Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 8, No. 3, pp. 317-334, August 2026

RESEARCH PAPER

OPEN ACCES

Optimization of Random Forest for Outpatient Cancellation Prediction Using PSO and Behavioral Feature Engineering

Azmi Abiyyu Dzaky[1], Farrikh Alzami, and Pujiono[3]

Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Semarang, Indonesia

Abstract

Outpatient cancellations and no-shows, defined respectively as proactive withdrawals and silent attendance failures, were consolidated into a binary target variable representing failed appointments that result in a completed visit. This issue significantly reduces healthcare efficiency and increases operational costs. While machine learning is widely applied, the existing methods face critical trade-offs between computational resource demands and data integrity, often relying on synthetic resampling such as SMOTE that introduces artificial noise. This study further aims to develop an optimized, lightweight Decision Support System designed as a core engine for potential future Hospital Digital Twin and AI Command Center architectures. The primary contribution is Behavioral Feature Engineering utilizing a 12-month retrospective patient history as a natural data enrichment strategy, offering a robust alternative to artificial data generation. The proposed framework employs an optimized Random Forest classifier combined with Target Encoding, Mutual Information-based feature selection, and Particle Swarm Optimization for hyperparameter tuning. To ensure computational efficiency, the optimization was performed on a stratified 2% sample of 10,690 records from a large-scale dataset containing 686,387 records from three hospitals. The final model achieved an accuracy of 0.871, an F1-score of 0.870, and a ROC-AUC of 0.940, effectively surpassing recent 2024 state-of-the-art benchmarks even under significant class imbalance. Furthermore, Particle Swarm Optimization achieved a 75.6% reduction in tree complexity, requiring only 122 estimators instead of 500, while preserving peak performance. Multi-scheme validation under strict chronological temporal splits and patient-grouped cross-validation proved high generalization stability, yielding robust F1-scores of 0.855 and 0.867, with a narrow 95% confidence interval of [0.8692, 0.8728]. The entire end-to-end pipeline was completed in under 60 minutes using standard CPU resources. In conclusion, the proposed framework provides a robust and sustainable predictive engine suitable for proactive clinical resource management and future hospital digital integration.

Paper History

Received May 30, 2026
Revised June 20, 2026
Accepted June 28, 2026
Published July 8, 2026

Keywords

Behavioral Feature Engineering;
Decision Support System;
Hospital Digital Twin;
Particle Swarm Optimization;
Random Forest

Author Email

azmiabiyyudzaky456@gmail.com
alzami@dsn.dinus.ac.id
pujiono@dsn.dinus.ac.id

I. Introduction

In the context of digital healthcare, the efficiency of outpatient flow is a critical factor in determining the performance of healthcare facilities [1][2]. Effective scheduling systems play a key role in managing patient arrivals [3][4][5]. However, patient cancellations and incomplete appointments remain as significant operational challenges [6][7]. In clinical scheduling literature, a no-show refers to a failure to attend a confirmed appointment without prior notification, while a cancellation refers to a confirmed appointment withdrawal in which the patient proactively notifies the facility. These two behaviors carry distinct operational consequences and may respond to different predictive signals. Previous studies have shown that no-show rates are still relatively high. For example, when the waiting time is only one day, the cancellation and no-show rates are approximately 12.3% and 10.3%, respectively. Meanwhile, when the waiting time is extended to one and a half months, the cancellation rate can increase to 42%, while the no-show rate increases to 16.3% [8]. This situation leads to wasted clinical time, increased operational costs, and difficulties in resource reallocation [9][10][11]. Therefore, hospital administrations require a reliable Decision Support System (DSS) to identify high-risk patients early and mitigate these losses [12][13][14][15], particularly to identify high-risk cancellations [16][17]. The integration of Machine Learning (ML) enhances these systems by offering more adaptive predictions compared to conventional rule-based approaches [18][18][20]. Various algorithms, including K-Nearest Neighbors (KNN) and Support Vector Machines (SVM), have been utilized in this domain [21][22]. Among these, Random Forest (RF) has

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

emerged as a widely adopted method [23] due to its proficiency in processing large tabular datasets through ensemble learning, which stabilizes predictions and minimizes overfitting [24]. However, the efficacy of RF is closely linked to hyperparameter configuration and data quality [25][26][27], necessitating robust feature engineering to improve data representation [28]. Despite these advances, existing models still face critical trade-offs between computational efficiency and predictive reliability. For example, Dashtban & Li [29]. proposed a complex Deep Learning (SDAE) model that achieved a relatively low F1-score of 0.288 and an AUC of 0.704, suggesting that complex architectures may underperform when input features provide limited predictive signals. In contrast, Yi Yang et al. [30]. achieved a high accuracy of 0.899 using RF, but their model required up to 137 hours of training time, which significantly hinders real-time applicability. To address class imbalance, Ocampo Osorio et al. [31]. employed Bagging RF combined with SMOTE, reaching an accuracy of 0.848; however, this approach relies on synthetic data generation that can introduce artificial noise into natural behavior patterns. Similarly, the PSO-RF hybrid framework proposed by Singh et al. [32]. also depends on SMOTE-based resampling. Furthermore, while Porto and Fogliatto [33] incorporated time-related features, their optimization process demanded substantial computational resources.

Despite these advances, two primary gaps remain unaddressed in the literature. First, the effectiveness of retrospective behavioral feature engineering as a systematic substitute for synthetic resampling in large-scale clinical data has not been fully evaluated. Second, the integration of a PSO-based optimization within an efficient workflow suitable for periodic retraining has not been demonstrated at this data scale. This study addresses these gaps through a structured 26-scenario ablation design applied to 686,387 real appointment records from three Indonesian private hospitals. To enhance operational efficiency while reducing computational costs and training time, this study proposed an optimized prediction framework designed as a conceptual core engine for potential future Hospital Digital Twin (HDT) and AI Command Center (AICC) architectures [34][35]. By accurately predicting cancellations, this model supports the strategic shift of Customer Relationship Management (CRM) from reactive to proactive patient engagement [36][37][38][39]. In this context, Particle Swarm Optimization (PSO) provides an efficient approach to explore complex parameter spaces compared to exhaustive search methods [32]. Unlike Genetic Algorithms, which involve complex crossover and mutation operations, or Bayesian Optimization, which assumes a smooth objective surface that may not hold for discrete spaces, PSO utilizes simple velocity-position updates that scale effectively for Random Forest configurations [40].

The contributions of this research are twofold. First, we implemented behavioral feature engineering utilizing a 12-month retrospective patient history as a natural data enrichment strategy. This approach demonstrates that enhancing original data quality provides more robust predictive signals than synthetic resampling techniques such as SMOTE, which may introduce artificial noise. Second, we integrated Target Encoding, Mutual Information, and Particle Swarm Optimization (PSO) to construct a streamlined and lightweight predictive framework. Rather than focusing on marginal accuracy gains through complex ensemble architectures, this research emphasized model simplification and computational efficiency. The proposed framework successfully identified an optimal, low-complexity structure that maintains peak performance while reducing tree complexity, thereby minimizing the computational footprint and enabling an end-to-end pipeline execution in under 60 minutes. Such efficiency is highly suitable for periodic retraining and confirms the model's viability for future integration within Hospital AI Command Center environments.

The primary objective of this study is to evaluate whether behavioral feature engineering provides stronger predictive signals than raw operational features, and to determine if PSO-based optimization can recover model performance after noise-reducing feature selection while yielding a simpler model structure. Furthermore, the study assessed whether the resulting framework satisfied the performance and efficiency requirements for periodic deployment in a real-world hospital operational setting.

This study is structured as follows: Section II describes the research methodology, including the dataset characteristics, the proposed research workflow, and the development of the optimized PSO-RF framework. Section III presents the experimental results and the detailed ablation study. Section IV provides an analytical discussion on the impact of behavioral data, benchmarking against state-of-the-art studies, and practical implications for healthcare management. Finally, Section V concludes the study and outlines future work.

