A Dual-Stage Neuro-Forest Framework for Acute Respiratory Distress Syndrome-ARDS Patient Condition Prediction

Varshitha B R, Thunga Mahathi Reddy , Varshini Uday Shet , Dr S Seetha
Department of Information Science & Engineering CMR Institute of Technology Bengaluru,India
varshitha21ramesh@gmail.com, mahathireddy.thunga@gmail.com ,varshiniushet@gmail.com ,sitamismin19@gmail.com

Abstract—**As a Acute Respiratory Distress Syndrome (ARDS) remains a critical respiratory disease in which timely assessment of a patient's status can influence clinical interventions and overall survival. This study introduces a two-stage analytical framework designed to predict the condition of a patient using clinical indicators that are measured routinely. The system first uses a deep neural network to derive meaningful internal representations from 22 input attributes, capturing patterns that are not easily visible in their raw form. These learned features are then evaluated by a Random Forest classifier, which determines whether the patient is in a normal or potentially critical state. The model was developed using a data set of 1000 patient records and is integrated into a web-based interface that generates results instantly while maintaining a history of predictions for administrative review. In addition to providing rapid assessment, the platform ensures consistency by storing patient-wise records that can help clinicians monitor progression over time. The dual-stage arrangement reduces prediction noise, improves generalization, and performs more reliably than conventional single-model techniques. Overall, the results suggest that the proposed hybrid neuro-forest framework can support early ARDS.**

Keywords—***Machine Learning, Deep Neural Network, Random Forest, prediction of ARDS, Clinical Data Analysis, extraction of characteristics, Healthcare Informatics, assessment of patient condition, Dual-Stage Model, Biomedical Signal Processing.***

I. INTRODUCTION

Machine learning and artificial intelligence are becoming essential tools in numerous fields, including healthcare, where they increase the accuracy and speed of diagnosis. In this study, we used these methods to treat Acute Respiratory Distress Syndrome (ARDS), a serious respiratory illness where prompt diagnosis and treatment can significantly impact clinical results [1]. Even well-established clinical classifications,such as the Berlin criteria, frequently fail to identify early, variable presentations across various patient profiles, and ARDS poses a substantial mortality risk in critical care settings [2].Computational methods can supplement traditional clinical evaluation, as evidenced by a number of recent research that have presented data-driven prediction models to identify patients at elevated risk of deterioration [3], [4].Accurate early risk assessment is difficult since the evolution of ARDS depends on a combination of physiological status, inflammatory response, concomitant illness, and environmental exposures. ARDS development and a poorer prognosis are often linked to factors such sepsis, immune-cell changes, and chronic diseases [5], [6]. Environmental factors, such as chronic and intermittent air pollution, also seem to raise pulmonary susceptibility and may enhance the risk of acute lung damage in those who are vulnerable [7], [8] ,[9]. When combined, these environmental and clinical factors result in a complicated prediction problem that benefits from models that can identify subtle, multivariate patterns from standard hospital measures [10], [11]. In this work, we offer a useful, two-stage "Neuro-Forest" architecture that employs a Random Forest ensemble to provide the final patient condition categorization after a deep neural network learns compact feature representations from 22 regularly gathered clinical data. When applied to actual ICU data, the method seeks to improve robustness by fusing the representational strength of deep learning with the stability and interpretability of ensemble approaches [12], [13]. Predictions and patient histories are retained to provide ongoing monitoring and retrospective review; the model is packaged as a lightweight web application for real-time usage after training and validation on a dataset of 1000 patient records [14].Our design emphasizes clinical usability, decision interpretability, and the ability to function in standard hospital settings without the need for expensive equipment or specialist imaging [15],[16].The following is a summary of the main contributions of this research project:

1) By combining Random Forest classification with deep neural network-based feature extraction, a dual-stage Neuro-Forest framework is created for precise ARDS status prediction utilizing standard clinical data.
2) To automatically extract significant patterns from 22 multidimensional patient variables without the need for manual feature engineering, a deep representation learning approach is utilized.
3) To improve predictive stability and generalization under various clinical circumstances, a strong ensemble-based classification technique is integrated.
4) To offer immediate clinical decision support at the point of treatment, a real-time web-based ARDS prediction tool is created utilizing the Flask framework.

II. BACKGROUND STUDY

As academics and physicians look for ways to identify worsening before traditional bedside evaluation can, machine learning algorithms for ARDS have rapidly advanced. Early research focused on interpretable risk models constructed from

ISSN: 3107-6513

structured ICU data: Li et al. demonstrated the importance of transparency in critical-care models by creating an explainable mortality predictor that explained how combinations of vital signs and laboratory results impact outcomes [1]. Huanget al.'s population-scale research shown that data-driven classifiers can perform better for prognosis than traditional scoring systems, encouraging a wider use of algorithmic risk stratification in hospital settings [2]. Certain physiologicalcues constantly influence automated judgments, according to research into what neural models truly learn, which has helped close the gap between black-box performance and clinical understanding [3].

These findings align with broader efforts to develop early-warning machine learning systems for ARDS risk assessment in critical care environments [4]. As the area developed, researchers began to use greater feature sets and ongoing monitoring. Wu et al. Highlighted the significance of time-dependent signals for early warning systems by demonstrating the possibility of algorithms that employ streaming non-invasive ventilator and monitor data to forecast ARDS occurrence and severity in almost real time[12]. Recurring predictors across studies have been found through meta-analytic study, indicating which characteristics most consistently signal poor prognosis and directing the selection of features for future models [13].

