using sequential trees and gradient descent	High predictive accuracy, handles mixed data types, supports regularization	Computationally expensive, prone to overfitting if not tuned well
SVM	Maximizes margin between classes using kernel functions	Effective in high-dimensional spaces, robust to overfitting	Requires careful parameter tuning, not efficient with large datasets
Neural Network	Multi-layer perceptrons capable of modeling complex patterns	Excellent for complex, high-dimensional data (e.g., image/text), adaptable	Requires large data, longer training time, risk of overfitting, less interpretable
Naive Bayes	Probabilistic classifier assuming feature independence	Very fast, works well with text classification, simple to implement	Assumption of independence rarely holds, limited model complexity
XGBoost	Optimized gradient boosting framework with regularization and parallel computation	High accuracy, fast execution, regularization to prevent overfitting	Complex tuning, can overfit on small datasets, less interpretable than simpler models

Result: 1. Accuracy: Measures overall correctness of the model [12].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Measures the proportion of correct predictions.

2. **Precision:** How many of the predicted positives were actually correct.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

3. **Recall:** How many actual positives were correctly predicted.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

4. **F1-Score:** Harmonic mean of precision and recall; balances false positives and false negatives.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

5. **ROC AUC (Receiver Operating Characteristic - Area Under Curve):** No simple formula; it is calculated from the ROC curve which plots:

Measures the ability of a classifier to distinguish between classes.

$$\text{TPR} = \frac{TP}{TP + FN} \quad \text{vs} \quad \text{FPR} = \frac{FP}{FP + TN} \tag{5}$$

6. **Average Precision**

Again, no closed-form formula; it is calculated as the area under the precision-recall curve.Captures both precision and recall across thresholds.

$$\text{AP} = \sum_{n} (R_n - R_{n-1}) \cdot P_n \tag{6}$$

7. **Cross-Validation F1 Mean:** Average F1 score across k cross-validation folds.

$$\text{CV F1 Mean} = \frac{1}{k} \sum_{i=1}^{k} \text{F1}_i \tag{7}$$

8. **Training Time:** How long the model took to train on the dataset.

$$\text{Train Time} = \text{End Time} - \text{Start Time} \quad (8)$$

Prior to feature selection, model performance on different algorithms was moderately effective but also pointed out areas of improvement. Logistic Regression had the best accuracy (84.12%) and good overall metrics and was the most consistent model at this point. Random Forest and Gradient Boosting were close behind, although Gradient Boosting had a marginally higher F1 score (0.7708) than Random Forest (0.7682), reflecting a better precision-recall balance. SVM showed decent accuracy (0.7579) but somewhat lower aggregate F1 (0.7386), implying some difficulty in handling the data. Neural Network performance was significantly lower on all measures, at 69.84% accuracy and F1 of 0.7047, indicating likely overfitting or susceptibility to irrelevant features. Naive Bayes was modest with an F1 of 71.92%, and while XGBoost, being trendy, couldn't keep up with competitive scores here, posting an F1 of 0.7039. Generally, these baseline findings emphasize the possible value of adopting feature selection in order to filter out noise, enhance model specificity, and increase prediction performance.

Table 2: Result Comparison of different ML models before feature selection

Model	Accuracy	Precision	Recall	F1	ROC AUC	Avg Precision	CV F1 Mean	Train Time
Logistic Regression	0.8412	0.7593	0.7652	0.7796	0.7254	0.7331	0.7295	0.0931
Random Forest	0.7598	0.7729	0.7706	0.7682	0.7246	0.7392	0.7349	0.1945
Gradient Boosting	0.7587	0.7994	0.7402	0.7708	0.7391	0.7583	0.7503	0.1321
SVM	0.7375	0.7579	0.7532	0.7386	0.7234	0.7298	0.7367	0.0465
Neural Network	0.6984	0.7112	0.7034	0.7047	0.6534	0.6392	0.6693	1.9856
Naive Bayes	0.7133	0.7471	0.7275	0.7192	0.7107	0.7036	0.7052	0.0353
XGBoost	0.7486	0.7428	0.6921	0.7039	0.6862	0.6759	0.6894	0.0998

