Society of Medical Education & Research

2026, Volume 6, Issue 1, Page No: 93-106
ISSN: 3108-4826

Journal of Medical Sciences and Interdisciplinary Research

An Ensemble Machine Learning Framework for Mortality Risk Prediction in Idiopathic Pulmonary Fibrosis

Sara Ben Youssef[1]*, Amal Trabelsi[1], Karim Boudiaf[2], Nabil Jebali[3]

[1] Department of Interstitial Lung Diseases and ML Prediction, Faculty of Medicine, University of Tunis, Tunis, Tunisia.
[2] Department of Idiopathic Pulmonary Fibrosis and Risk Stratification, Faculty of Medicine, University of Sousse, Sousse, Tunisia.
[3] Department of Ensemble Learning and Clinical Outcomes, Faculty of Medicine, University of Sfax, Sfax, Tunisia.

***E-mail** ✉ sara.benyoussef@fmed.rnu.tn

Abstract

Idiopathic pulmonary fibrosis (IPF) ranks among the most frequent forms of interstitial lung disease and is characterized by progressive scarring (fibrosis) of lung tissue. Affected individuals are generally advised to pursue lung transplantation; failure to do so often leads to ongoing, irreversible pulmonary deterioration and eventual mortality. With advanced, irreversible damage, reliable forecasting of patient survival becomes critically important. Conventional clinical approaches typically lack robust predictive instruments for this purpose. Nevertheless, the proven capacity of artificial intelligence to address complex medical scenarios has opened the door to mortality prediction through machine learning methodologies. The present study introduced a soft voting ensemble classifier based on the 30 most informative clinical features to estimate mortality risk in patients diagnosed with idiopathic pulmonary fibrosis. Five distinct machine learning models were integrated for this task: random forest (RF), support vector machine (SVM), gradient boosting machine (GBM), XGBoost (XGB), and multi-layer perceptron (MLP). Integration of the classifiers via the soft voting ensemble yielded an accuracy of 79.58%, sensitivity of 86%, F1-score of 84%, prediction error of 0.19, and responsiveness of 0.47. Conclusions: The developed model can assist physicians in clinical decision-making, longitudinal disease monitoring, and risk mitigation. Ultimately, it supports efforts to lower mortality rates, optimize patient health status, and improve risk stratification.

Keywords: Artificial intelligence, Idiopathic pulmonary fibrosis disease, Machine learning, Mortality risk, Prediction

Introduction

Idiopathic pulmonary fibrosis (IPF) falls within the broader category of interstitial lung diseases (ILDs), a diverse group of pulmonary parenchymal disorders marked by varying degrees of fibrosis and inflammation [1]. Among these conditions, IPF is the most frequent and clinically aggressive subtype. It is characterized by relentless progression toward respiratory failure, continuous worsening of pulmonary function, and substantial mortality rates [2]. While some patients exhibit identifiable triggers, a considerable proportion of IPF cases arise without any clear underlying cause [3, 4]. Although environmental exposures (such as air pollution), certain pharmacological agents, and infectious processes can increase the likelihood of developing IPF, the fundamental mechanisms responsible for the disease remain poorly understood [5, 6]. Established risk factors include advancing age, sex, hereditary elements, tobacco use, and gastro-oesophageal reflux disease [5, 7]. Typical presenting features involve a persistent dry cough and progressive shortness of breath on exertion. Affected individuals often report pronounced fatigue, unintended weight loss, and joint and muscle discomfort. After diagnosis, patients require

Access this article online https://smerpub.com/

Received: 19 November 2025; **Accepted:** 27 February 2026

How to cite this article: Ben Youssef S, Trabelsi A, Boudiaf K, Jebali N. An Ensemble Machine Learning Framework for Mortality Risk Prediction in Idiopathic Pulmonary Fibrosis. J Med Sci Interdiscip Res. 2026;6(1):93-106. https://doi.org/10.51847/AYtcETOX5X

regular clinical oversight. No definitive cure is currently available; therefore, management centers on vigilant monitoring and supportive care [8]. Existing interventions primarily relieve breathing difficulties or slow functional decline. Consequently, patients must adapt to living with this chronic and debilitating illness.

In cases of advanced lung destruction, lung transplantation offers the potential to extend life expectancy [9]. However, the high cost of this procedure leaves many patients without viable options, exposing them to fatal outcomes. Mortality associated with IPF remains elevated. Historical analyses conducted by Ref. [10] indicate a median post-diagnosis survival period of roughly 2–3 years. Supporting these findings, Ref. [11] documented mortality rates of approximately 64.3 deaths per million among men and 58.4 deaths per million among women. Investigations by Ref. [12] further emphasize the heightened vulnerability of adults aged sixty years and above. Despite considerable research efforts, prognostication in IPF continues to pose difficulties for clinicians, since baseline pulmonary function parameters alone prove inadequate for reliably forecasting mortality [13].

The expanding role of digital health technologies and artificial intelligence (AI) applications holds considerable promise for improving the longitudinal care of individuals diagnosed with IPF. For example, Bhatt *et al.* [14] introduced an AI-enabled federated learning system designed for stroke risk prediction in a Healthcare 4.0 environment underpinned by 5G connectivity, with potential integration into wearable platforms for continuous cardiac surveillance. In a related development, author [15] created Eboost, a new AI-based tool for forecasting COVID-19 outcomes. Likewise, Ref. [16] proposed an integrated framework merging AI with computer vision techniques to enable real-time surveillance, adherence monitoring, and mitigation of infectious disease transmission. The medical literature contains extensive work on predictive modeling for various pulmonary and respiratory disorders, including IPF.

In earlier periods, IPF diagnosis relied predominantly on conventional techniques that were often laborious, only modestly effective, and limited in diagnostic precision. A representative non-data-driven strategy, grounded in statistical techniques such as regression or Cox proportional hazards modeling, was advanced by Refs. [13] and [17]. Numerous additional initiatives have been undertaken, yet many have fallen short in delivering robust accuracy or practical utility. AI methodologies now offer new opportunities to assess disease severity and predict survival probabilities in IPF. Sophisticated AI detection algorithms allow healthcare professionals to identify IPF cases with greater precision and speed [18]. Beyond early identification, AI-supported diagnostic tools equip patients with better strategies to control their condition and sustain an acceptable quality of life [18]. One notable example is the Fibresolve platform, which has exhibited strong performance as a complementary diagnostic aid, attaining 87% specificity and outperforming standard approaches in sensitivity [19]. Similarly, the ZCoR-IPF instrument uses comorbidity profiles extracted from clinical records to improve diagnostic accuracy, supporting earlier recognition of IPF and initiation of appropriate therapies [20]. By identifying intricate patterns in clinical datasets and evaluating longitudinal medical histories, AI systems offer a compelling pathway toward more reliable and efficient IPF diagnosis. This advancement facilitates earlier clinical intervention and improved long-term patient prognosis [21].