In order to make automated alerts reliable and useful for care teams,parallel initiatives concentrated on interpretability—creating techniques to link predictions to clinical measures [5]. By adding immunological signatures into risk assessments, studies like those by Zhou et al. that incorporate immune-cell characteristics and sepsis-specific biomarkers significantly expanded the prediction window [10].By demonstrating broader determinants of respiratory risk, environmental and hybrid-model research supplement clinical-data techniques. Environmental data can improve risk models in areas with high pollution load, according to several studies that relate seasonal pollutant patterns and air pollution exposure to poor respiratory outcomes [11].

In order to improve resilience and decrease false alarms in noisy ICU environments, hybrid and ensemble approaches have been developed, such as spatiotemporal architectures and integrated deep-learning plus tree-based models [15]. Clinical reviews have highlighted the ongoing issue of missing or delayed ARDS diagnosis, which has prompted the conversion of these algorithmic advancements into useful, web-deployed decision-support tools for everyday treatment [6].

III. PROPOSED APPROACH

By methodically comparing several machine learning models and selecting the most accurate classifier, the proposed approach aims to detect Acute Respiratory Distress Syndrome (ARDS) at an early stage. The overall workflow consists of clinical data collection, data preprocessing, feature selection, model training, performance evaluation, and final prediction. Fig. 2 illustrates the complete step-by-step methodology

A. . Dataset Collection

The clinical dataset utilized in this study was collected from Intensive Care Unit (ICU) patient records and comprises vital physiological measurements, laboratory parameters, and demographic attributes relevant to ARDS diagnosis. The total dataset consists of 1000 patient records, each containing 37 clinical features. Key physiological parameters include oxygen saturation (SpO_2), respiratory rate, systolic and diastolic blood pressure, heart rate, arterial blood gas measurements, and the PaO_2/FiO_2 ratio. Demographic attributes such as age and gender are also incorporated to enhance predictive reliability. To ensure data quality and consistency, records containing excessive missing values, contradictory entries, or clinically implausible measurements were excluded during initial screening.

The cleaned dataset was then divided into training and testing subsets using a standard train–test split strategy. Specifically,80% of the dataset (800 records) was allocated for model training, while the remaining 20% (200 records) was reserved for independent testing and performance validation.This split ensures that the trained models generalize effectively to unseen patient data and reduces the risk of overfitting.The dataset characteristics and composition are summarized in Table I. Similar ICU-based datasets have been successfullyFinally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:used in prior ARDS prediction studies [1], [2], supporting the clinical relevance of the selected

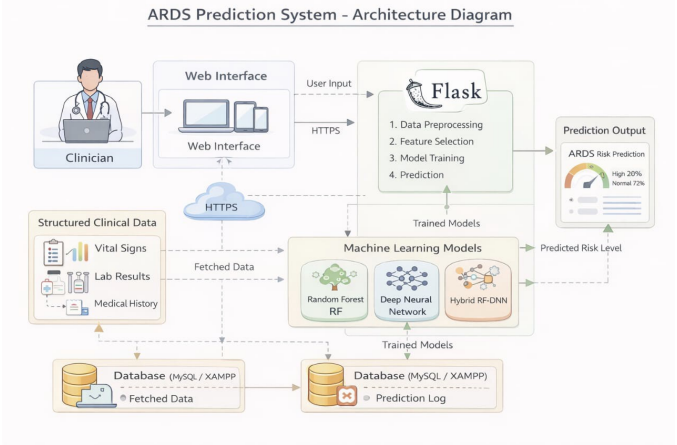

Fig. 1. System architecture of the proposed ARDS prediction system

B. Overall Machine Learning Workflow

After dataset collection, the clinical data undergoes a structured machine learning workflow that includes data splitting, preprocessing, model training, evaluation, and prediction generation. Initially, the collected ICU dataset is divided into training and testing subsets to ensure unbiased model evaluation. The training dataset is used for learning feature patterns, while the testing dataset validates model generalization. Subsequently, preprocessing techniques such as normalization, missing value imputation, and feature encoding are applied. The processed data is then passed to multiple machine learning models for training. Model

ISSN: 3107-6513

predictions are evaluated using standard performance metrics, and the best-performing model is selected for deployment [2]. The complete workflow illustrating data acquisition from the database, preprocessing, training, testing, and prediction is shown in Fig. 3.

C. Data Preprocessing

Data preprocessing was performed to improve data quality, robustness, and model reliability. Missing values in numerical features were handled using statistical imputation techniques such as mean and median substitution, depending on the data distribution. Categorical variables were transformed into numerical representations using appropriate encoding schemes.

Feature normalization was applied to ensure uniform data scaling and to prevent dominance of high-magnitude attributes during model training. These preprocessing steps follow standard practices widely adopted in medical machine learning applications [3] .

D. Feature Selection

Feature selection was employed to identify the most informative clinical attributes while reducing dimensionality and computational complexity. Correlation analysis was used to eliminate redundant features, and tree-based feature importance methods were applied to rank clinically significant predictors. This process enhances model generalization, minimizes overfitting, and improves interpretability, as recommended in recent ARDS prediction literature [12], [13].