II. Materials and Method

A. Dataset

The dataset for this study was sourced from internal operational reports of three private hospitals located within the same district in Indonesia. These facilities provide services across various categories, including national health insurance (BPJS), Intermediate, and Premium classes. The raw data comprises transaction records collected over a one-year observation period for outpatient appointments, totaling 686,387 records with a file size of 113 MB. This dataset reflects natural variations in demographic and clinical characteristics across the patient population. In this study, the target variable, *is_cancel_appointment*, is defined to identify outpatient appointments that do not result in a completed patient visit. Positive cases were derived from two specific operational outcomes: (1) No-Shows, where patients booked appointments via the mobile application but failed to attend without prior notification (accounting for approximately 97.1% of positive cases); and (2) Cancellations are where appointments are formally

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

Table 1. Dataset Overview

No	Category	Description	Column
1	Demographics	Non-personal identity and background	Patient Age, Gender, Payer Type, Is New Patient, Contact ID (Hashed)
2	Appointment Details	Scheduling and execution timing	Appointment Id, Booking Date, Appointment Date, Day Type, Lead Time (Diff Day)
3	Clinical and Operasional	Medical service provider details	Doctor ID (Hashed), Specialty Name, Hospital Class, Booking Source
4	Enviromental	Weather conditions on appointment day	Weather Status, Temperature, Humidity, Wind Speed, Rain
5	Behavioral History	Patient interaction track record (12 Months)	Total Booking, Total Cancel, Total Not Cancel (1 Year)
6	Target	Classification label	Is_cancel_appointment (0 = No Cancel, 1 = Cancel)

withdrawn either through the mobile app or the front office (representing approximately 2.9% of positive cases). While some cancellations may stem from provider-related changes, such as physician absences, these instances represent a negligible fraction (less than 1% of the total dataset). Since both outcomes result in an unused appointment slot, they were consolidated into a single binary target variable. The final dataset consists of 486,865 completed appointments (70.93%) and 199,522 incomplete appointments (29.07%), as illustrated in Fig. 1.

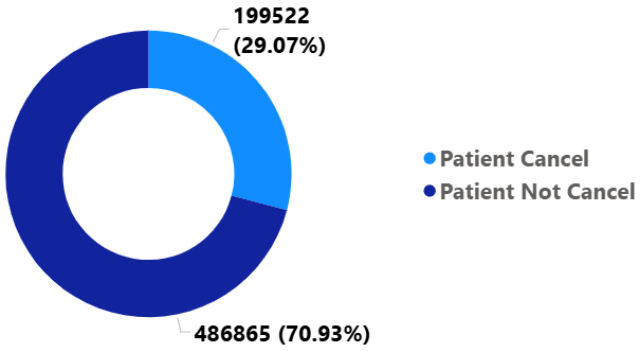

Fig. 1. Distribution of Dataset

Beyond standard operational records, the dataset was enriched with external and historical attributes. Based on previous research [8][37], which indicates that environmental conditions and attendance history significantly influence patient behavior, additional variables were integrated into the analysis. It includes local weather data based on hospital coordinates and approximately one year of retrospective transaction history for each patient. Based on institutional database flags, the dataset contains 24,712 records (3.6%) identified as new patients and 661,675 records (96.4%) as returning patients. For model training, behavioral history features were treated as continuous numerical variables to capture the full spectrum of interaction density. In cases where prior transactional signals were absent, a zero-baseline imputation was applied to represent a neutral behavioral profile. It ensures that the Random Forest model can effectively process both first-time attendees and frequent visitors within a unified predictive framework. The complete feature set used in this study is presented in Table 1.

B. Research Workflow

The research workflow was systematically designed as a multi-layered framework, as illustrated in Fig. 2 The process initiated at the Datasource Layer, where 686,387 records were integrated with environmental data. In the Enrichment & Cleaning Layer, comprehensive feature engineering was performed to enrich the dataset with behavioral context. A core contribution of this phase was the retrospective tracing of patient activity over a 12-month period to generate historical behavioral features, including total bookings, cumulative cancellations, and successful attendances. To prevent temporal leakage, a strict historical partitioning strategy was applied, dividing the dataset into two non-overlapping windows: a Historical Baseline Window (January – December 2024) and a Target Prediction Window (January – November 2025). Behavioral features were compiled exclusively from the 2024 window, ensuring that future transactional information from 2025 could not influence the baselines. For new patients with no records in the 2024 window, behavioral counts were systematically imputed with 0, representing a neutral profile. Additionally, environmental variables were integrated through the OpenWeather API, followed by a cleaning process to address outliers and remove irrelevant features [41].

At the Data Preprocessing Layer, categorical variables were transformed into numerical formats suitable for machine learning [42]. Binary and ordinal columns utilized direct mapping, while high-cardinality features were processed using Target Encoding (TE). Crucially, TE parameters were calculated exclusively on the training set to prevent data leakage and ensure model integrity [43]. Following preprocessing, the dataset was divided using a stratified 80:20 hold-out ratio for training and testing,

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

preserving the original 70.93/29.07 class distribution across both subsets.

Within the Ablation Study & Learning Layer, a 2% stratified sample of the training set (approximately 10,690 records) was utilized during the hyperparameter optimization phase. This strategy enabled Particle Swarm Optimization (PSO) and RandomizedCV to identify optimal parameters more efficiently, significantly reducing computational overhead. Once the optimal parameters were identified, the final Random Forest model was trained using the complete training set (534,480 records). This phase included an extensive ablation study covering 26 scenarios designed to evaluate the effectiveness of encoding methods (LE vs. TE), resampling strategies (Raw Data vs. SMOTE), feature enhancements (Behavioral and Weather), and optimization techniques (Default vs. PSO). Model performance was evaluated using F1-score and ROC-AUC metrics.

To guarantee robustness and clinical viability, the best-performing model underwent evaluation in the Model Evaluation Layer, which included an interpretability audit using SHAP (SHapley Additive exPlanations) and Gini Importance to validate the clinical logic of the predictors. Subsequently, in the Model Validation Layer, the model was subjected to three strict validation schemes: (1) an overfitting analysis to measure performance gaps between training and testing; (2) a chronological temporal split validation (Training: January–September 2025, Testing: October–November 2025); and (3) a 5-fold patient-grouped cross-validation (GroupKFold) grouped by Contact Id to prevent patient-level data leakage.

Finally, a Deployment Concept was established to bridge the gap between research and practical application. This conceptual framework illustrates how the optimized engine serves as a core for a Hospital AI Command Center (AICC) and Digital Twin infrastructure. It includes an AICC Predictive Engine for real-time probability scoring, an Intelligence Dashboard for operational monitoring, Scalable Infrastructure for data architecture, and an Actionable CRM system for proactive patient reminders and slot reallocation. The detailed logic and implementation steps of the proposed method are provided in Table 2 (Algorithm 1).

C. Random Forest (RF)

Random Forest (RF) was used in this study as an ensemble learning method to identify patterns in historical appointment data [44]. RF is preferred over a single decision tree due to its ability to reduce overfitting through bagging and bootstrap aggregation, where multiple trees are trained on randomly sampled data to achieve lower variance and more stable predictions [45]. Fundamentally, RF operates by training independent decision trees and applying majority voting for the final decision [46]. The core component of this algorithm is the decision tree, which recursively partitions data from the root node to the leaf nodes to produce more homogeneous clusters [47]. The quality of each partition was evaluated using an impurity measure, where lower values indicate better attributes for data partitioning [48] In decision tree theory, Entropy $H(S)$ is a fundamental measure of impurity or uncertainly in a dataset S, as defined in Eq. (1) [49].

Table 2. Algorithm 1

Proposed PSO-RF Framework
BEGIN
Input : Raw Transactional Data Records (D_{raw})
Output: Optimized RF Model (θ_{opt}), performance metrics, and stability
Stage I: Data Preparation & Feature Enrichment
Generate retrospective behavioral features Fb (bookings, cancellations, attendances)
Impute new patients (F_b=*0*) and merge environmental weather data F_e via API
Stage II: Preprocessing & Feature Selection
Drop noise-inducing temporal columns & Apply Target Encoding (TE)
Execute feature selection using Mutual Information threshold (MI > 0.5)
Perform stratified 80:20 split to generate Training Set D_{train} & Test Set D_{test}
Stage III: Metaheuristic Optimization (*PSO*)
Draw stratified 2% sample D_{sample} from D_{train}
Initialize swarm particles P with 10 particles over spaces S
n_estimators ∈ [50,300], max_depth ∈ [5,30], min_split ∈ [2,20], min_leaf ∈ [1,10]
Set maximum iterations T = 10
While $t \leq T$ **Do:**
For each particle $p_i \in P$ **Do**:
Evaluate fitness $f(p_i)$ = mean F1-Weighted via 3-fold CV on D_{sample}
Update personal best P_{best} and global best G_{best} based on $f(p_i)$
Update velocity and position
End For
$t = t1$
End While
Retrieve optimal hyperparameters $\theta_{opt} = G_{best}$
Stage IV: Final Training & Multi-Scheme Validation
Train final RF model on full D_{train} using θ_{opt}
Validate D_{test} to calculate metrics (Acc, Prec, Rec, F1, ROC-AUC, PR-RUC)
Perform 5-fold GroupKFold (by Contact Id)
Perform chronological Temporal Validation (Train Jan-Sep 2025; Test Oct-Nov 2025)
Return Final_Model, Metrics, and Stability_Score
END

$$H(S) = \sum_{i=1}^{c} p_i log_2(p_i) \quad (1)$$

Based on entropy, Information Gain (IG) can be used to estimate the reduction in uncertainty after splitting the data based on attribute A, as expressed in Eq. (2) [50]

$$IG(S,A) = H(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} H(S_v) \quad (2)$$