The current investigation applies artificial intelligence techniques, notably machine learning (ML) algorithms, to forecast mortality risk in IPF patients. The resulting model is designed to function as a practical decision-support instrument for clinicians, enabling more informed management, effective risk stratification, and ultimately reduced mortality. Unlike prior statistically oriented models, the proposed framework is fully data-driven and ML-based, promising superior reliability, effectiveness, and predictive performance.

The structure of this paper is as follows: Section 2 surveys relevant prior studies. Section 3 details the data acquisition process and the methodological approach adopted. Section 4 reports the findings generated by the proposed model. Finally, Sections 5 and 6 discuss the results and overall conclusions, respectively.

Related work

In recent years, investigators have examined the clinical status of patients with idiopathic pulmonary fibrosis (IPF) to identify episodes of disease exacerbation and to estimate survival probabilities. A wide range of methodologies and techniques have been applied to tackle these challenges. This section reviews selected prior studies on IPF management.

Evgeni *et al*. [22] examined the use of artificial intelligence (AI) methods, including machine learning,

within respiratory medicine. The authors concluded that although AI cannot fully substitute for physicians, such models serve as valuable adjuncts in clinical decision-making, enhancing patient care while minimizing human error and reducing treatment expenses. AI has emerged as a promising approach for resolving various healthcare challenges. This perspective was reinforced by a systematic review by S. Soffer *et al*. [23], which emphasized AI's contribution to assessing interstitial lung disease (ILD) on chest computed tomography. Their analysis demonstrated that AI has transformed medical image interpretation, thereby advancing the quality of care for patients with ILD.

More recently, Mäkelä *et al*. [24] introduced an AI model based on a convolutional neural network (CNN) that can quantify fibroblast foci and inflammatory cells. Tested on a cohort of 71 IPF patients, the model successfully identified histological features that are challenging to measure manually, providing new insights for prognostic evaluation. In a comparable effort, Handa *et al*. [25] developed an AI-driven system for analyzing chest computed tomography scans in IPF. This approach supplemented conventional gender, age, and lung physiology variables with imaging data to improve prediction of disease exacerbation. Yu *et al*. [26] constructed a machine learning model using a hybrid algorithm combining quantum particle swarm optimization and random forest (QPSO-RF) in a study involving 50 IPF patients. The model achieved an overall accuracy of 82.1% in forecasting disease progression from computed tomography data. Employing lung CT scans from IPF patients, Mandal *et al*. [27] identified the most suitable machine learning model for predicting lung function in advanced disease stages. The system evaluated multiple ML algorithms, generated confidence scores, and estimated final forced vital capacity values for individual patients. Ziyuan Wang [28] applied a deep learning (DL) strategy to automatically predict IPF. The model utilized a DenseNet neural network architecture, integrating CT scan images with clinical variables to forecast changes in lung function and to determine the presence of IPF. These investigations primarily relied on medical imaging, particularly CT scans.

Although CT imaging has been extensively used for diagnosis, some studies have explored using clinical data alone to assess IPF. For example, Shi *et al*. [26] introduced a soft-voting ensemble model to predict IPF severity early. This ensemble combined three base classifiers (GBM, RF, and XGB), attaining an accuracy of 71%, precision of 64%, recall of 71%, and F1 score of 66%. The model helps clinicians and patients evaluate disease severity in early phases, supporting more effective disease management. This study uses the same dataset but focuses on predicting patient survival. It incorporates comprehensive feature engineering and detailed model evaluation to enhance predictive performance and reliability. Additional research on mortality prediction in IPF includes Wu *et al*. [29], who developed the CTPF risk prediction model. This model was trained on CT images alongside physiological parameters such as FVC, DLco, SpO2, age, and gender using deep learning techniques. Results indicated a positive association between high-risk classification and increased mortality. Thillai *et al*. [30] also addressed mortality prediction in IPF patients through CT scan analysis. Their objective was to automate imaging biomarker extraction via deep learning-based segmentation, achieving over 97.8% success across training sets and revealing a strong correlation with forced vital capacity (FVC), a key measure of pulmonary function. The resulting tool enables rapid CT segmentation and provides accurate short- and long-term prognostic assessments.

In contrast to these earlier approaches, the current method relies solely on clinical data to predict survival in IPF patients, using the dataset's reported average survival duration as a reference. Extensive data preprocessing and feature engineering were performed to ensure data integrity and model robustness. Given the critical implications of survival predictions for patient outcomes, a rigorous performance evaluation was conducted to confirm the validity and trustworthiness of the proposed AI model's results. The dataset's characteristics further distinguish this research and the developed methodology, as elaborated in subsequent sections.

Materials and Methods

This section details the materials and methodological framework adopted for the current investigation.

Data source and implementation environment

The data for this study were drawn from the Korean Interstitial Lung Disease Cohort (KICO). The cohort includes records from 2424 patients and incorporates 502 distinct features associated with IPF. All individuals provided informed consent to participate. The dataset offers extensive patient profiles covering demographic

details, prior medical evaluations related to IPF, pharmacological histories, survival and mortality outcomes, instances of disease exacerbation, pulmonary function metrics, and various other clinical parameters.

While multiple strategies—from conventional statistical techniques to advanced AI methods—have been applied to IPF prediction, reliable analysis requires both accessible high-quality data and suitable processing. Ali *et al*. [31] previously analyzed this dataset with a different focus, assessing disease severity and assigning patients to one of four exacerbation classes (class 0 to class 3). The present work, however, pursues a distinct research direction by investigating survival and mortality in IPF patients in relation to a variety of clinical characteristics. These characteristics provide valuable indications of patient condition and, in cases of death, help clarify potential contributing factors, including acute exacerbation, pneumonia, trans-infection or sepsis, pulmonary embolism, lung cancer, heart disease, or heart failure. The dataset is labeled with three outcome categories: "Survival," "Death," and "Tracking Failed." The "Survival" category identifies patients who remained alive, the "Death" category denotes those who died due to IPF, and "Tracking Failed" refers to individuals lacking sufficient follow-up data to establish a definitive outcome. This study concentrates on predicting survival versus mortality using clinical features and omits records classified as "Tracking Failed," given their limited utility for determining patient status.