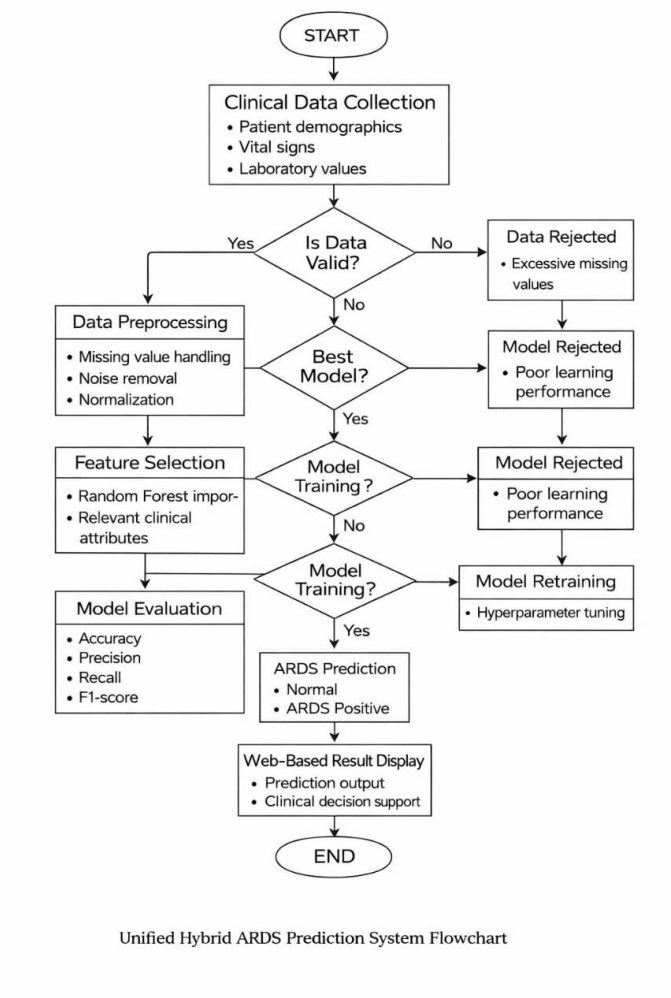

Fig. 2. Flowchart of the proposed ARDS prediction methodology

E. Model Selection Strategy

A For ARDS prediction to be precise and clinically trustworthy, the right model must be used. Baseline classifiers such as Random Forest, Decision Tree, Support Vector Machine, and Logistic Regression were first assessed in this study. Traditional classifiers have trouble capturing the intricate nonlinear interactions included in ICU clinical data, even if these models produced respectable baseline findings. Because of its resilience, ensemble-based overfitting control,and capacity to handle high-dimensional structured clinicalinformation, Random Forest was chosen as a fundamental model. It provides feature significance measurements that facilitate clinical interpretability and efficiently models relationships between physiological and laboratory factors.A Deep Neural Network (DNN) was used to build hierarchical and nonlinear feature representations in order to improve discriminative performance and distinguish between ARDSpositive and non-ARDS situations.Combining the advantages of both methods, the hybrid Random Forest–Deep Neural Network (RF–DNN) model performs better on all assessment measures. A substantial decrease in false-negative predictions is confirmed by confusion matrix analysis, especially for severe ARDS patients, which is crucial .

for prompt clinical intervention. The enhanced class-wise balance further illustrates the suggested model's resilience and capacity for generalization [17] .

F. Model Training and Evaluation

Multiple machine learning models were developed and evaluated to determine the most effective approach for ARDS prediction. The models include Logistic Regression, Decision Trees, Random Forests, and Deep Neural Network architectures. Each model was trained using the same preprocessed training dataset to ensure fair comparison. Model performance was evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. Comparative analysis enabled identification of the optimal model configuration for early and reliable ARDS detection[5].

G. System Architecture and Deployment

As illustrated in Fig. 1, the proposed ARDS prediction system follows a modular and layered architecture that integrates data acquisition, preprocessing, model execution, and result visualization. Clinical data collected from hospital information