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

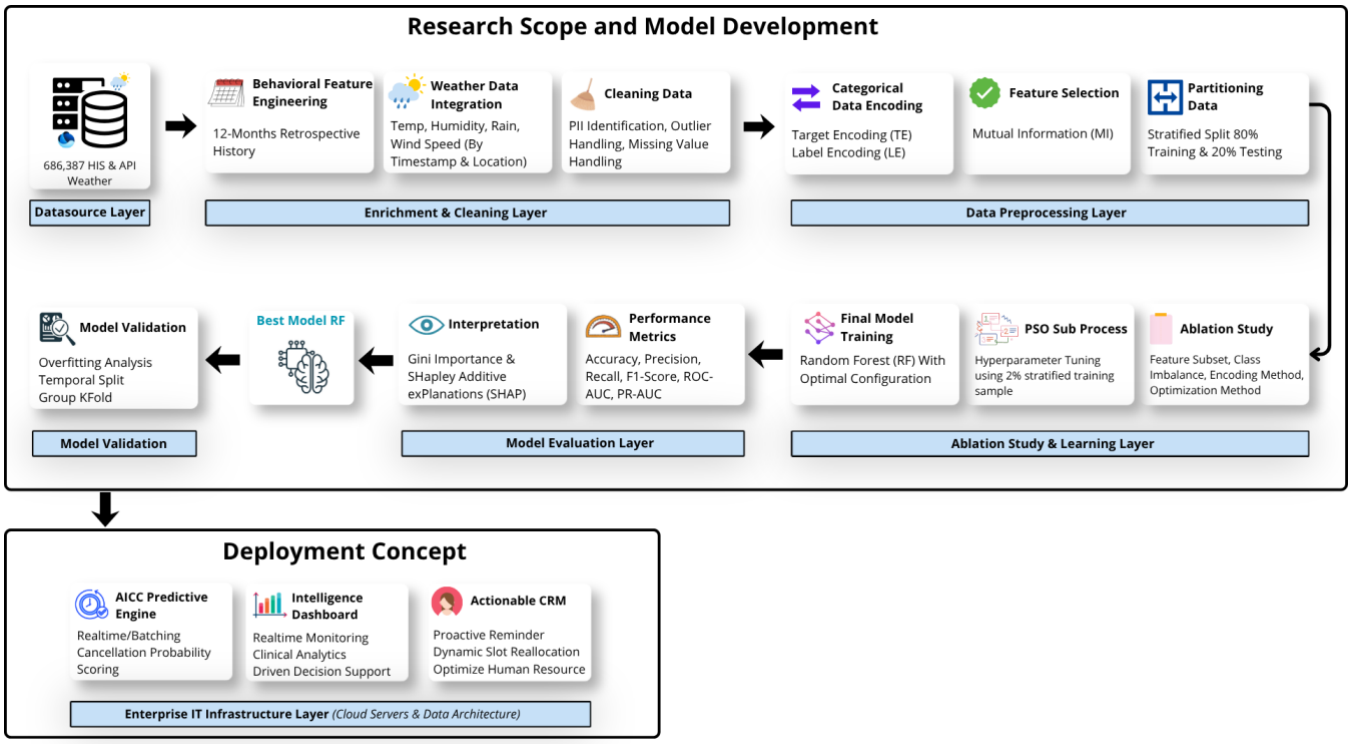

Fig. 2. Research Workflow Diagram

However, to enhance computational efficiency for large-scale datasets, this study utilized the Gini Index as the primary impurity measure [51]. The Gini Index calculates the probability of misclassification without the need for logarithmic calculations. The Gini impurity for a node is expressed in Eq. (3)

$$Gini(S) = 1 - \sum_{i=1}^{c} {p_i}^2 \quad (3)$$

where c represents the number of classes and P_i is the proportion of data in class i. During the tree construction, the attribute that minimizes the Weighted Gini Index Eq. (4) was chosen as the optimal split

$$Gini_{split}(S, A) = \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Gini(S_v) \quad (4)$$

Once the forest is constructed, the individual predictions from each tree T are aggregated. The final classification $\hat{y}RF$ was determined via majority voting, as shown in Eq. (5).

$$\hat{y}RF = mode(\hat{y}_1, \hat{y}_2, \hat{y}_3 . \hat{y}_T) \quad (5)$$

Additionally, the model utilized a confidence score to determine the binary classification threshold, with a default probability threshold of 0.5 applied to distinguish between completed visits and the cancellation class, aligning with standard predictive modeling practices in clinical scheduling [52]. The effectiveness of this model is significantly influenced by hyperparameter configurations, such as the number of trees and tree depth [53]. Therefore, metaheuristic optimization is required to balance predictive performance and computational complexity.

D. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a population-based metaheuristic algorithm designed to identify global optimum solutions by mimicking the social behavior of individuals, known as particles, within a swarm. Each particle represents a candidate solution that traverses the multi-dimensional search space based on its personal best performance (P_{best}) and the overall success of the entire swarm (G_{best}) [54]. The primary advantages of PSO include its ease of implementation, steady convergence, and efficiency in handling non-linear hyperparameter spaces for complex machine learning models [40]. The movement of each particle is governed by velocity and position update equations. The velocity v_i^{t+1} incorporates the particle's inertia, cognitive experience, and social influence, as defined in Eq. (6), while the position v_i^{t+1} is updated in Eq. (7) [55].

$$v_i^{t+1} = w \cdot v_i^t + c_1 \cdot r_1 \cdot (P_{best} - x_i^t) + c_2 \cdot r_2 \cdot (G_{best} - x_i^t) \quad (6)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (7)$$

In this equation, w represents the inertia weight, while c_1 and c_2 denote the cognitive and social acceleration constants, respectively. The variables $r_1$and r_2 are pseudo-random numbers between 0 and 1 that ensure stochastic exploration to prevent the swarm from being trapped in local optima. To evaluate the quality of each particle's position, a Fitness Function $f(x_1)$ was utilized. In this study, the fitness was measured using the mean F1-weighted score derived from K-Fold cross-validation on the training sample as expressed in Eq. (8)

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

$$f(x_i) = \frac{1}{K}\sum_{k-1}^{K} F1_{weighted,k} \quad (8)$$

where $K = 3$ represents the 3-fold cross-validation used during the tuning phase. PSO was configured with 10 particles and 10 iterations, with $w = 0.7$ and $c_1 = c_2 = 1.5$. The search bounds for the hyperparameters were: *n_estimators* ∈ [50, 300], *max_depth* ∈ [5, 30], *min_samples_split* ∈ [2, 20] and *min_samples_leaf* ∈ [1, 10]. To maintain computational efficiency, the fitness function was evaluated on a stratified 2% sample of the training set. The representativeness of this sample size was validated using the Margin of Error (E) formula for finite population, defined in Eq. (9)

$$E = Z\sqrt{\frac{p(1-p)}{n} \times \frac{N-n}{N-1}} \quad (9)$$

Given a population $N = 534{,}480$, a sample $n = 10{,}690$, and a confidence level $Z = 2.576\ (99\%)$, the Margin of Error is approximately 1.23%, ensuring sufficient statistical representativeness. This sampling strategy is computationally pragmatic; executing a 10-particle, 10-iteration PSO with 3-fold cross-validation requires training the Random Forest model 300 times. On the full dataset, this process would exceed 12+ hours of high-CPU computation, which is impractical for periodic retraining. When the optimal parameters identified, θ_{opt} are then retrieved as shown in Eq. (10) before being applied to the final model.