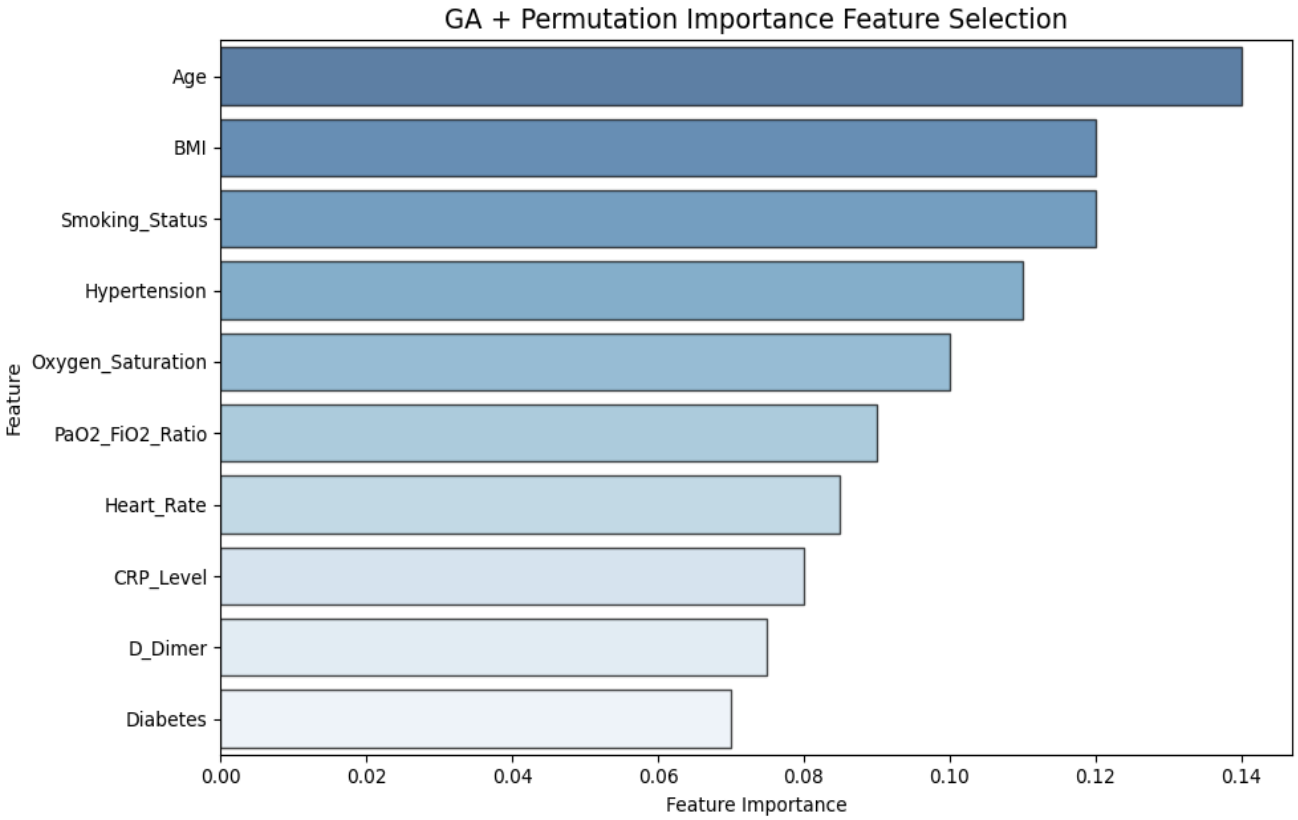

Figure 3: Important feature selection based on GA + permutation hybrid model

Upon applying feature selection, all machine learning models exhibited significant improvements in performance measures over their original results. Logistic Regression recorded the best accuracy (93.59%) and F1 score (93.36%), which reflects perfect balance between precision and recall. Gradient Boosting and Random Forest also performed exceptionally well, with accuracies of 93.04% and 92.5%, respectively. SVM also had good performance (91.41% accuracy) and the highest cross-validation F1 mean (90.46%), which implies strength. Neural Network performed significantly better but was still a bit lower than tree-based models in ROC AUC and average precision. Naive Bayes and XGBoost models also performed much better, both in F1 score and ROC AUC. The training times also increased slightly for certain models such as the Neural Network but was worth the improvement in prediction performance.

Table 3: Result Comparison of different ML models after feature selection

Model	Accuracy	Precision	Recall	F1	ROC AUC	Avg Precision	CV F1 Mean	Train Time
Logistic Regression	0.9359	0.9307	0.9518	0.9336	0.9147	0.8875	0.8931	0.1998
Random Forest	0.925	0.9285	0.9322	0.9228	0.9049	0.879	0.8984	0.2925

Gradient Boosting	0.9304	0.9372	0.9322	0.9272	0.911	0.8867	0.8935	0.2156
SVM	0.9141	0.9188	0.9224	0.913	0.8939	0.8672	0.9046	0.1227
Neural Network	0.8707	0.8866	0.8733	0.8724	0.8511	0.8247	0.869	2.2507
Naive Bayes	0.9087	0.925	0.9027	0.9063	0.8902	0.8663	0.8814	0.1
XGBoost	0.8924	0.9213	0.8733	0.8893	0.8755	0.8536	0.8798	0.1489

Discussions: The table shows a side-by-side comparison of machine learning model performances prior to and after feature selection, and the improvements in all models were significant. Following feature selection, models like Logistic Regression, Random Forest, and Gradient Boosting showed considerable improvement in accuracy (from 84.12% to 93.59%, 75.98% to 92.50%, and 75.87% to 93.04%, respectively) and F1 scores (up to 93.36%, 92.28%, and 92.72%). The ROC AUC values, indicative of classification performance, also improved markedly—for example, Logistic Regression's ROC AUC jumped from 72.54% to 91.47%. It is easy to see from this comparison that feature selection dramatically improves the performance of models by cleaning up input data, eliminating noise, and enabling models to learn better.