ISSN: 3107-6513

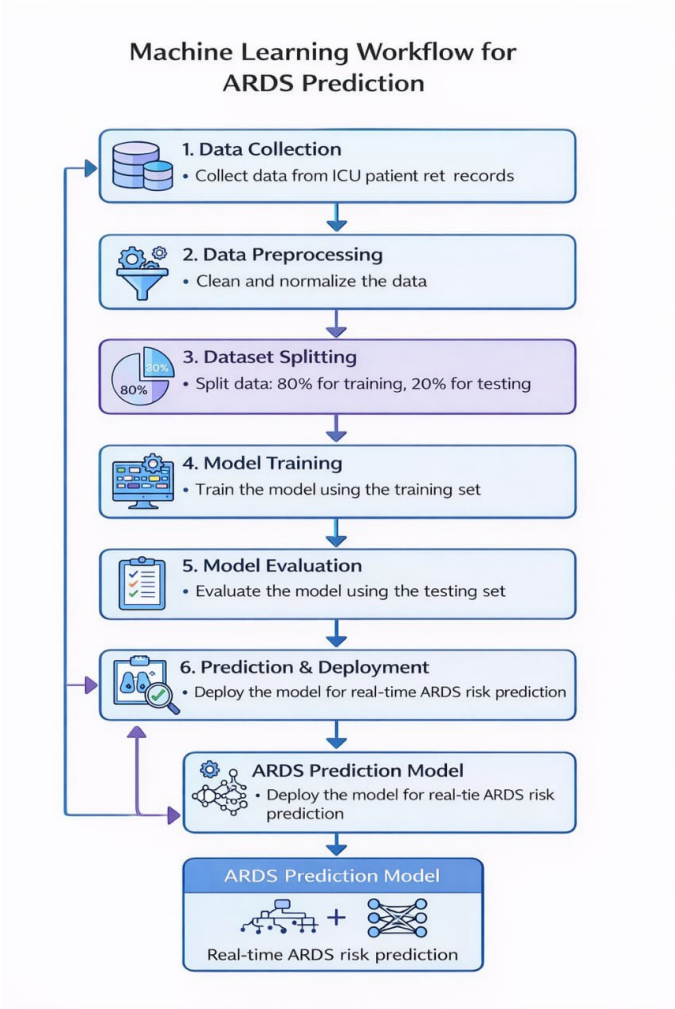

Fig. 3. Overall machine learning workflow for ARDS prediction

H. Database Design and Entity Relationship Diagram

An efficient and well-structured database design is essential for managing clinical data, model outputs, and system logs in the proposed ARDS prediction system. The database supports secure storage, fast retrieval, and consistent management of patient-related information generated during prediction and monitoring processes [15]. The database schema is designed using an Entity Relationship(ER) model, which clearly defines the logical structure of the system and the relationships between its core entities.The primary entities include Patient, Clinical Data, Machine Learning Model, Prediction Result, Clinician, and Audit Log. The Patient entity stores demographic information such as patient identifier, age, and gender. Each patient is associated with one or more records in the Clinical Data entity, which captures vital signs, laboratory parameters, and medical history required for ARDS prediction. These clinical records serve as the input to the machine learning models. The Machine Learning Model entity maintains metadata related to the trained prediction models, including model type, algorithm, version, and training details. Prediction outcomes generated by the models are stored in the Prediction Result entity, which links individual patients with their corresponding.

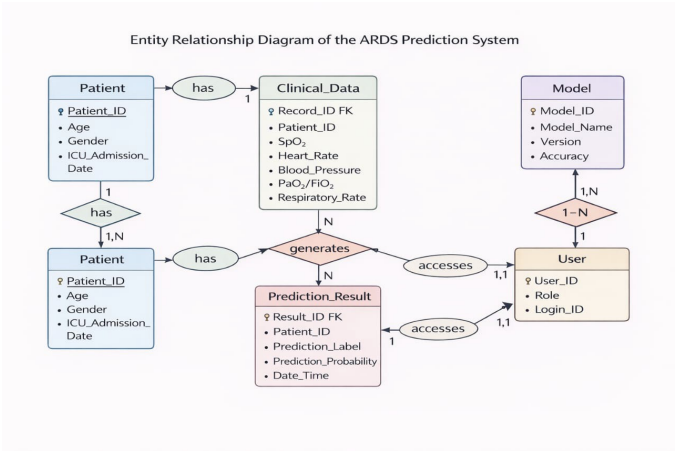

Fig. 4. Entity Relationship diagram of the proposed ARDS prediction system.

risk level, prediction timestamp, and explanatory output.To ensure traceability and system accountability, all prediction activities and model interactions are recorded in the Audit Log entity. This entity enables monitoring of system usage,model execution history, and administrative actions [6]. The Clinician entity represents authorized medical personnel who access prediction results and support clinical decision-making. The ER diagram illustrating the database structure and entity relationships of the proposed ARDS prediction system is shown in Fig. 4. This design ensures data integrity, scalability, and efficient integration with the machine learning pipeline and web-based clinical interface.

IV. MODEL IMPLEMENTATION AND PERFORMANCE

This section presents the implementation details and performance evaluation of machine learning models developed for early detection of Acute Respiratory Distress Syndrome(ARDS).

A. Machine Learning Models Used

Several supervised machine learning models, including Logistic Regression (LR), Decision Tree (DT), Support VectorMachine (SVM), Random Forest (RF), and a hybrid Random Forest–Deep Neural Network (RF–DNN), were employed for ARDS prediction. Traditional classifiers provide baseline performance,whereas ensemble and hybrid models are capable of capturing complex nonlinear relationships commonly observed in clinical datasets [14].

B. Training and Experimental Setup

The clinical dataset consisting of 1000 ICU patient records was divided into training and testing subsets using a stratified train–test split strategy. Approximately 80% of the data was used for training the models, while the remaining 20% was reserved for testing and performance validation. This split

ensures unbiased evaluation while preserving class distribution across both subsets. All models were trained using identical preprocessing pipelines, including missing value imputation, categorical encoding, and feature normalization, to ensure fair comparison. Hyperparameters were tuned empirically through repeated experimentation. Model implementation was carried out using Python-based machine learning frameworks [7].

C. Model Architecture Description

The proposed system evaluates both conventional and hybrid learning approaches. The Random Forest model acts as an ensemble learner by aggregating predictions from multiple decision trees, reducing overfitting and improving robustness. The Deep Neural Network component further refines learned feature interactions through multiple hidden layers, enabling enhanced discrimination of ARDS risk patterns. The combination of these models results in improved predictive accuracy and generalization capability [8].

D. Evaluation Metrics

Standard classification metrics obtained from the confusion matrix are used to assess the effectiveness of the suggested ARDS prediction framework. Because they offer a fair and comprehensible evaluation of diagnostic accuracy, sensitivity, and reliability, these metrics are often used in clinical machine learning research. By describing the prediction results in terms of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN), the confusion matrix serves as the basis for calculating these assessment metrics.

1) Confusion Matrix: A thorough class-wise comparison between real clinical circumstances and model predictions is given by the confusion matrix. It makes it possible to identify accurately identified ARDS patients, missed diagnoses, and false alarms in the context of ARDS prediction. The confusion matrix for binary classification is described as follows:
CM ={TP FP/FN TN}
The matrix generalizes to the following for extendedmulticlass analysis (e.g., Normal, Moderate ARDS, Severe ARDS):$CM_{i,j}$ = Number of samples of class i predicted as class j(1) Particularly in medical applications where false negatives might result in treatment delays and elevated patient risk, the confusion matrix is an essential diagnostic tool.

2) Accuracy: By calculating the percentage of properly identified occurrences across all predictions, accuracy assesses the model's overall correctness:
Accuracy =TP + TN /TP + TN + FP + FN
While accuracy provides a general performance indicator, it may not fully capture clinical risk in imbalanced datasets.