All computational experiments for model development were conducted on a workstation operating under Windows 10, featuring an Intel (R) Core (TM) i7-10700 processor and 16 GB RAM. Programming was performed in Python 3.8 using the Jupyter Notebook platform, with primary libraries consisting of TensorFlow v2.13, Keras 2.13, and Scikit-learn.

Data preprocessing

Thorough preprocessing of the dataset is critical for securing strong performance from AI models. Standard preprocessing operations include managing missing values, normalizing or scaling variables, selecting features, and carefully including or excluding cases [32, 33]. The following subsections describe the key preparatory procedures applied to the data.

Feature selection

Datasets with a large number of variables frequently include features that add little predictive value and can occasionally undermine model effectiveness. Feature selection aims to retain only the most informative attributes, thereby enhancing performance, lowering model complexity, and decreasing computational requirements [34, 35]. This study tested three prominent feature selection approaches on this dataset: feature importance ranking [36], recursive feature elimination with cross-validation [37, 38], and the Boruta automated selection method [39]. Recursive feature elimination with cross-validation emerged as the most suitable technique for the present data. This method is favored in machine learning applications because it systematically identifies relevant predictors while helping to prevent overfitting. The procedure used a gradient boosting classifier with 5-fold cross-validation, using accuracy as the optimization metric to maximize correct classifications. This technique identified an optimal subset of 30 features, as shown in **Figure 1**.

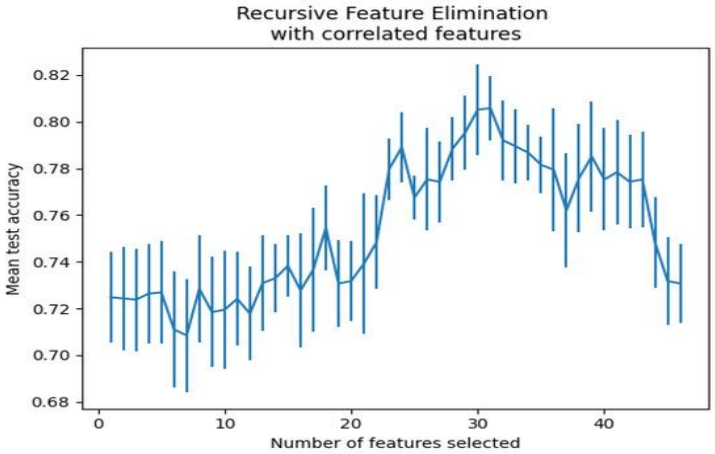

Figure 1. Number of features vs. cross-validation scores.

Complementing the algorithmic outcomes, discussions with clinical specialists reinforced the selection of these 30 most pertinent features based on their relative importance, as shown in **Figure 2**.

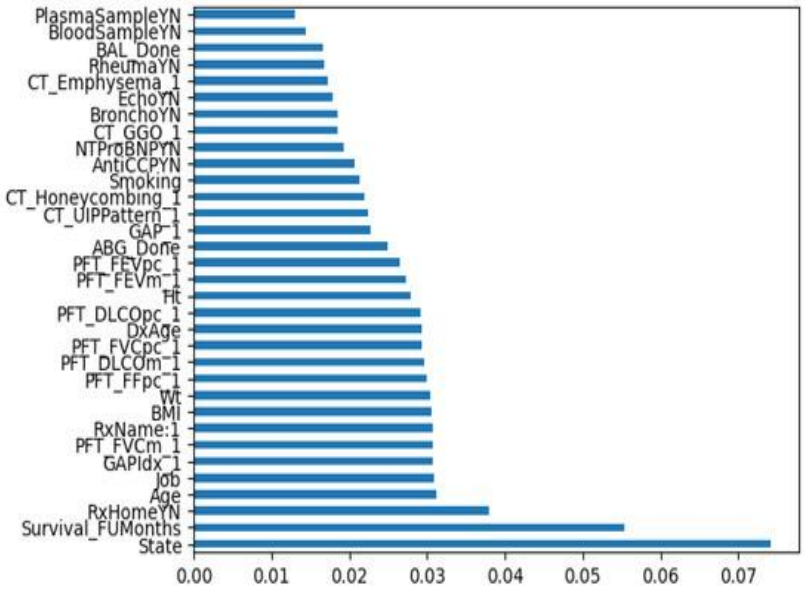

Figure 2. Feature classification according to their importance.

Table 1 provides a brief overview of the 30 selected features.

Table 1. Selected features and descriptions.

No.	Feature	Description	Values
1	Ht	Patient height	Numerical
2	wt	Patient weight	Numerical
3	BMI	Body mass index (BMI)	Numerical
4	Age	Patient age	Numerical
5	Job	Occupational category	1 = Housewife, 2 = Office worker, 3 = Commerce, 4 = Construction site
6	Smoking	Smoking status	1 = Never smoked, 2 = Current smoker, 3 = Former smoker
7	DxAge	Age at diagnosis	Numerical
8	ABG_Done	Arterial blood gas test performed	1 = Performed, 0 = Not performed
9	PFT_FEVm	Forced expiratory volume (measured) from pulmonary function test	Numerical
10	PFT_FEVpc	Predicted forced expiratory volume (%)	Numerical
11	PFT_FVCm	Measured forced vital capacity	Numerical
12	PFT_FVCpc	Predicted forced vital capacity (%)	Numerical
13	PFT_FFpc	Pulmonary function test free-flow parameter	Numerical
14	PFT_DLCOm	Measured diffusing capacity of the lungs for carbon monoxide	Numerical
15	PFT_DLCOpc	Predicted diffusing capacity of the lungs for carbon monoxide	Numerical
16	CT_UIPPattern	UIP pattern identified on CT scan	1 = Definite, 2 = Probable, 3 = Inconsistent
17	CT_Honeycombing	Presence of honeycombing on CT	Yes = 1, No = 0

18	CT_GGO_1	Presence of ground-glass opacity (GGO)	Yes = 1, No = 0
19	BloodSampleYN	Blood sample available	Yes = 1, No = 0
20	PlasmaSampleYN	Plasma specimen available	Yes = 1, No = 0
21	SerumSampleYN	Serum specimen available	Yes = 1, No = 0
22	NTProBNPYN	NT-proBNP assessment performed	1 = Performed, 0 = Not performed
23	NTProBNP	NT-proBNP measurement	Numerical
23	GAP_1	GAP stage classification	Categorical
24	EchoYN	Echocardiography performed	1 = Performed, 0 = Not performed
25	GAPIdx_1	GAP index classification	Categorical
27	RxName	Prescribed medication (active ingredient)	0 = None, 1 = Corticosteroid, 2 = N-acetylcysteine, 3 = Pirfenidone, 4 = Nintedanib
28	RxHomeYN	Home oxygen therapy	Yes = 1, No = 0
29	Survival_FUMonths	Follow-up duration (months)	Categorical
30	State	Disease exacerbation status	0 = No exacerbation, 1 = First-stage exacerbation, 2 = Second-stage exacerbation, 3 = Third-stage exacerbation