$$\theta_{opt} = argmax\ f(x_i) \quad (10)$$

E. Data Encoding & Feature Selection

A major challenge in processing hospital datasets is the presence of categorical variables with high cardinality, such as specialty_name and booking_source. Traditional One-Hot Encoding is often inefficient for handling these variables due to the curse of dimensionality, while Label Encoding (LE) can introduce ordinal bias by implying non-existent relationships between categories [56][57]. To address these limitations, this study applied Target Encoding (TE) to features with high cardinality. TE replaced each category with the average probability of the target variable for that category, enabling the model to better capture the relationship between features and cancellation outcomes. The application of TE followed strict protocols to prevent overfitting and data leakage. Specifically, the encoding parameters were calculated exclusively using the training set and subsequently mapped to the test set. For a category c_i the encoded value $\hat{x}_c$ was calculated as the mean of the target variable y for all instances n_c belonging to that category, as shown in Eq. (11), isolated within training data to prevent leakage.

$$\hat{x}_c = \frac{1}{n_c}\sum_{i \in n_c} y_i \quad (11)$$

For categories present in the test data but absent during training, a Global Mean Fallback technique was utilized. This approach assigned the overall average cancellation rate of the training population to unknown categories, ensuring the encoding process remains stable and robust against out-of-distribution samples. Furthermore, to ensure that the model focuses on the most predictive signals and reduces noise, this study implemented an information-theoretic feature selection approach using Mutual Information (MI). MI measures the strength of the relationship between a feature X and the target variable Y by quantifying the amount of information obtained about Y through X. This relationship was founded on Entropy $H(Y)$, which represents the uncertainty of the target variable, defined in Eq. (12)

$$H(Y) = -\sum_{y \in Y} p(y) log\ \mathrm{p(y)} \quad (12)$$

The dependency between the feature and the target was then formally expressed as the Mutual Information $I(X;Y)$, as defined in Eq. (13) [58][59].

$$I(X;Y) = \sum_{x \in X}\sum_{y \in Y} (p(x,y) log\left(\frac{p(x,y)}{p(x) \cdot p(y)}\right) \quad (13)$$

where $p(x,y)$ is the joint probability distribution of X and Y, while $p(x)$ and $p(y)$ are their respective marginal probability distributions. In this study, features with an MI score below a threshold of 0.05 were excluded. This process ensures a streamlined feature subset, where the reduction in uncertainty of the target variable Y given feature X is maximized, as represented by the relationship in Eq. (14).

$$I(X;Y) = H(Y) - H(Y|X) \quad (14)$$

F. Matrix Evaluation

Evaluating model performance is crucial for assessing the effectiveness of the Random Forest model in predicting patient cancellations. While Accuracy in Eq. (15) provides a general overview of correct predictions, it can be misleading in imbalanced datasets [60]. Therefore, this study prioritized metrics that focus on the minority class, including Precision in Eq. (16), which measures the proportion of true positive predictions, and Recall in Eq. (17), which assesses the model's ability to identify all actual cancellation instances [61][62]. The F1-Score in Eq. (18) was employed as the harmonic mean to provide a robust balance between these two metrics [63].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$F1\ Score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \quad (18)$$

The model's discriminative capacity across various thresholds was evaluated using the Receiver Operating Characteristic (ROC-AUC) and Precision-Recall (PR-

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

AUC) curves. The False Positive Rate (FPR), essential for ROC measurement, is defined in Eq. (19), while the ROC-AUC is mathematically represented as the area under the curve in Eq. (20).

$$FPR = \frac{FP}{FP + TN} \quad (19)$$

$$ROC - AUC = \int_0^1 TPR(FPR^{-1}(t))dt \quad (20)$$

To quantify the structural efficiency achieved through PSO, the Tree Complexity Reduction (R) percentage was calculated using Eq. (21), comparing the optimized tree count (T_{opt}) with the default ensemble size (T_{def}).

$$R = \left(\frac{T_{def} - T_{opt}}{T_{def}}\right) \times 100\% \quad (21)$$

Finally to ensure the statistical stability of the results, a 95% Confidence Interval (CI) for the F1-score was calculated based on the standard error and the Z-score for a normal distribution, as expressed in Eq. (22).

$$CI = \hat{p} \pm Z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (22)$$

III. Results

A. Experimental Setup

The experiments in this study were executed within the Google Colab cloud platform utilizing a 12 GB RAM environment. The computational system ran on an Intel Xeon CPU @ 2.20GHz, no GPU or hardware accelerators were used as the Random Forest model was executed on CPU cores. The software stack was built on Python (v3.12.13), utilizing Pandas (v2.2.2), NumPy (v2.0.2), Scikit-Learn (v1.6.1) for machine learning, Imbalanced-Learn (v0.14.2) for resampling, SHAP (0.52.0) and PySwarms (v1.3.0) for particle swarm optimization. A fixed random seed (random_state=42) was applied across all data splitting, sampling, and model training operations to guarantee consistent result replication. An ablation study comprising 26 experimental scenarios was conducted to assess the contribution of each component. The results are summarized in Table 3.

B. Baseline Evaluation and Synthetic Resampling

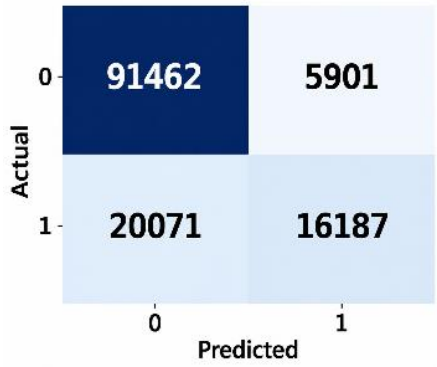

Fig. 3. Confusion Matrix Scenario 2

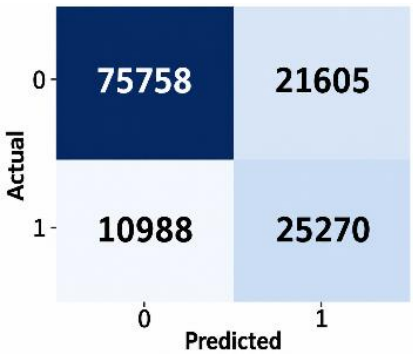

Fig. 4. Confusion Matrix Scenario 3

The initial phase of the ablation study (Scenarios 1–6) was conducted to establish baseline performance using raw operational data without engineered behavioral or environmental features. This stage specifically evaluated the effectiveness of Target Encoding (TE) compared to Label Encoding (LE), along with the impact of SMOTE in handling class imbalance. As presented in Table 3, The default Random Forest model using LE (Scenario 1) achieved a baseline accuracy of 0.795. Optimization using RandomizedSearchCV (Scenario 2) slightly improved this to 0.806, the integration of SMOTE in Scenario 3 led to a performance degradation, with accuracy and F1-score dropping to 0.756 and 0.760, respectively.

Analysis of the confusion matrices in Fig. 3 and Fig. 4 quantifies this trade-off: while SMOTE successfully reduced false negatives by 45.2% (from 20,071 to 10,988), it simultaneously triggered a 266% surge in false positives, jumping from 5,901 to 21,605. This confirms that synthetic noise severely blurred the model's decision boundaries. From a clinical decision-support perspective, this error pattern is counterproductive. An excessive volume of false positives leads to 'alert fatigue' among administrative staff, who may waste resources by proactively contacting patients who actually intend to attend. Conversely, false negatives represent a direct financial loss, as they lead to unused appointment slots that could have been reallocated to other patients.

Target Encoding (TE) demonstrated superior robustness, with Scenario 5 achieving an accuracy of 0.805 and a ROC-AUC of 0.851. Unlike the LE framework, SMOTE combined with TE (Scenario 6) remained stable at 0.800 accuracy, avoiding the sharp decline observed in Scenario 3. This stability indicates that TE maintains categorical integrity by mapping categories directly to target probabilities rather than arbitrary rankings. However, the performance ceiling of approximately 80% suggests that raw operational data alone is insufficient for high-precision modeling, highlighting the necessity of integrating a broader behavioral context in subsequent stages.