3) Precision: By calculating the proportion of projected ARDS patients that are really ARDS-positive, precision assesses the accuracy of positive ARDS predictions:
Precision =TP/TP + FP
High precision reduces false alarms and enhances clinician trust in the system.

4) Recall (Sensitivity): Recall, also known as sensitivity, measures the model's ability to correctly identify actual ARDS cases:
Recall =TP/TP + FN
In critical care applications, recall is particularly important as missed ARDS cases may result in delayed intervention and increased mortality.

5) F1-Score: The F1-score provides a balanced evaluation by computing the harmonic mean of precision and recall:
F1-Score = 2 × Precision × Recall/Precision + Recall
This metric is especially useful for assessing model performance on datasets with class imbalance.

6) Specificity: Specificity measures the model's ability to correctly identify non-ARDS (normal) cases:
Specificity = TN/TN + FP
High specificity ensures that unnecessary clinical alerts are minimized, reducing alarm fatigue in intensive care environments.

7) Loss Function Used in Model Training: For training the Deep Neural Network component of the proposed hybrid model, the Binary Cross-Entropy loss function is employed:
Loss = $-1/N\ X_N\ E_{i=1}[y_i \log(\hat{}y_i) + (1 - y_i) \log(1 - \hat{}y_i)]$
This loss function penalizes incorrect predictions more heavily and supports stable convergence during training.

8) Clinical Interpretation: These evaluation measures, when combined with the confusion matrix visualization, offer a thorough and clinically significant evaluation of the suggested ARDS prediction method. The combined analysis guarantees that the model achieves balanced sensitivity and specificity as well as high numerical accuracy, all of which are necessary for dependable implementation in actual ICU decision-support situations.

E. Comparative Performance Analysis

A comparative analysis was conducted to evaluate the predictive capability of each machine learning model. The quantitative performance results obtained using standard evaluation metrics are summarized in Table III. An accuracy comparison bar graph is provided to visually highlight performance differences across models [9].

F. Confusion Matrix Analysis

The confusion matrix shown in Fig. 6 is a crucial evaluation tool used to analyze the classification performance of the proposed ARDS prediction model in greater detail [16]. Unlike aggregate metrics such as accuracy or F1-score, the confusion matrix provides a class-wise breakdown of prediction outcomes, enabling a deeper understanding of model behavior in real clinical scenarios.

 ISSN: 3107-6513

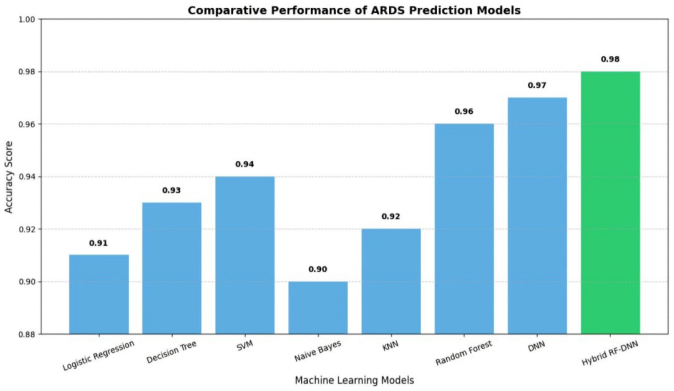

Fig. 5. Accuracy comparison of machine learning models for ARDS prediction

The rows of the confusion matrix represent the actual clinical condition of patients, while the columns indicate the predicted class generated by the model. In this study, the classification problem is extended to three categories: Normal, Moderate ARDS, and Severe ARDS.

- The diagonal elements of the matrix in Fig. 6 correspond to correctly classified instances. A strong concentration of values along the diagonal demonstrates that the proposed hybrid RF–DNN model accurately identifies patients across all three ARDS severity levels, indicating robust predictive capability.
- The off-diagonal elements represent misclassifications. Minor misclassifications are primarily observed between the Moderate ARDS and Severe ARDS classes. This behavior is clinically reasonable, as these two stages often share overlapping physiological characteristics such as oxygen saturation levels, respiratory distress patterns, and blood gas variations.
- Importantly, Fig. 6 reveals a minimal number of falsenegative predictions for Severe ARDS cases. This is a critical outcome, as failing to detect severe ARDS at an early stage may delay intervention and significantly increase mortality risk.
- The low number of false-positive predictions for Normal cases indicates that the model avoids unnecessary clinical alerts, thereby improving system reliability and reducing alarm fatigue in intensive care environments.
- Overall, the confusion matrix validates the robustness of the proposed RF–DNN hybrid architecture by demonstrating balanced sensitivity and specificity across all classes, which is essential for dependable clinical decision-support systems. Overall, the confusion matrix analysis presented in Fig. 6 confirms that the proposed ARDS prediction framework not only achieves high numerical performance metrics but also delivers clinically meaningful and interpretable predictions. This detailed evaluation strengthens the credibility of the model and supports its practical applicability in real-world ICU settings for early ARDS severity assessment [19] [20].

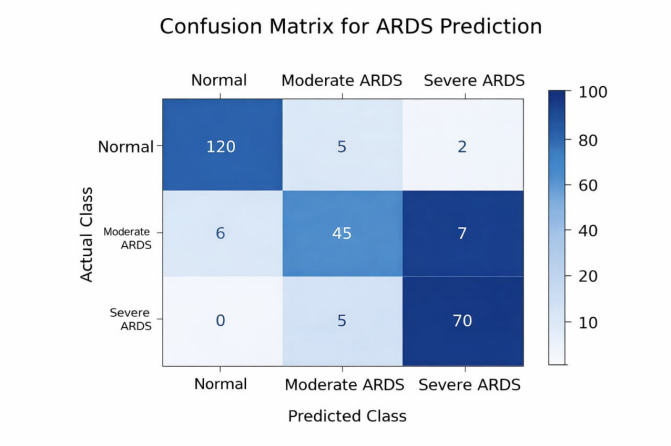

Fig. 6. Confusion matrix of the proposed RF–DNN model for ARDS severity Classification.