Data sampling

Data sampling constitutes a statistical procedure that involves selecting, processing, and examining a representative portion of a larger dataset to uncover underlying patterns and trends. It is also an effective strategy for managing class imbalance within datasets [40, 41]. Class imbalance occurs when the distribution of observations across target categories is uneven—in this case, the "survival" class [32]. Such imbalance poses a significant challenge because many machine learning algorithms tend to overlook the minority class, even though accurate recognition of this class is crucial for overall model performance. To mitigate the imbalance in the current dataset, we oversampled the minority class using the Synthetic Minority Oversampling Technique (SMOTE), a widely recognized and effective oversampling method [41]. SMOTE generates synthetic examples for the minority class by interpolating between a given instance and one of its k-nearest neighbors, thereby producing new data points that enhance class representation. This technique has proven reliable for creating balanced datasets [32] and helps improve model learning by reducing bias associated with unequal class distributions [41].

Proposed machine-learning model

The present study evaluated several machine learning models to predict survival outcomes (survival or death) among patients with IPF. Specifically, the following established classifiers were implemented: multi-layer perceptron (MLP), random forest (RF), support vector machine (SVM), gradient boosting machine (GBM), and XGBoost (XGB). These models were applied to determine the survival and mortality status of IPF patients with high accuracy. However, individual classifiers may not always yield optimal results due to variations in input hyperparameters. **Table 2** summarises the hyperparameters examined for each classifier along with their assigned values, selected to construct a robust and efficient predictive system.

Table 2. Specifications of the classifiers for IPF mortality risk prediction.

Classification model	Configured hyperparameters
Multi-Layer Perceptron (MLP)	hidden_layer_sizes = (500,500,500), max_iter = 1000, random_state = 42
Random Forest	n_estimators = 500, criterion = 'gini', max_depth = 20, min_samples_split = 5, min_samples_leaf = 5, random_state = 0
Support Vector Machine (SVM)	kernel = 'linear', degree = 7, gamma = 5.9, C = 30, decision_function_shape = 'ovr', probability = True, random_state = 0
Gradient Boosting Machine (GBM)	learning_rate = 0.1, n_estimators = 500, max_depth = 15, min_samples_split = 5, min_samples_leaf = 5, subsample = 1, max_features = 'sqrt', random_state = 10
Extreme Gradient Boosting (XGBoost)	objective = 'binary:logistic', learning_rate = 0.01, n_estimators = 500, max_depth = 15, gamma = 7, colsample_bytree = 0.4, subsample = 0.9, scale_pos_weight = 2, reg_alpha = 0.1, silent = False, random_state = 0

To leverage the strengths of multiple classifiers and minimize the overfitting risk inherent in any single model, we adopted a soft voting ensemble approach. This technique integrates predictions from several base classifiers to form a more effective overall model. **Figure 3** illustrates the fundamental architecture of the proposed ensemble classifier. In a soft voting ensemble, diverse machine learning classifiers or varied training subsets are combined to generate final predictions [32]. The ensemble aggregates probability outputs from each constituent classifier before making a collective decision on unseen data. The primary rationale for employing an ensemble strategy is to reduce generalization error in predictions. By combining diverse and independent base learners, the approach limits individual model weaknesses and yields more stable results. Although the ensemble incorporates multiple base classifiers, it operates as a unified predictive system. Ensemble modeling represents a common and effective practice in applied machine learning.

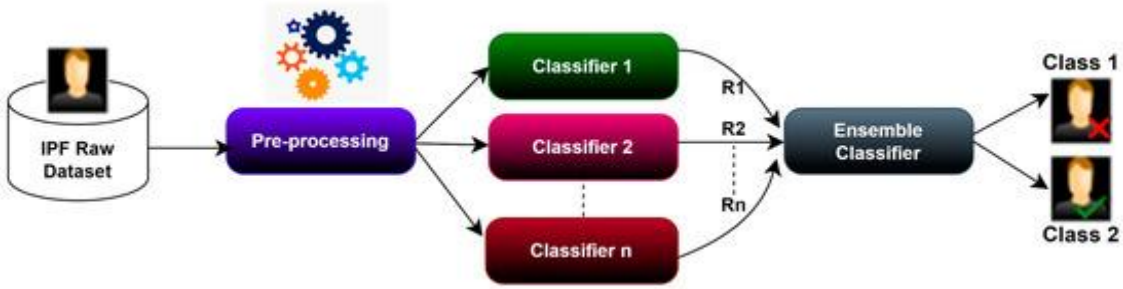

Figure 3. Basic architecture of the proposed ensemble classifier with the arrows highlighting the participation of different classifiers in determining the best predictor.

Performance evaluation metrics

The performance of the proposed ELIPF model was assessed using several key indicators, including model accuracy (A+), sensitivity or recall (R+), average precision (P+), F1-score (F1), reliability and error-related measures (MSE, MCC, K, and ROC-AUC), as well as response time or latency (Time). Brief explanations of these evaluation metrics are provided in Eq. (1)–Eq. (8). Accuracy represents the overall correctness of the ELIPF model in classifying both "Survived" and "Death" outcomes. It is computed as shown in Eq. (1) and reflects the proportion of correct predictions across all cases.

$$Accuracy\ (A^{+}) = \frac{TP\ +\ TN}{TP\ +\ TN\ +\ FP\ +\ FN} \quad (1)$$

TP: True Positive; TN: True Negative; FN: False Negative; FP: False Positive.

Recall (also known as sensitivity) quantifies the proportion of actual positive cases that the model correctly identifies. In this study, recall indicates how effectively the model recognizes samples belonging to each class. The average recall across classes was calculated according to Eq. (2).

$$Recall\ (R^{+}) = \frac{TP}{TP\ +\ FN} \quad (2)$$

Precision evaluates the proportion of positive predictions that are actually correct. In this research, precision was determined using Eq. (3) to gauge the reliability of the model's assignment of instances to particular classes.

$$Precision\ (P^{+}) = \frac{TP}{TP\ +\ FP} \quad (3)$$

F1-Score (F1) provides a balanced single measure that combines recall and precision. It is calculated via Eq. (4) and accounts for both positive and negative samples, applying a penalty for misclassifications within each class.

$$F1 - Score\ (F1) = 2 * \frac{Precision\ *\ Recall}{Precision\ +\ Recall} \quad (4)$$

To further assess the reliability and effectiveness of the proposed model, we employed additional quantitative metrics. These measures offer deeper insights into model behavior and highlight areas for potential refinement.