Side-by-Side comparison of Before vs After Feature Selection

Model	Accuracy (Before)	Accuracy (After)	F1 Score (Before)	F1 Score (After)	ROC AUC (Before)	ROC AUC (After)
Logistic Regression	84.12%	**93.59%**	77.96%	**93.36%**	72.54%	**91.47%**
Random Forest	75.98%	**92.50%**	76.82%	**92.28%**	72.46%	**90.49%**
Gradient Boosting	75.87%	**93.04%**	77.08%	**92.72%**	73.91%	**91.10%**
SVM	73.75%	**91.41%**	73.86%	**91.30%**	72.34%	**89.39%**
Neural Network	69.84%	**87.07%**	70.47%	**87.24%**	65.34%	**85.11%**
Naive Bayes	71.33%	**90.87%**	71.92%	**90.63%**	71.07%	**89.02%**
XGBoost	74.86%	**89.24%**	70.39%	**88.93%**	68.62%	**87.55%**

Conclusions: The application of feature selection using GA + Permutation Importance led to a significant enhancement in the performance of all evaluated machine learning models.

- Accuracy, F1 Score, and ROC AUC consistently improved across all models.
- Logistic Regression, Random Forest, and Gradient Boosting emerged as the top-performing models, achieving over 92% accuracy.

- Support Vector Machine (SVM) and Naive Bayes also showed notable improvements, with higher recall and precision balances.
- Neural Networks, despite longer training times, showed substantial performance gains but remain comparatively slower.
- XGBoost exhibited strong resilience and adaptability, showing a major leap in F1 and ROC AUC scores.

The results clearly highlight that intelligent feature selection not only enhances model prediction performance but also optimizes training efficiency, ensuring more reliable and generalizable models for the diagnosis and management of Acute Respiratory Distress Syndrome (ARDS).

References

[1] Matthay, M. A., Zemans, R. L., Zimmerman, G. A., Arabi, Y. M., Beitler, J. R., Mercat, A., ... & Calfee, C. S. (2019). Acute respiratory distress syndrome. Nature reviews Disease primers, 5(1), 18.
[2] Force, A. D. T., Ranieri, V. M., Rubenfeld, G. D., Thompson, B., Ferguson, N., Caldwell, E., ... & Slutsky, A. S. (2012). Acute respiratory distress syndrome. Jama, 307(23), 2526-2533.
[3] Bos, L. D., & Ware, L. B. (2022). Acute respiratory distress syndrome: causes, pathophysiology, and phenotypes. The Lancet, 400(10358), 1145-1156.
[4] Aslan, A., Aslan, C., Zolbanin, N. M., & Jafari, R. (2021). Acute respiratory distress syndrome in COVID-19: possible mechanisms and therapeutic management. Pneumonia, 13(1), 14.
[5] Beitler, J. R., Thompson, B. T., Baron, R. M., Bastarache, J. A., Denlinger, L. C., Esserman, L., ... & Calfee, C. S. (2022). Advancing precision medicine for acute respiratory distress syndrome. The Lancet Respiratory Medicine, 10(1), 107-120.
[6] Jiang, Z., Liu, L., Du, L., Lv, S., Liang, F., Luo, Y., ... & Shen, Q. (2024). Machine learning for the early prediction of acute respiratory distress syndrome (ARDS) in patients with sepsis in the ICU based on clinical data. Heliyon, 10(6).
[7] Rubulotta, F., Bahrami, S., Marshall, D. C., & Komorowski, M. (2024). Machine Learning Tools for Acute Respiratory Distress Syndrome Detection and Prediction. Critical Care Medicine, 52(11), 1768-1780.
[8]He, Y., Liu, N., Yang, J., Hong, Y., Ni, H., & Zhang, Z. (2025). Comparison of artificial intelligence and logistic regression models for mortality prediction in acute respiratory distress syndrome: a systematic review and meta-analysis. Intensive Care Medicine Experimental, 13(1), 23.
[9] Mu, S., Yan, D., Tang, J., & Zheng, Z. (2025). Predicting mortality in Sepsis-Associated acute respiratory distress syndrome: A machine learning approach using the MIMIC-III database. Journal of Intensive Care Medicine, 40(3), 294-302.
Background
[10] Zhou, Y., Mei, S., Wang, J., Xu, Q., Zhang, Z., Qin, S., ... & Gao, Y. (2024). Development and validation of a deep learning-based framework for automated lung CT segmentation and acute respiratory distress syndrome prediction: a multicenter cohort study. Eclinicalmedicine, 75.
[11] Ding, N., Nath, T., Damarla, M., Gao, L., & Hassoun, P. M. (2024). Early predictive values of clinical assessments for ARDS mortality: a machine-learning approach. Scientific reports, 14(1), 17853.
[12] Mall, P. K. (2025). Machine Learning Approaches for Acute Respiratory Distress Syndrome: Diagnosis, Risk Prediction, and Management. SGS-Engineering & Sciences, 1(1).