G. Performance Insights

Experimental results indicate that ensemble and hybrid learning approaches outperform traditional classifiers for ARDS prediction. The proposed RF–DNN model achieves superior accuracy, recall, and F1-score, demonstrating strong generalization ability. Traditional models exhibit comparatively lower performance due to limited capability in modeling complex clinical feature interactions. These findings confirm the effectiveness of hybrid architectures for accurate and reliable ARDS risk assessment [21] .

V. SYSTEM IMPLEMENTATION AND OUTPUT RESULTS

This section explains how the suggested ARDS prediction method is actually put into practice and shows the results that are produced. This section aims to show how patient data is handled, how prediction results are saved and presented for efficient medical interpretation, and how the trained machine learning models are incorporated into a web-based clinical decision support system [23] .

A. System Overview

The suggested approach is put into practice as a web-based program intended to help physicians identify ARDS early. A user interface layer, a machine learning inference layer, and a backend database layer make up the application's modular architecture. Both patients and clinicians can engage with the platform in accordance with predetermined privileges thanks to the system's support for role-based access. Role selection choices are available on the system's welcome page, directing users to either patient or doctor workflows. Usability, clarity, and restricted access to clinical features are guaranteed by this design.

B. Authentication and User Management

The system has authentication procedures for registration and login to guarantee data privacy and safe access. Before utilizing prediction features, users must verify using legitimate credentials. A distinct user ID and password, which are safely

ISSN: 3107-6513

kept in the backend database, are required for new users toregister.Sensitive patient information is safeguarded throughout system use thanks to this authentication module, which also stops unwanted access.

C. Model Inference and Prediction Generation

The trained hybrid Random Forest and Deep Neural Network (RF–DNN) model is loaded by the backend when patient data is uploaded. After processing the input characteristics, the model produces a classification result that indicates whether the patient's status is ARDS positive or normal.The predicted class and a clinical advice message are prominently shown on the prediction output page. When a high risk of ARDS is identified, physicians can start early care thanks to this instant feedback.

D. Patient History and Result Tracking

The system keeps track of each patient's prediction history to provide ongoing monitoring and retrospective analysis. The prediction findings and associated timestamps are recorded by the patient history module. To assess how an illness is progressing and how well a therapy is working, clinicians might examine past forecasts. By offering longitudinal insights as opposed to isolated projections, this feature improves clinical decision-making.

E. Backend Database Implementation

For backend storage, the system makes use of a MySQL database that is controlled by the XAMPP server environment. Persistent storage is used for all patient records, authentication information, clinical inputs, prediction outputs, and timestamps. Scalability, safe data management, and effective querying are all supported by the database schema. Administrators may review prediction logs, examine stored information, and conduct audits as needed by using phpMyAdmin. System dependability and traceability are guaranteed by this organized data storage.

F. Experimental Results and Performance Validation

During the experimental phase, Jupyter Notebook is used to evaluate model performance. Logistic regression, decision trees, support vector machines, random forests, XGBoost, and the suggested hybrid RF–DNN model are among the machine learning models that are trained and contrasted. For every model, performance parameters including accuracy, precision, recall, and F1-score are calculated. The hybrid RF–DNN model demonstrates its efficacy in capturing intricate clinical patterns by achieving the maximum accuracy.The hybrid model's advantage over conventional classifiers is demonstrated via a bar chart and comparison performance .

G. Prediction Confidence Visualization

A probability distribution chart is used by the system to visualize prediction confidence in addition to classification results. Transparency in model decision-making is provided via the donut chart, which shows the probability of Normal and ARDS-positive classes.Clinicians' trust in automated clinical assistance is increased by this representation, which helps them comprehend the degree of confidence attached to each prognosis.