Mean Squared Error (MSE) quantifies the average squared deviation between predicted and actual values, as defined in Eq. (5) and Eq. (6).

$$Error_i = Predicted_i - Actual_i \quad (5)$$

$$\mathrm{MSE} = \sum_{i=1}^{n} \frac{(\mathrm{Error_i})^2}{n} \quad (6)$$

MSE is an important indicator of model reliability, reflecting the model's accuracy and capacity for generalization. Lower MSE values correspond to higher predictive reliability.

Matthews Correlation Coefficient (MCC) incorporates true positives, true negatives, false positives, and false negatives to provide a comprehensive evaluation of binary classification performance. It provides a balanced assessment even with class imbalance, making it particularly suitable for the current study. Given the dataset's imbalanced nature, MCC was used to validate classifier quality and complement insights from accuracy and F1-score. By considering all elements of the confusion matrix, MCC reveals the interrelationships among the factors influencing the ELIPF model's final predictions. The coefficient is calculated as presented in Eq. (7).

$$\mathrm{MCC} = \frac{(\mathrm{TP} \times \mathrm{TN}) - (\mathrm{FN} \times \mathrm{FP})}{\sqrt{(\mathrm{TP} + \mathrm{FP}) \times (\mathrm{FP} + \mathrm{TN}) \times (\mathrm{TP} + \mathrm{FN}) \times (\mathrm{FN} + \mathrm{TN})}} \quad (7)$$

Cohen's Kappa Coefficient (K) assesses the level of agreement between observed and expected classifications, accounting for chance agreement in classification tasks. It is derived from the values in the confusion matrix.

$$\mathrm{K} = \frac{2 \times (\mathrm{TP} \times \mathrm{TN} - \mathrm{FN} \times \mathrm{FP})}{(\mathrm{TP} + \mathrm{FP}) \times (\mathrm{FP} + \mathrm{TN}) \times (\mathrm{TP} + \mathrm{FN}) \times (\mathrm{FN} + \mathrm{TN})} \quad (8)$$

Latency or Time (s) refers to the inference time required by the ELIPF model to process input data and generate predictions, thereby indicating the model's responsiveness.

Process workflow

Figure 4 illustrates the complete workflow for constructing the predictive system to classify IPF patients into the two outcome categories—"Survived" and "Death"—. The procedure consists of six primary phases, several of which have been described in detail in preceding subsections. These phases encompass data acquisition, preprocessing, implementation of base classifiers, development of the proposed ensemble model, performance evaluation, and final prediction. Clinical specialists gathered the IPF patient data, which formed the basis of the dataset. During preprocessing, key operations such as feature selection and KNN imputation were performed. The processed data were subsequently partitioned into training and testing sets using an 80:20 ratio in accordance with the Pareto principle. Given the dataset's inherent class imbalance, which could lead to model overfitting, the Synthetic Minority Oversampling Technique (SMOTE) was applied to balance the classes. Multiple base classifiers were trained on the balanced training data, from which five models—MLP, RF, XGBoost, SVM, and Gradient Boosting—demonstrated the strongest performance with optimally tuned hyperparameters. These were integrated into the ensemble model. The efficacy of the proposed ensemble was assessed through metrics including accuracy, recall, precision, confusion matrix, ROC, and AUC to reliably distinguish between survival and mortality states in IPF patients.

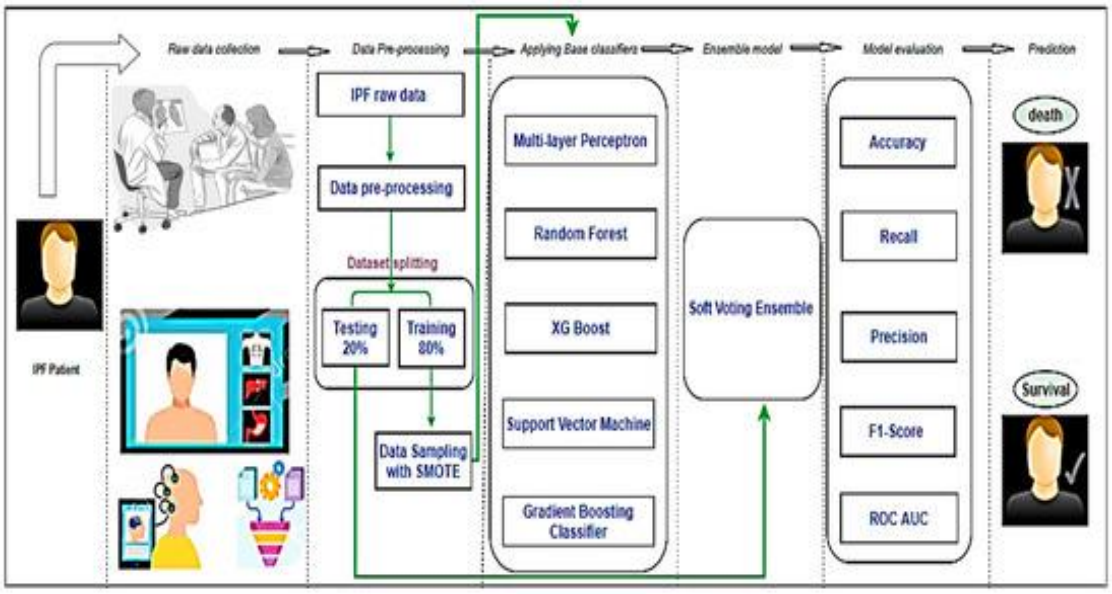

Figure 4. Complete Process flow of the experiment.

Experimental results

Five machine learning classifiers were combined to form an ensemble model capable of accurately predicting survival outcomes in patients with IPF. The classifiers were trained on 80% of the original dataset after application of the SMOTE technique to address class imbalance. Model performance was evaluated on the remaining 20% reserved for testing, with the outcomes summarised in **Table 3**.

Table 3. Comparative analysis of classifiers' performance.