C. The Dominance of Behavioral Feature

To address the limitations of raw data, the experiment was expanded to include behavioral feature engineering (Scenarios 7–12). These features capture 12 months of historical patterns, such as cumulative cancellations and

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

Table 3. Ablation Experiment Result

No	Scenario	Encode	Column Behavior	Column Weather	Acc	Prec	Rec	F1 Score	ROC AUC
1	Default RF	LE	No	No	0.795	0.780	0.800	0.780	0.839
2	Enhanced RF (RandomizedCV)	LE	No	No	0.806	0.800	0.810	0.790	0.806
3	SMOTE + Enhanced RF (RandomizedCV)	LE	No	No	0.756	0.780	0.760	0.760	0.840
4	Default RF	TE	No	No	0.796	0.780	0.800	0.790	0.796
5	Enhanced RF (RandomizedCV)	TE	No	No	0.805	0.800	0.810	0.790	0.851
6	SMOTE + Enhanced RF (RandomizedCV)	TE	No	No	0.800	0.790	0.800	0.790	0.851
7	Default RF	LE	Yes	No	0.868	0.870	0.870	0.870	0.936
8	Enhanced RF (RandomizedCV)	LE	Yes	No	0.870	0.870	0.870	0.870	0.939
9	SMOTE + Enhanced RF (RandomizedCV)	LE	Yes	No	0.850	0.870	0.850	0.850	0.934
10	Default RF	TE	Yes	No	0.867	0.870	0.870	0.870	0.936
11	Enhanced RF (RandomizedCV)	TE	Yes	No	0.871	0.870	0.870	0.870	0.940
12	SMOTE + Enhanced RF (RandomizedCV)	TE	Yes	No	0.867	0.870	0.870	0.870	0.938
13	Default RF	LE	No	Yes	0.805	0.790	0.800	0.790	0.847
14	Enhanced RF (RandomizedCV)	LE	No	Yes	0.808	0.800	0.810	0.790	0.852
15	SMOTE + Enhanced RF (RandomizedCV)	LE	No	Yes	0.758	0.780	0.760	0.760	0.838
16	Default RF	TE	No	Yes	0.805	0.790	0.810	0.790	0.848
17	Enhanced RF (RandomizedCV)	TE	No	Yes	0.808	0.800	0.810	0.790	0.852
18	SMOTE + Enhanced RF (RandomizedCV)	TE	No	Yes	0.799	0.790	0.800	0.790	0.848
19	Default RF	LE	Yes	Yes	0.868	0.870	0.870	0.870	0.937
20	Enhanced RF (RandomizedCV)	LE	Yes	Yes	0.871	0.870	0.870	0.870	0.939
21	SMOTE + Enhanced RF (RandomizedCV)	LE	Yes	Yes	0.848	0.870	0.850	0.850	0.933
22	Default RF	TE	Yes	Yes	0.868	0.870	0.870	0.870	0.937
23	Enhanced RF (RandomizedCV)	TE	Yes	Yes	0.871	0.870	0.870	0.870	0.940
24	SMOTE + Enhanced RF (RandomizedCV)	TE	Yes	Yes	0.868	0.870	0.870	0.870	0.938
25	Enhanced RF (RandomizedCV) + Mutual Information	TE	Yes	Yes	0.868	0.870	0.870	0.870	0.938
26	[***Proposed Method***] Enhanced RF (PSO) + Mutual Information	TE	Yes	Yes	0.871	0.870	0.870	0.870	0.940

lead times, to evaluate the impact of data enrichment on predictive performance. As documented in Table 3, the inclusion of behavioral context significantly improved model performance without requiring complex algorithmic modifications. Integrating behavioral signals into the baseline RF model (Scenario 7) increased accuracy to 0.868, a substantial 7.3% improvement from Scenario 1. Peak performance was achieved in Scenario 11, combining TE, RandomizedSearchCV, and behavioral data, resulting in an accuracy of 0.871, an F1-score of 0.870, and a ROC-AUC of 0.940.

This performance improvement is visualized in Fig. 5 which presents a comparative bar chart of key evaluation

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

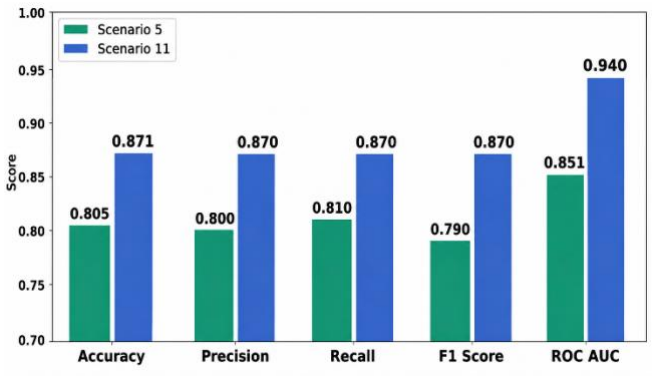

Fig. 5. Comparison of Matrix Evaluation Scenario 5 vs Scenario 11

metrics between the baseline (Scenario 5) and the behaviorally enriched model (Scenario 11). The Y-axis represents the metric score (ranging from 0.70 to 1.00), illustrating an expansion in ROC-AUC from 0.851 to 0.940. These results indicate that historical behavior offers a more robust predictive signal for patient outcomes

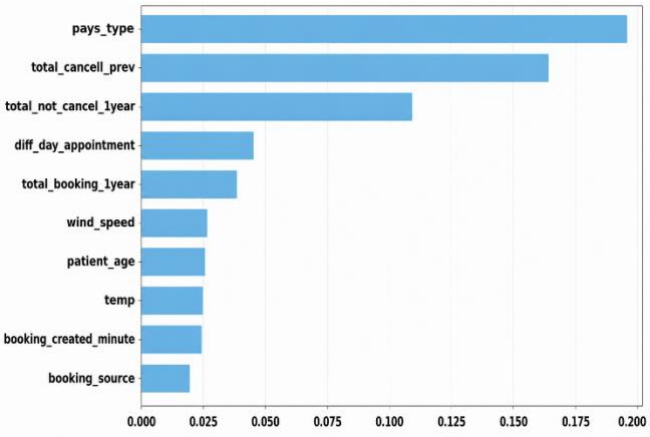

Fig. 6. Random Forest Scenario 23 Feature Importance

than standard data collected at the time of booking. Compared to the baseline models, which exhibited limited class separation, the behaviorally enriched model demonstrates a stronger capacity to distinguish between positive and negative outcomes. These findings confirm that in clinical scheduling environments, long-term patient history provides a more reliable foundation for prediction than variables restricted to a single transaction event.

D. Analysis of Environmental Variables

Following the validation of behavioral features, the influence of external environmental factors, specifically temperature, humidity, and wind speed was examined in Scenarios 13 to 18. Integrating these variables yielded only marginal performance gains; for instance, the baseline accuracy of the TE-based model (Scenario 5) increased slightly from 0.805 to 0.808 in Scenario 17. This finding suggests that while environmental conditions provide additional context, they do not serve as primary predictors for appointment outcomes in this specific clinical setting. The full integration phase (Scenarios 19 to 24) combined raw data, behavioral history, and weather variables, with Scenario 23 achieving a peak accuracy of 0.871 and a ROC-AUC of 0.940.

Despite the strong metrics observed in Scenario 23, a deeper examination of the feature importance ranking in Fig. 6, where the X-axis quantifies the relative Gini importance score for each variable identified a critical issue regarding model interpretability. While behavioral variables remained the most influential, the variable booking_created_minute appeared as the ninth most important feature. From both clinical and operational perspectives, the specific minute an appointment is booked is unlikely to have a causal relationship with patient attendance behavior. The prominence of this variable suggests that the model might be capturing fine-grained transactional noise that may fail to generalize.

To verify this concern, an interpretability audit was conducted using SHAP (SHapley Additive exPlanations), as illustrated in Fig. 7. The SHAP beeswarm plot visualizes the distribution of impact each feature has on the model output; the X-axis represents the SHAP value, where positive values (points to the right of the center line) increase the probability of cancellation. Furthermore, the color gradient reflects the feature value, where red indicates high values and blue indicates low values. In stark contrast to the Gini ranking, the SHAP audit reveals that *booking_created_minute* contributes minimally to the final prediction, appearing at the bottom of the hierarchy with a near-zero effect across the stratified samples. This inconsistency confirms that the high Gini ranking was an artifact of impurity-based bias toward high-cardinality noise.

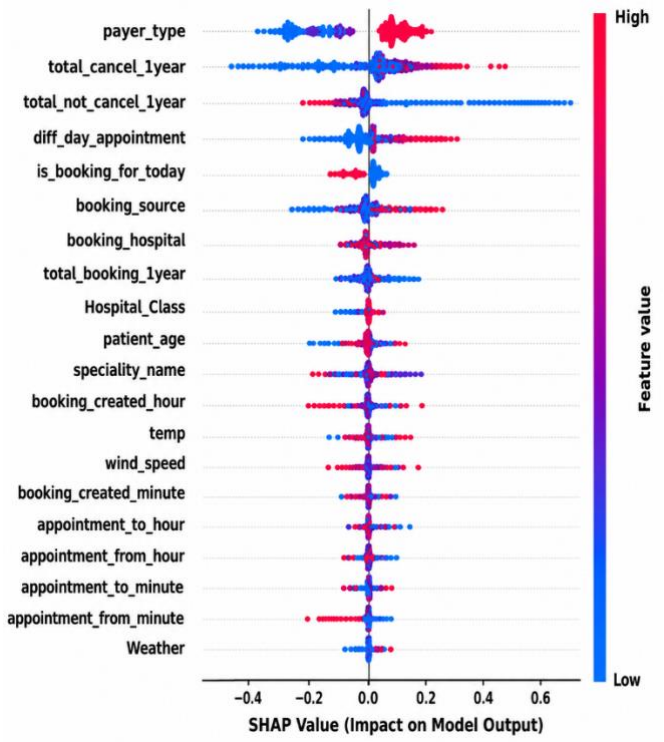

Fig. 7. SHAP beeswarm plot for global feature interpretation of Scenario 23.