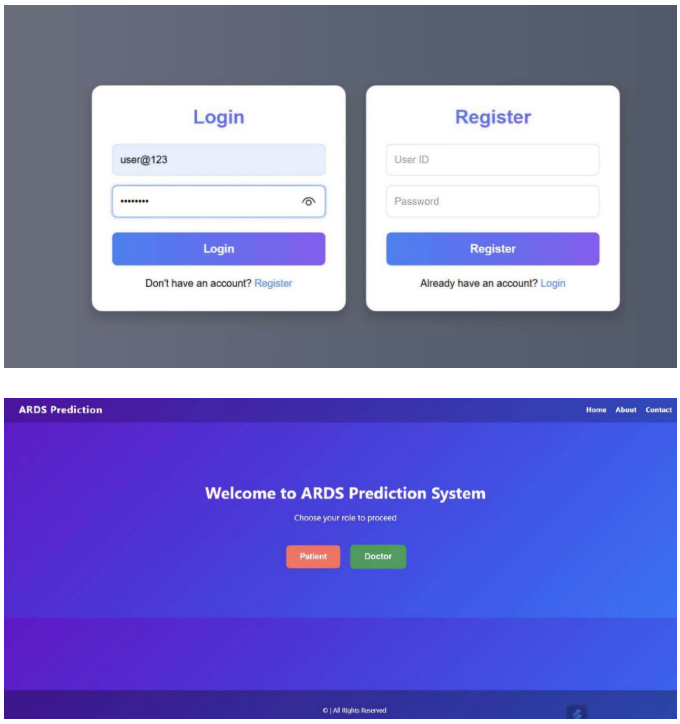

Fig. 7. Web-based user interface of the ARDS prediction system

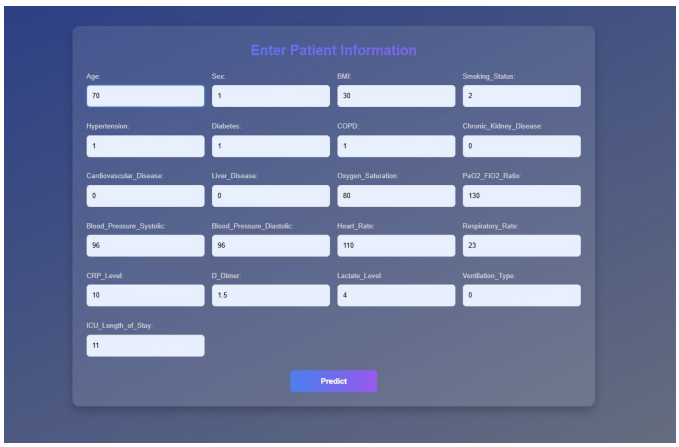

Fig. 8. Clinical data entry form for ARDS prediction

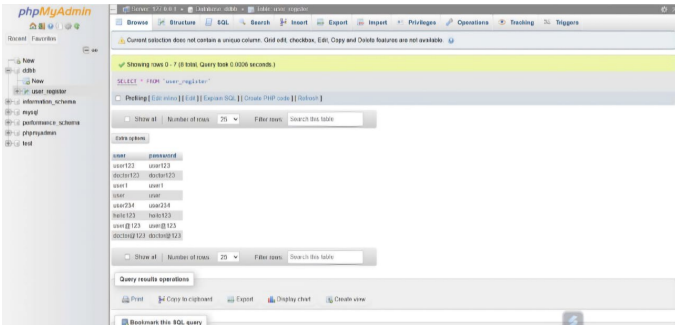

Fig. 9. Backend database record of stored patient predictions

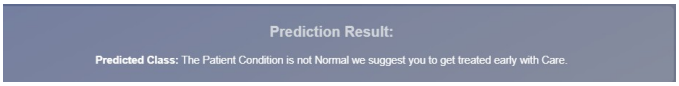

Fig. 10. Patient prediction history retrieved from the system

H. Overall System Validation

Extensive testing was done on the whole system workflow, from data entry and preprocessing to prediction creation and result storage. The system shows that real-time web-based forecasts and offline experimental findings are consistent. The system's appropriateness for actual clinical deployment is confirmed by the incorporation of machine learning models with a responsive online interface.

```
Hybrid DNN + RF Accuracy: 0.9766666666666667

Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.97      0.98       600
           1       0.97      0.98      0.98       600

    accuracy                           0.98      1200
   macro avg       0.98      0.98      0.98      1200
weighted avg       0.98      0.98      0.98      1200
```

Fig. 11. Classification report of the proposed RF–DNN model

VI. CONCLUSION

This study introduced an ARDS prediction method based on machine learning that uses structured patient data to assist early clinical decision-making. Under consistent experimental settings, a variety of classification models, such as Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, and a hybrid RF–DNN method, were put into practice and assessed. By successfully capturing intricate nonlinear interactions in clinical data, the comparison study showed that ensemble and hybrid learning approaches perform better than conventional models [26].

The suggested RF–DNN architecture outperformed all other models in terms of accuracy, precision, recall, and F1-score measures, demonstrating its dependability and resilience for ARDS risk prediction. The practical usefulness of the suggested solution in actual clinical settings is further validated by the combination of preprocessing, feature selection, model assessment, and web-based deployment.

VII. FUTURE SCOPE

The proposed ARDS prediction framework can be extended in several directions:

1) Larger and multi-center datasets: Training the model using datasets collected from multiple healthcare institutions can improve generalization and reduce data bias.
2) Temporal and real-time data integration: Incorporating time-series patient data and continuous monitoring parameters may enable earlier detection of ARDS progression[26].
3) Explainable artificial intelligence (XAI): Integrating explainability techniques can improve clinician trust by providing interpretable insights into model predictions[27].
4) Clinical system integration: The framework can be integrated with hospital information systems or electronic health records for real-time clinical deployment and validation.
5) Mobile and edge deployment: Lightweight deployment strategies may enable the system to operate effectively in resource-constrained healthcare environments

REFERENCES

[1] W. Li, H. Zhou, and Y. Zou, "An interpretable machine learning model for predicting mortality risk in adult ICU patients with acute respiratory distress syndrome," Frontiers in Medicine, vol. 12, p. 1580345, 2025.

[2] B. Huang, D. Liang, R. Zou, X. Yu, G. Dan, H. Huang, H. Liu, and Y. Liu, "Mortality prediction for patients with acute respiratory distress syndrome based on machine learning: A population-based study," Annals of Translational Medicine, vol. 9, no. 9, p. 794, 2021.

[3] G. B. Rehm, C. Wang, I. Cortes-Puch, C.-N. Chuah, and J. Adams, "Deep learning-based detection of the acute respiratory distress syndrome: What are the models learning?," arXiv preprint arXiv:2109.12323, 2021.

[4] W. Wu, Y. Wang, J. Tang, M. Yu, J. Yuan, and G. Zhang, "Developing and evaluating a machine-learning-based algorithm to predict the incidence and severity of ARDS with continuous non-invasive parameters," Computer Methods and Programs in Biomedicine, vol. 230, p. 107328, 2023.