Classifiers	Accuracy	Precision	Recall	F1-Score
MLP	0.7622	0.8016	0.8412	0.8250
Random Forest	0.7803	0.8105	0.8611	0.8300
XGBoost	0.7441	0.8624	0.7109	0.7864
SVM	0.7648	0.8232	0.8105	0.8103
GBM	0.7881	0.8135	0.8722	0.8401
SVE	**0.7958**	**0.8261**	**0.8637**	**0.8420**

The results indicate that the proposed soft voting ensemble (SVE) model achieved superior overall performance in distinguishing "Survival" and "Death" states among IPF patients. It recorded an accuracy of 79%, sensitivity of 86%, and F1-score of 84.2%. Although GBM attained a slightly higher recall of 87.2% than SVE's 86.3%, SVE performed better across most evaluation metrics. Similarly, while XGBoost achieved higher precision (86.2%), its accuracy (74%) and F1-score (78.6%) were lower than SVE's, confirming the ensemble as the preferred model for this prediction task.

A confusion matrix further illustrates the model's classification performance in a two-dimensional format. The x-axis denotes the actual classes (labeled 0 and 1, corresponding to survival states), while the y-axis shows the predicted labels. The diagonal elements represent correct classifications for each category and align well with expected performance levels. **Figure 5** displays the confusion matrix generated by the proposed soft voting ensemble model.

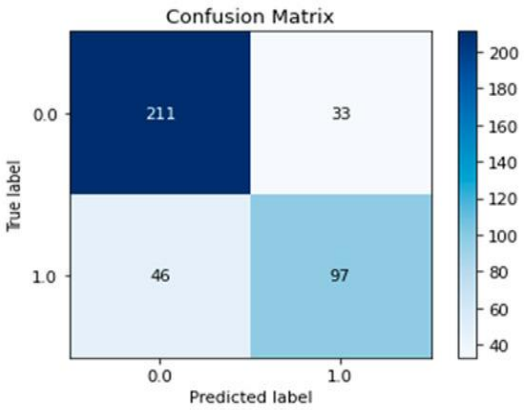

Figure 5. Confusion matrix of our proposed ensemble model.

As shown in **Figure 5**, the ELIPF model exhibited limited misclassification, correctly identifying 211 cases

as actual positives and 97 as actual negatives. These outcomes confirm the model's adequate capacity to determine the clinical status of IPF patients.

Table 4 presents additional indicators of the proposed model's predictive reliability and efficiency relative to the individual classifiers.

Table 4. Classifiers' performance evaluation.

Classifiers	Time (s)	MSE	MCC	K
MLP	0.7154	0.2532	0.0107	0.5229
Random Forest	0.2720	0.4677	0.0327	0.0428
XGBoost	0.0339	0.2222	0.0160	0.0348
SVM	2.3360	0.2144	0.0120	0.0662
GBM	0.0166	0.3927	0.0197	0.0170
SVE	0.4740	0.1925	0.5555	0.58484

In terms of computational efficiency, the GBM recorded the shortest inference time (0.0166 s), whereas the SVM required 2.3360 s. The SVE achieved a latency of 0.4740 s. Although the GBM offered faster predictions, the SVE demonstrated greater reliability in error reduction and generalization, with an MSE of 0.1925, MCC of 0.555, and K of 0.58484. Given the emphasis on precision in health monitoring applications, the SVE—with its balanced latency of 0.4740 s—is more dependable and effective than the GBM, which showed higher error (0.3927) and lower generalization values (0.0197 and 0.0170).

To further confirm the superiority of the SVE model over the constituent classifiers, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is presented in **Figure 6**.

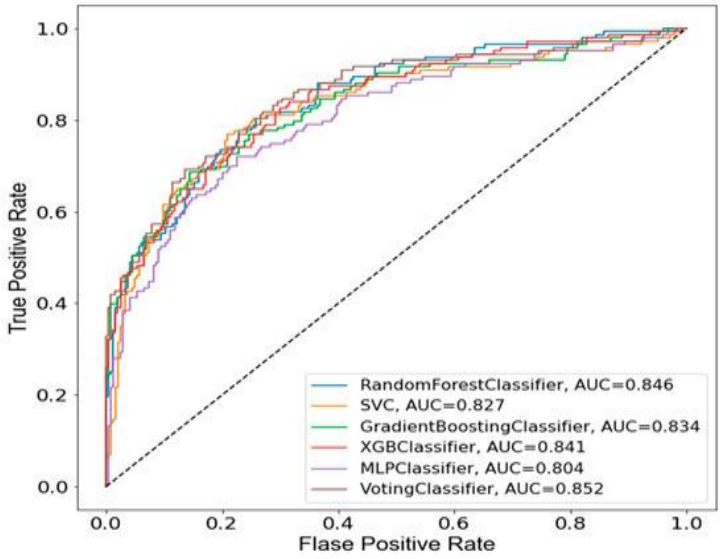

Figure 6. AUC-ROC of the proposed ensemble model.

The AUC-ROC curve graphically represents model performance across various classification thresholds.

$$\mathrm{AUC-ROC} = \int_0^1 \mathrm{TPR(x)dx} \quad (9)$$

It is computed as the integral of the true positive rate (TPR) with respect to the false positive rate (FPR) from 0 to 1. This probability curve illustrates the trade-off between sensitivity and specificity at different decision thresholds. As evident in **Figure 6**, the SVE model exhibits strong generalisability, achieving an AUC value of 0.852 across multiple thresholds.

Results and Discussion

This study introduced a predictive framework for determining "Survival" and "Death" outcomes in patients

with idiopathic pulmonary fibrosis (IPF) based exclusively on clinical data, rather than relying on CT scan images as in several previous investigations [23, 42, 43]. Using data from the Korean Interstitial Lung Disease Cohort (KICO) at Haeundae Paik Hospital, Busan, South Korea, the research applied several advanced machine learning algorithms to perform binary classification of patient survival status. Performance varied across the algorithms tested on the dataset. The five most effective models—multi-layer perceptron, random forest, XGBoost, support vector machine, and gradient boosting—were selected and integrated into an ensemble that surpassed the performance of any individual classifier, as detailed in **Table 3**. Both soft and hard voting ensemble strategies were examined. Hard voting determines the final classification by majority vote among the base models, whereas soft voting aggregates the predicted probabilities from each classifier. The soft voting approach was ultimately chosen due to its superior results.