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

E. Feature Selection and Proposed Method Performance Recovery

To enhance system reliability and mitigate the spurious correlations identified in the previous audit, a two-stage feature selection process was implemented. It involved an initial expert filtering to remove clinically irrelevant attributes (such as booking_minute), followed by an information-theoretic approach using Mutual Information (MI) to exclude variables with low predictive scores (< 0.05). In Scenario 25, this purification process led to a marginal reduction in Accuracy (from 0.871 to 0.868) and ROC-AUC (from 0.940 to 0.938) compared to Scenario 23. This performance dip confirms that the removal of noisy features, which previously acted as a 'statistical crutch', temporarily reduced the model's discriminatory power under standard hyperparameter configurations.

Table 4. Result of Final Hyperparameter Configuration

Hyperparameter	Scenario(23) RandomizedCV	Scenario(25) MI + RandomizedCV	Proposed Method(26) MI + PSO
n_estimators	500	459	122
max_depth	30	27	26
min_samples_split	2	5	17
min_samples_leaf	2	8	9

As shown on Table 4, the proposed PSO-based optimization in Scenario 26 successfully restored the Accuracy and ROC-AUC to their peak values of 0.871 and 0.940, respectively. The statistical significance of this recovery is supported by a tight 95% Confidence Interval (CI) calculated on the test partition (*N=133,621*) yielding [0.8692, 0.8728] for the F1-Score and [0.9387, 0.9413] for ROC-AUC. These narrow intervals confirm that the model's performance is statistically stable and not a result of stochastic variation. This recovery demonstrates that PSO can identify a more effective hyperparameter configuration that maximizes predictive signals from a refined, noise-free feature set.

A major operational advantage of this optimization is the substantial structural simplification of the model. While the RandomizedSearchCV approach (Scenario 23) required 500 estimators to achieve peak performance, the PSO-optimized model reached the same result using only 122 estimators. It represents a 75.6% reduction in tree complexity. Since memory footprint and computational overhead in Random Forest ensembles scale linearly with the number of trees, this reduction facilitates in serialized model size and CPU training requirements. Furthermore, PSO identified a more conservative min_samples_split value of 17, significantly higher than the value of 2 selected by RandomizedSearchCV. By converging on larger split and leaf boundaries (min_samples_leaf = 9), the PSO algorithm acts as an automated mathematical regularizer. This forces the trees to learn broader, group level behavioral patterns rather than overfitting to individual transactional anomalies. Consequently, the proposed framework successfully balances the bias variance trade-off, yielding a model that is not only computationally lean but also logically robust for deployment in high-volume hospital environments.

F. Overfitting Analysis

To verify that the performance of the PSO-optimized framework in Scenario 26 was driven by genuine patterns rather than overfitting, a rigorous generalization analysis was conducted. In large-scale healthcare informatics, a model is considered reliable if the performance gap between the training and testing sets remains within a conservative threshold, typically below 10%. Maintaining this stability is essential for clinical applications, ensuring the model can accurately process unseen patient data in real-world environments. As detailed in Table 5, the performance gap between the training set (Accuracy: 0.8990) and the testing set (Accuracy: 0.8707) is only 0.0283 (2.83%). A similar minimal variance was observed in the F1-score, with a gap of 0.0286 (2.86%). These results are significantly below the 10% industrial benchmark, confirming that the proposed framework is highly stable and resilient to overfitting. These findings highlight two critical factors that contributed to the model's robustness. First, the two-stage feature selection process using expert filtering and Mutual Information successfully eliminated spurious patterns, preventing the Random Forest from learning irrelevant data artifacts. Second, the conservative hyperparameter configuration identified by the PSO algorithm, particularly the *min_samples_split* value of 17, serves as a structural regularizer. It forces the model to capture broader, group-level behavioral trends rather than memorizing individual data points. Consequently, the framework demonstrates the consistent and predictable performance required for high-stakes deployment in hospital operational systems.

Table 5. Overfitting Analysis Result

Matrix	Training Score	Testing Score	Gap	Status
Accuracy	0.8990	0.8707	0.0283	Safe
F1-Score	0.8979	0.8693	0.0286	Safe

G. Robustness and Sensitivity Analysis

To address potential concerns regarding temporal dependency and patient-level data leakage—inherent challenges in clinical scheduling datasets a rigorous sensitivity analysis was performed on the proposed PSO-

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

Table 6. Robustness and Sensitivity Analysis of the Proposed Model

Validation Schema	Partition Method	Accuracy	F1 Score	ROC AUC	ΔF1 Score
Random Split (Baseline)	Stratified Random (80:20)	0.8707	0.8692	0.9400	Reference
Temporal Split (Test 1)	Chronological (Last 2 Months as Test Set)	0.8562	0.8551	0.9291	-0.0141
Patient Group K-Fold (Test 2)	Group Kfold by Contact Id (5 Folds)	0.8681	0.8669	0.9386	-0.0023

RF model. The framework was evaluated under two realistic deployment stress-tests: (1) Chronological Temporal Validation, where the model was trained on the first nine months of data and tested on the final two months to simulate prospective deployment; and (2) 5-Fold Patient-Grouped Cross-Validation (GroupKFold), where the split was strictly partitioned by patient identifier (Contact Id). It ensures that no individual patient's records appear in both the training and testing sets simultaneously, thereby preventing group-level leakage. The high stability observed in this per-patient validation confirms that the model effectively generalizes across diverse patient subgroups. Furthermore, since each fold implicitly encompasses varied insurance classes, age groups, and hospital locations, the minimal performance variance ensures that the predictive signal remains robust across heterogeneous populations.

As evidenced in Table 6 the model's F1-score experienced only a minor decrease of 1.41% (dropping from the 0.8692 baseline to 0.8551 in the chronological

Table 7. Comparison of the Proposed Framework with Existing Methods

Author	Study Setting	Sample Size	Method	Val Method	Computation-al Spec	Acc	ROC AUC	Training Time
Dashtban & Li [29]	Acute NHS Hospital, UK	3,747,285	SDAEs (Deep Learning)	Temporal Split (75:25)	CPU 4 Ghz, RAM 32GB, GPU NVIDIA GTX 1080Ti	Not Shown	0.704	Not Shown
Ocampo Osorio et al. [31]	Internal Medicine, Colombia	18,587	Bagging RF With SMOTE	5-Fold Cross-Validation	RAM 32GB, NVIDIA Quadro M2000 GPU	0.848	0.890	Not Shown
Yi Yang et al. [30]	MDHB Hospital, New Zealand	649,758	RF With Oversampling	10-Fold Cross-Validation	Not Shown	0.890	0.922	>137 Hours
Alshehri et al. [20]	Primary Care, Saudi Arabia	5,379,649	Random Forest	Holdout (70:30)	Not Shown	0.765	0.852	Not Shown
Deina et al. [21]	Hospitals in Brazil	16,088	Symbolic Regression + IHT	Double z-Fold CV (100 reps)	Not Shown	0.819	0.810	Not Shown
Proposed Method	Three Private Hospitals, Indonesia	686,387	PSO-RF With Behavior Features	Stratified 80:20 (GroupK Fold & Temporal Validation)	CPU Intel Xeon, RAM 12 GB Without GPU/TPU	0.871	0.940	< 60 minutes

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

split). Moreover, the patient-grouped validation maintained a high F1-score of 0.8669, representing a negligible difference of only 0.23% from the random split baseline. These results proved that the engineered behavioral features and the PSO-optimized structure prevent over-optimistic performance estimations and guarantee robust generalization. Crucially, the consistency of these metrics across all validation schemes validates the sampling strategy used during the optimization phase; it demonstrates that the 2% stratified training sample provided a statistically sufficient representation of the full dataset distribution to identify globally optimal hyperparameters.