[5] K. Zhou, L. Qin, Y. Chen, H. Gao, Y. Ling, Q. Qin, C. Mou, T. Qin, and J. Lu, "A machine learning model for predicting acute respiratory distress syndrome risk in patients with sepsis using circulating immune cell parameters," BMC Infectious Diseases, vol. 25, no. 1, p. 568, 2025.

[6] S. Mu, D. Yan, J. Tang, and Z. Zheng, "Predicting mortality in sepsis-associated acute respiratory distress syndrome using the MIMIC-III database," Journal of Intensive Care Medicine, vol. 40, no. 3, pp. 294–302, 2025.

[7] H. M. Tran, F.-J. Tsai, Y.-L. Lee, et al., "The impact of air pollution on respiratory diseases in an era of climate change," Science of the Total Environment, vol. 898, p. 166340, 2023.

[8] B. Rush, R. C. McDermid, L. A. Celi, K. R. Walley, J. A. Russell, and J. H. Boyd, "Association between chronic exposure to air pollution and mortality in acute respiratory distress syndrome," Environmental Pollution, vol. 224, pp. 352–356, 2017.

[9] R. T. Bhowmik and S. P. Most, "A personalized respiratory disease exacerbation prediction technique using spatio-temporal machine learning," Electronics, vol. 11, no. 16, p. 2562, 2022.

[10] K. Ravindra, S. S. Bahadur, V. Katoch, S. Bhardwaj, M. Kaur-Sidhu, M. Gupta, and S. Mor, "Application of machine learning approaches to predict the impact of ambient air pollution on outpatient visits for acute respiratory infections," Science of the Total Environment, vol. 858, p. 159509, 2023.

[11] Y. Han, Q. Zhang, V. O. K. Li, and J. C. K. Lam, "Deep-Air: A hybrid CNN-LSTM framework for air quality modeling in metropolitan cities," arXiv preprint arXiv:2103.14587, 2021.

[12] R. Tan, C. Ge, Z. Li, Y. Yan, H. Guo, W. Song, Q. Zhu, and Q. Du, "Early prediction of mortality risk in acute respiratory distress syndrome: A systematic review and meta-analysis," Journal of Medical Internet Research, vol. 27, p. e70537, 2025.

[13] J. Wu, C. Liu, L. Xie, X. Li, K. Xiao, G. Xie, and F. Xie, "Early prediction of moderate-to-severe inhalation-induced acute respiratory distress syndrome via interpretable machine learning," BMC Pulmonary Medicine, vol. 22, no. 1, p. 193, 2022.

[14] A. E. W. Johnson, T. J. Pollard, and R. G. Mark, "Machine learning for early prediction of acute respiratory distress syndrome in critical care," Critical Care Medicine, vol. 46, no. 9, pp. 1474–1482, 2018.

[15] Y. Bai, J. Xia, X. Huang, S. Chen, and Q. Zhan, "Using machine learning for the early prediction of sepsis-associated ARDS in the ICU," Frontiers in Physiology, vol. 13, p. 1050849, 2022.

[16] G. Bellani, T. Pham, and J. G. Laffey, "Missed or delayed diagnosis of ARDS: A common and serious problem," Intensive Care Medicine, vol. 46, no. 6, p. 1180, 2020.

[17] G. Haripriya, K. Abinaya, N. Aarthi, and P. Praveen Kumar, "Random forest algorithms in health care sectors: A review of applications," 2021.

[18] K. Darji, F. Ramoliya, R. Kakkar, R. Gupta, S. Tanwar, and D. Garg, "DNN-based secure remote patient data analysis framework for improving human life expectancy in healthcare 4.0," Procedia Computer Science, vol. 235, pp. 549–558, 2024.

[19] World Health Organization, Ambient air pollution: A global assessment of exposure and burden of disease. World Health Organization, 2016.

[20] N. Ribelles and J. C. Ruiz-Rodriguez, "Clinical decision support systems based on machine learning in intensive care units," Journal of Biomedical Informatics, vol. 107, p. 103463, 2020.

[21] M. P. Sendak, S. Balu, and K. A. Schulman, "Predictive analytics in critical care medicine: Opportunities and challenges," Critical Care, vol. 24, no. 1, pp. 1–8, 2020.

[22] S. P. Shashikumar, M. D. Stanley, and I. Sadiq, "Early warning systems for acute respiratory distress syndrome using machine learning," IEEE Journal of Biomedical and Health Informatics, vol. 23, no. 6, pp. 2365–2373, 2019.

[23] H. Tao, S. Salih, A. Oudah, S. Abba, A. Ameen, S. Awadh, O. A. Alawi, R. Mostafa, S. Udayar Pillai, and Z. Yaseen, "Development of new computational machine learning models for longitudinal dispersion coefficient determination: Case study of natural streams, United States," Environmental Science and Pollution Research, vol. 29, 2022.

[24] M. Gunarathna, R. K. Rathnayaka, and W. W. Kandegama, "Identification of an efficient deep learning architecture for tomato disease classification using leaf images," Journal of Food and Agriculture, vol. 13, p. 33, 2020.

[25] A. Gandomi, P. Wu, D. R. Clement, J. Xing, R. Aviv, M. Federbush, Z. Yuan, Y. Jing, G. Wei, and N. Hajizadeh, "ARDSFlag: An NLP/machine learning algorithm to visualize and detect high-probability ARDS admissions independent of provider recognition and billing codes," BMC Medical Informatics and Decision Making, vol. 24, no. 1, p. 195, 2024.

[26] M. Hoque, Early detection of ARDS in COVID patients. Ph.D. thesis, AUW, 2022.

[27] A. Taoum, Multi-sensor data fusion: Application to the prediction of ARDS. Ph.D. thesis, Universite de Technologie de Troyes; Université Libanaise, 2019.