The proposed model was rigorously assessed using the performance indicators presented in **Table 3**. The findings show that the soft voting ensemble provides the most effective solution for predicting survival outcomes in IPF patients in this dataset, offering strong predictive capability, reliability, and practical efficiency. Thus, developing a robust early-warning system for mortality risk assessment in IPF using machine learning and readily available clinical features is feasible. In comparison with related studies, the current work specifically targets binary survival versus mortality prediction, differing from approaches focused on survival analysis. Moreover, the dataset consists solely of clinical variables without accompanying imaging data, which distinguishes this research. A notable limitation, however, is that the model's overall performance could be further improved with a larger volume of data. High-accuracy machine learning models generally require extensive training samples, which were not fully available in the present dataset. Future efforts will focus on expanding the dataset, retraining the model, and minimizing missing values during collection to improve outcomes.

Conclusion

The present study developed a machine learning model designed to identify "Survival" and "Death" states in IPF patients utilizing clinical characteristics. The model performs automated binary classification between these two outcomes. Thirty highly relevant features were selected from the dataset of 2424 patients through recursive feature elimination to optimize model fit. Preprocessing steps included imputing missing values via the K-nearest neighbors method and correcting class imbalance using the SMOTE oversampling technique. Multiple base classifiers were evaluated on the prepared data, and the five strongest performers were combined into a soft-voting ensemble that delivered the highest overall results. This ensemble achieved an accuracy of 79.58%, recall of 86%, and F1-score of 84%, outperforming the individual state-of-the-art classifiers and providing reliable predictions and classifications.

The proposed model supports clinicians by facilitating informed decision-making, effective disease monitoring, reduced mortality risk, improved patient health management, and better risk stratification. Future work will involve gathering additional data and exploring more advanced artificial intelligence methods, such as deep learning, to further refine the model and increase predictive accuracy.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Kalchiem-Dekel O, Galvin JR, Burke AP, Atamas SP, Todd NW. Interstitial lung disease and pulmonary fibrosis: A practical approach for general medicine physicians with a focus on the medical history. J Clin Med. 2018;7(12):476. doi:10.3390/jcm7120476
2. Barratt SL, Creamer A, Hayton C, Chaudhuri N. Idiopathic pulmonary fibrosis (IPF): An overview. J Clin Med. 2018;7(8):201. doi:10.3390/jcm7080201
3. Katzenstein ALA, Myers JL. Idiopathic pulmonary fibrosis: Clinical relevance of pathologic classification. Am J Respir Crit Care Med. 1998;157(4):1301-15. doi:10.1164/ajrccm.157.4.9706039
4. American Thoracic Society. ATS/ERS international consensus statement: Idiopathic pulmonary fibrosis:

Diagnosis and treatment. Am J Respir Crit Care Med. 2000;161(2):646-64. doi:10.1164/ajrccm.161.2.ats3-00
5. Ley B, Collard HR. Epidemiology of idiopathic pulmonary fibrosis. Clin Epidemiol. 2013;5:483-92. doi:10.2147/CLEP.S54815
6. King TE Jr, Pardo A, Selman M. Idiopathic pulmonary fibrosis. Lancet. 2011;378(9807):1949-61. doi:10.1016/S0140-6736(11)60052-4
7. Song JW, Hong SB, Lim CM, Koh Y, Kim DS. Acute exacerbation of idiopathic pulmonary fibrosis: Incidence, risk factors and outcome. Eur Respir J. 2011;37(2):356-63. doi:10.1183/09031936.00159709
8. Selman M, Pardo A. Idiopathic pulmonary fibrosis: An epithelial/fibroblastic cross-talk disorder. Respir Res. 2001;3(1):3. doi:10.1186/rr178
9. Kistler KD, Nalysnyk L, Rotella P, Esser D. Lung transplantation in idiopathic pulmonary fibrosis: A systematic review of the literature. BMC Pulm Med. 2014;14:139. doi:10.1186/1471-2466-14-139
10. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011;183(6):788-824. doi:10.1164/rccm.2009-040GL
11. Nalysnyk L, Cid-Ruzafa J, Rotella P, Esser D. Incidence and prevalence of idiopathic pulmonary fibrosis: Review of the literature. Eur Respir Rev. 2012;21(126):355-61. doi:10.1183/09059180.00002512
12. Raghu G, Weycker D, Edelsberg J, Bradford WZ, Oster G. Incidence and prevalence of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2006;174(7):810-6. doi:10.1164/rccm.200602-163OC
13. King TE Jr, Tooze JA, Schwarz MI, Brown KR, Cherniack RM. Predicting survival in idiopathic pulmonary fibrosis: Scoring system and survival model. Am J Respir Crit Care Med. 2001;164(7):1171-81. doi:10.1164/ajrccm.164.7.2003140
14. Bhatt H, Jadav NK, Kumari A, Gupta R, Tanwar S, Polkowski Z, et al. Artificial neural network-driven federated learning for heart stroke prediction in healthcare 4.0 underlying 5G. Concurr Comput Pract Exp. 2024;36(8):e7911. doi:10.1002/cpe.7911
15. Vekaria D, Kumari A, Tanwar S, Kumar N. ξboost: An AI-based data analytics scheme for COVID-19 prediction and economy boosting. IEEE Internet Things J. 2020;8(21):15977-89. doi:10.1109/JIOT.2020.3024258
16. Ajakwe SO, Arkter R, Ahakonye LAC, Kim DS, Lee JM. Real-time monitoring of COVID-19 vaccination compliance: A ubiquitous IT convergence approach. In: Proceedings of the 2021 International Conference on Information and Communication Technology Convergence (ICTC); 2021; Jeju Island, Republic of Korea;2021. p. 440-5. doi:10.1109/ICTC52510.2021.9620910
17. Ryerson CJ, Hartman T, Elicker BM, Ley B, Lee JS, Abbritti M, et al. Clinical features and outcomes in combined pulmonary fibrosis and emphysema in idiopathic pulmonary fibrosis. Chest. 2013;144(1):234-40. doi:10.1378/chest.12-2403
18. Alsomali H, Palmer E, Aujayeb A, Funston W. Early diagnosis and treatment of idiopathic pulmonary fibrosis: A narrative review. Pulm Ther. 2023;9(2):177-93. doi:10.1007/s41030-023-00225-7
19. Ahmad Y, Mooney J, Allen IE, Seaman J, Kalra A, Muelly M, et al. A machine learning system to indicate diagnosis of idiopathic pulmonary fibrosis non-invasively in challenging cases. Diagnostics. 2024;14(8):830. doi:10.3390/diagnostics14080830
20. Onishchenko D, Marlowe RJ, Ngufor CG, Faust LJ, Limper AH, Hunninghake GM, et al. Screening for idiopathic pulmonary fibrosis using comorbidity signatures in electronic health records. Nat Med. 2022;28(10):2107-16. doi:10.1038/s41591-022-02010-y
21. Dack E, Christe A, Fontanellaz M, Brigato L, Heverhagen JT, Peters AA, et al. Artificial intelligence and interstitial lung disease: Diagnosis and prognosis. Investig Radiol. 2023;58(8):602-9. doi:10.1097/RLI.0000000000000988
22. Mekov E, Miravitlles M, Petkov R. Artificial intelligence and machine learning in respiratory medicine. Expert Rev Respir Med. 2020;14(6):559-64. doi:10.1080/17476348.2020.1743189
23. Soffer S, Morgenthau AS, Shimon O, Barash Y, Konen E, Glicksberg BS, et al. Artificial intelligence for interstitial lung disease analysis on chest computed tomography: A systematic review. Acad Radiol. 2021;29(Suppl 4):S226-35. doi:10.1016/j.acra.2021.07.022