IV. Discussion

A. The Impact of Behavioral Data and Model Optimization

The findings of this study confirm that predictive performance in healthcare scheduling is more influenced by feature quality than model architecture complexity. An increase in the ROC-AUC from 0.851 to 0.940 indicates that 12 months behavioral features provide more informative signals than raw operational data. This is because patient behavior in clinical settings is often habitual, where historical attendance and cancellation patterns more effectively reflect patient commitment than single-time booking records. Furthermore, behavioral features may partially capture underlying socioeconomic factors or institutional access barriers, consistent performance across multiple validation schemes suggests that these features serve as stable, high-level predictors of patient intentions. By integrating historical habits, the model captures complex, latent behavioral drivers without requiring sensitive or intrusive personal socioeconomic data, thus preserving patient privacy while enhancing predictive reliability.

While methods such as that of Ocampo Osorio et al. [31] emphasizing synthetic augmentation through SMOTE to address class imbalance, the results of this study indicate that in our specific experimental setting, increasing sensitivity through synthetic generation can increase the potential for false positives. We acknowledge that SMOTE performance is highly sensitive to environmental settings and hyperparameter optimization; however, for this large-scale dataset, replacing synthetic resampling with Target Encoding (TE) and behavioral features successfully maintained the natural decision boundary in clinical data. This interpretation is supported by the PSO-based configuration, which favors a more general model structure by selecting a higher min_samples_leaf value of 17. Consequently, the system remains stable and reliable when applied to a hospital setting, reducing the risk of overfitting on noisy or artificial data points.

B. Comparative Context with Previous Research

To provide a broader context for the proposed framework's performance, we compared our results with five representative studies as summarized in the following Table 7. While direct head-to-head comparisons are inherently limited by the heterogeneity of geographic settings and class distributions, these benchmarks highlight the technical and operational advantages of our PSO-RF model. A critical finding in this study is the "*Imbalance Paradox*" which was observed when comparing our results with large-scale national studies. For instance, Alshehri et al. [20] utilized a massive dataset of 5.3 million records from Saudi Arabia, which benefited from a nearly balanced distribution (55.07% no-show). Generally, balanced datasets present a simpler classification task. However, Alshehri's best model only achieved an accuracy of 0.765 and ROC-AUC of 0.852. In contrast, our model navigated a significantly more challenging and extreme imbalance (only 29.07% cancellation) yet achieved a substantially higher accuracy of 0.871 and ROC-AUC of 0.940. This disparity proves that Behavioral Feature Engineering captures a much higher signal-to-noise ratio than the demographic and weather features used in Alshehri's work.

Furthermore, our approach maintains higher data integrity compared to studies that rely on artificial data manipulation. Ocampo Osorio et al. [31] applied Bagging RF combined with SMOTE on a small dataset of 18,587 records, achieving an accuracy of 0.848. Similarly, Deina et al. [21] utilized Symbolic Regression and Instance Hardness Threshold (IHT) resampling on a dataset of 16,088 records, reaching an accuracy of 0.819. Our framework demonstrates superior scalability; by leveraging a 12-month historical trace as a natural data enrichment strategy, we achieved a higher accuracy of 0.871 on a dataset of 686,387 records which is more than 40 times larger than the datasets used by Ocampo and Deina without the artificial noise of SMOTE or the information loss inherent in IHT undersampling.

Computational sustainability also represents a major advantage of the proposed framework. While the high-performing model by Yi Yang et al. [30] required an impractical training time of over 137 hours, our system completed the entire pipeline from data preprocessing to multi-scheme validation in under 60 minutes using standard CPU resources. This efficiency is driven by the PSO algorithm's ability to identify an optimal, low-complexity model structure (reducing tree count by 75.6%) on a stratified 2% sample. These comparative results indicate that even when compared to more recent and complex architectures such as Deep Learning [29] or Symbolic Regression [21], an optimized Random Forest remains a superior choice for clinical Decision Support Systems. By prioritizing feature quality and metaheuristic tuning over model depth, our framework provides a realistic, lightweight, and cost-effective solution for Hospital AI Command Center (AICC) environments.

The comparative results presented in Table 7 are intended to provide a contextual benchmark rather than a direct claim of superiority. Given the inherent heterogeneity in clinical settings, geographic locations, and computing infrastructure across different studies, the proposed framework demonstrates its value as a resource-efficient alternative that balances high predictive

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

performance with practical operational feasibility in a private hospital setting.

C. Practical and Corporate Implication

The findings of this study provide direct actionable insights for hospital management and operational efficiency. From an operational perspective, the high precision of the proposed model significantly mitigates the risk of 'alert fatigue' among administrative staff. This is achieved by minimizing False Positives (FP) instances where a patient is incorrectly flagged for cancellation but actually intends to attend. High FP rates would cause staff to waste valuable clinical hours on unnecessary outbound confirmation calls, potentially leading to patient annoyance. By relying on authentic behavioral history, the proposed model ensures that front-office interventions are concentrated on truly high-risk patients, thereby optimizing workforce productivity. This targeted approach supports a proactive Customer Relationship Management (CRM) strategy, enabling a transition from reactive handling to proactive slot management [36][37].

From a corporate and financial standpoint, the framework addresses the critical issue of False Negatives (FN) instances where the model fails to identify a patient who will eventually cancel. Operationally, FNs lead to *'perishable capacity loss'*, where appointment slots remain unused and clinical personnel stay idle, resulting in direct revenue loss that cannot be recovered. The proposed PSO-RF model, which maintains a robust F1-score of 0.870, strikes an effective balance between administrative efficiency and financial protection. Furthermore, the primary corporate advantage lies in the model's retraining efficiency. While theexisting high-accuracy models may require over 137 hours for retraining [30], our framework completed this cycle in under 60 minutes. This rapid turnaround ensures system relevance without the need for expensive high-performance computing (HPC) infrastructure or excessive energy consumption. Furthermore, the model's lightweight architecture is compatible with standard enterprise servers, making it a fiscally sustainable solution for future Hospital AI Command Center (AICC) and Digital Twin platforms.

In practical clinical deployments within an AICC environment, a rigid default decision threshold (e.g., 0.5) often fails to account for fluctuating daily staffing levels or clinic demands. Since our Random Forest model outputs continuous probability scores, hospital administrators can dynamically adjust the decision threshold to balance the trade-off between Precision and Recall. For instance, in high-demand outpatient clinics where vacant slots represent a significant loss of "perishable" capacity, the AICC can lower the threshold (e.g., to 0.3) to aggressively identify potential cancellations for double-booking. Conversely, during periods of limited administrative capacity, the threshold can be raised (e.g., to 0.7) to trigger alerts only for patients with a very high probability of non-attendance, ensuring that outbound calls remain resource-efficient. To support this flexibility, the model achieves a high Precision-Recall Area Under the Curve (PR-AUC) of 0.8538, confirming that its predictive stability is maintained across various operational decision boundaries.

D. Conceptual Integration and CRM Workflow

To guide the potential implementation of the proposed model, this study presents a conceptual enterprise data architecture and Customer Relationship Management (CRM) workflow. Built on the Microsoft Fabric platform using a Medallion Architecture, raw transactional and scheduling events from patient-facing mobile applications are planned to be ingested into the Bronze Layer. These records would then be refined and structurally organized within the Silver Layer through automated ETL snapshots. In the Gold Layer, these datasets were combined with historical behavioral features and local weather data to form an AI-ready feature store. The optimized PSO-RF model would function as a stateless predictive engine within this environment, evaluating each incoming booking to generate cancellation probability scores in real-time. These predictive insights were designed to be integrated into PowerBI dashboards for clinical monitoring and a Fabric IQ semantic ontology. By leveraging Copilot, front-office operators could interact with these insights via natural language queries, enabling immediate, AI-assisted decision support for daily scheduling adjustments. This integration transforms mathematical predictions into accessible intelligence for hospital administrative staff.

Rather than relying on a fully automated system, the proposed workflow adopts a “Human-in-the-Loop” approach. Hospital administrators would retain the authority to adjust decision thresholds dynamically via the PowerBI interface based on real-time clinic demands and regional demographics. Once a high-risk alert is triggered, the CRM strategy is segmented to optimize resource allocation. For younger, digitally-literate patients, the system would trigger automated push notifications or WhatsApp reminders at critical touchpoints, such as three days or one day prior to the appointment. Conversely, for elderly patients or those with lower digital access, the AICC operator would initiate manual outbound confirmation calls prompted by the system. This demographic-sensitive CRM strategy aims to ensure optimal capacity recovery while maximizing institutional efficiency and respecting individual patient communication preferences.