24. Mäkelä K, Mäyränpää MI, Sihvo HK, Bergman P, Sutinen E, Ollila H, et al. Artificial intelligence identifies inflammation and confirms fibroblast foci as prognostic tissue biomarkers in idiopathic pulmonary fibrosis. Hum Pathol. 2021;107:58-68. doi:10.1016/j.humpath.2020.10.004
25. Handa T, Tanizawa K, Oguma T, Uozumi R, Watanabe K, Tanabe N, et al. Novel artificial intelligence-based technology for chest computed tomography analysis of idiopathic pulmonary fibrosis. Ann Am Thorac Soc. 2022;19(3):399-406. doi:10.1513/AnnalsATS.202102-170OC
26. Shi Y, Wong WK, Goldin JG, Brown MS, Kim GHJ. Prediction of progression in idiopathic pulmonary fibrosis using CT scans at baseline: A quantum particle swarm optimization-Random forest approach. Artif Intell Med. 2019;100:101709. doi:10.1016/j.artmed.2019.101709
27. Mandal S, Balas VE, Shaw RN, Ghosh A. Prediction analysis of idiopathic pulmonary fibrosis progression from OSIC dataset. In: Proceedings of the 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON); 2020 Oct 2-4; Greater Noida, India. doi:10.1109/GUCON48875.2020.9231203
28. Wang Z. Deep learning approach for auto-detecting idiopathic pulmonary fibrosis prediction. In: Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID); 2021 May 28-30; Guangzhou, China. p. 283-90. doi:10.1109/AIID51893.2021.9456465
29. Wu X, Yin C, Chen X, Zhang Y, Su Y, Shi J, et al. Idiopathic pulmonary fibrosis mortality risk prediction based on artificial intelligence: The CTPF model. Front Pharmacol. 2022;13:878764. doi:10.3389/fphar.2022.878764
30. Thillai M, Oldham JM, Ruggiero A, Kanavati F, McLellan T, Saini G, et al. Deep learning-based segmentation of CT scans predicts disease progression and mortality in IPF. Am J Respir Crit Care Med. 2024. [Epub ahead of print]. doi:10.1164/rccm.202311-2185OC
31. Ali S, Hussain A, Aich S, Park MS, Chung MP, Jeong SH, et al. A soft voting ensemble-based model for the early prediction of idiopathic pulmonary fibrosis (IPF) disease severity in lungs disease patients. Life. 2021;11(10):1092. doi:10.3390/life11101092
32. Ajakwe SO, Ajakwe IU, Jun T, Kim DS, Lee JM. CIS-WQMS: Connected intelligence smart water quality monitoring scheme. Internet Things. 2023;23:100800. doi:10.1016/j.iot.2023.100800
33. Huang J, Li YF, Xie M. An empirical analysis of data preprocessing for machine learning-based software cost estimation. Inf Softw Technol. 2015;67:108-27. doi:10.1016/j.infsof.2015.07.004
34. Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: A data perspective. ACM Comput Surv. 2017;50(6):94. doi:10.1145/3136625
35. Ajakwe SO, Ihekoronye VU, Ajakwe IU, Jun T, Kim DS, Lee JM. Connected intelligence for smart water quality monitoring system in IIoT. In: Proceedings of the 13th International Conference on Information and Communication Technology Convergence (ICTC); 2022; Jeju Island, Republic of Korea; 2022. p. 2386-91. doi:10.1109/ICTC55196.2022.9952837
36. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. Bioinformatics. 2007;23(19):2507-17. doi:10.1093/bioinformatics/btm344
37. Akhtar F, Li J, Pei Y, Xu Y, Rajput A, Wang Q. Optimal features subset selection for large for gestational age classification using gridsearch based recursive feature elimination with cross-validation scheme. In: Proceedings of the International Conference on Frontier Computing; 2019; Kitakyushu, Japan. Springer; 2019. doi:10.1007/978-981-15-3250-4_102
38. Ihekoronye VU, Ajakwe SO, Kim DS, Lee JM. Hierarchical intrusion detection system for secured military drone network: A perspicacious approach. In: Proceedings of the MILCOM 2022 - 2022 IEEE Military Communications Conference; 2022; Rockville, MD, USA; 2022. p. 336-41. doi:10.1109/MILCOM55135.2022.10017460
39. Yahaya CAC, Firdaus A, Mohamad S, Ernawan F, Razak MFA. Automated feature selection using Boruta algorithm to detect mobile malware. Int J Adv Trends Comput Sci Eng. 2020;9(6):9029-36. doi:10.30534/ijatcse/2020/01962020
40. Han W, Huang Z, Li S, Jia Y. Distribution-sensitive unbalanced data oversampling method for medical diagnosis. J Med Syst. 2019;43(2):39. doi:10.1007/s10916-019-1157-9

41. Ajakwe SO, Deji-Oloruntoba O, Olatubosun S, Duorinaah F, Bayode IA. Multidimensional perspective to data preprocessing for model cognition verity. In: Recent Trends and Future Direction for Data Analytics. Hershey (PA): IGI Global; 2024. doi:10.4018/979-8-3693-3609-0.ch006
42. Juarez MM, Chan AL, Norris AG, Morrissey BM, Albertson TE. Acute exacerbation of idiopathic pulmonary fibrosis. Chest. 2007;132(5):1652-8. doi:10.1378/chest.07-0299
43. Lee HJ, Im JG, Ahn JM, Yeon KM. Lung cancer in patients with idiopathic pulmonary fibrosis: CT findings. J Comput Assist Tomogr. 1996;20(6):979-82. doi:10.1097/00004728-199611000-00018