E. Research Limitations

Despite the high predictive performance achieved, this study acknowledges several limitations that define its scope and generalizability. First, the dataset was sourced exclusively from three private hospitals belonging to the same healthcare group in Indonesia. Consequently, the findings may reflect specific institutional cultures and patient demographics that might not be entirely representative of public healthcare facilities or different geographic regions. Future research is required to validate the model's performance in more diverse clinical settings. Second, the framework is subject to the 'cold-start problem'. Since the predictive power is heavily driven by 12-month behavioral features, the model’s efficacy may decrease for first-time patients who lack a

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

transactional history. While a zero-baseline imputation was applied, this subgroup represents a challenge for behavior-based prediction engines. Furthermore, the reliance on retrospective transaction data may introduce latent biases, as it does not account for external factors such as patient socio-economic status, transportation barriers, or personal emergencies that are not captured in the hospital information system. Third, in terms of methodological scope, this research focused primarily on optimizing Random Forest to satisfy the clinical requirement for model interpretability. While we compared PSO-RF against several baseline models, a more exhaustive comparison with advanced boosting algorithms, such as XGBoost or LightGBM, was not conducted. Future studies should investigate whether these high-performance models offer marginal accuracy gains that justify the potential loss in interpretability.

Finally, the current validation strategy including chronological temporal splits remains retrospective in nature. Although it effectively simulates prospective deployment, the study lacks real-world implementation testing in a live clinical workflow. Practical challenges, such as system latency, institutional approval, and human-resource readiness within a live AI Command Center (AICC) environment, remain to be explored. Establishing pilot trials for real-time operational evaluation is therefore a critical next step to further validate the framework's effectiveness in reducing actual no-show rates.

V. Conclusion

This study aimed to develop an optimized, lightweight predictive framework for outpatient appointment cancellations by integrating behavioral feature engineering and Particle Swarm Optimization (PSO). The results demonstrate that behavioral feature engineering is the most significant driver of performance, increasing the baseline accuracy from 0.795 (Scenario 1) to 0.871 (Scenario 26). Crucially, the proposed framework proved effective in performance recovery, while the removal of clinically irrelevant features (noise filtering) initially caused a minor performance dip, the PSO-based optimization successfully restored metrics to their peak values of 0.871 accuracy and 0.940 ROC-AUC. This recovery was achieved while maintaining higher data integrity compared to synthetic resampling (SMOTE) and surpassing recent benchmarks. Furthermore, the PSO algorithm successfully identified an optimal model structure, achieving a 75.6% reduction in tree complexity (requiring only 122 estimators instead of 500) without sacrificing accuracy. The entire pipeline is computationally sustainable, completing in under 60 minutes using standard CPU resources. This makes the system a viable core engine for Hospital AI Command Centers (AICC) and Digital Twin architectures. Robustness analysis confirmed high stability, with a maximum performance degradation of only 1.41% under strict temporal validation. Future work will focus on addressing the cold-start problem for new patients such as incorporating socio-economic and geolocation data. Additionally, subsequent research should include prospective validation in streaming environments, testing across public hospital settings, and the integration of real-time public transportation data to further refine cold-start patient modeling.

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial, or not for profit sectors.

Conflict of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions: Azmi Abiyyu Dzaky was responsible for conceptualization, data acquisition, preprocessing, machine learning development, and experiment implementation. Farrikh Alzami contributed to methodology validation, result interpretation, and manuscript revision. Pujiono provided research supervision, and methodological guidance.

Data Availability: The dataset used in this study is proprietary and not publicly accessible due to institutional privacy policies. De-identified data may be made available from the corresponding author upon reasonable request and subject to institutional approval.

Code Availability: The source code used for model training and optimization is available from the corresponding author upon reasonable request.

Ethical Approval: This study utilized retrospective, de-identified secondary operational data. No direct interaction with human participants was conducted, and all personally identifiable information was removed at the source. Formal ethical approval is not required for this type of study.

Informed Consent: Informed consent is not required as the research used retrospective de-identified data and involved no direct patient participation.

Consent for Publication: Not applicable.

Acknowledgment: The authors extend their sincere gratitude to the participating hospital institutions in Indonesia for providing access to the retrospective outpatient appointment dataset used in this study.

References

[1] M. Yu, G. Zhao, and D. Tang, “The relationship between internal and external factors about the outpatients’ choice of hospital: A cross-sectional study from Jiaxing City, China,” *Health Science Reports*, vol. 5, no. 5, Sep. 2022, doi: 10.1002/hsr2.821.

[2] R. Mbau *et al.*, “Analysing the Efficiency of Health Systems: A Systematic Review of the Literature,” *Applied Health Economics and Health Policy*, vol. 21, no. 2, pp. 205–224, Mar. 2023, doi: 10.1007/s40258-022-00785-2.

[3] A. Ala and F. Chen, “Appointment Scheduling

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

Problem in Complexity Systems of the Healthcare Services: A Comprehensive Review," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–16, Mar. 2022, doi: 10.1155/2022/5819813.

[4] I. Rajković, D. Baričević, and M. Šimunić, "Overview Of Different It Solutions For Reducing No-Shows And Unexplained Patient Cancellations In Health Tourism," 2024, pp. 167–175. doi: 10.20867/thi.27.21.

[5] A. R. Shour, G. L. Jones, R. Anguzu, S. A. Doi, and A. A. Onitilo, "Development of an evidence-based model for predicting patient, provider, and appointment factors that influence no-shows in a rural healthcare system," *BMC Health Services Research*, vol. 23, no. 1, p. 989, Sep. 2023, doi: 10.1186/s12913-023-09969-5.

[6] M. Alturbag, "Factors and Reasons Associated With Appointment Non-attendance in Hospitals: A Narrative Review," *Cureus*, vol. 16, no. 4, Apr. 2024, doi: 10.7759/cureus.58594.

[7] K. Toker, K. Ataş, A. Mayadağlı, Z. Görmezoğlu, I. Tuncay, and R. Kazancıoğlu, "A Solution to Reduce the Impact of Patients' No-Show Behavior on Hospital Operating Costs: Artificial Intelligence-Based Appointment System," *Healthcare*, vol. 12, no. 21, p. 2161, Oct. 2024, doi: 10.3390/healthcare12212161.

[8] G. Leeftink, G. Martinez, E. W. Hans, M. Y. Sir, and K. S. Pasupathy, "Optimising the booking horizon in healthcare clinics considering no-shows and cancellations," *International Journal of Production Research*, vol. 60, no. 10, pp. 3201–3218, May 2022, doi: 10.1080/00207543.2021.1913292.

[9] D. Schwalbe, M. Sodemann, M. Iachina, B. M. Nørgård, N. H. Chodkiewicz, and J. Ammentorp, "Causes of Patient Nonattendance at Medical Appointments: Protocol for a Mixed Methods Study," *JMIR Research Protocols*, vol. 12, no. 1, p. e46227, Nov. 2023, doi: 10.2196/46227.

[10] J. Lopes, T. Miranda, R. Sousa, and J. Machado, "Primary Health Care Appointments and Hospital Stay: An Impact Analysis," *Procedia Computer Science*, vol. 251, no. 2019, pp. 696–702, 2024, doi: 10.1016/j.procs.2024.11.171.

[11] M. R. Mazaheri Habibi *et al.*, "Evaluation of no-show rate in outpatient clinics with open access scheduling system: A systematic review," *Health Science Reports*, vol. 7, no. 7, Jul. 2024, doi: 10.1002/hsr2.2160.

[12] L. J. Basile, N. Carbonara, R. Pellegrino, and U. Panniello, "Business intelligence in the healthcare industry: The utilization of a data-driven approach to support clinical decision making," *Technovation*, vol. 120, no. February 2022, p. 102482, Feb. 2023, doi: 10.1016/j.technovation.2022.102482.

[13] R. T. Sutton, K. D. Chappell, D. Pincock, D. Sadowski, D. C. Baumgart, and K. I. Kroeker, "The Effect of an Electronic Medical Record–Based Clinical Decision Support System on Adherence to Clinical Protocols in Inflammatory Bowel Disease Care: Interrupted Time Series Study," *JMIR Medical Informatics*, vol. 12, pp. e55314–e55314, Mar. 2024, doi: 10.2196/55314.

[14] N. Wiwatkunupakarn *et al.*, "The Integration of Clinical Decision Support Systems Into Telemedicine for Patients With Multimorbidity in Primary Care Settings: Scoping Review," *Journal of Medical Internet Research*, vol. 25, p. e45944, Jun. 2023, doi: 10.2196/45944.

[15] K. M. Toffaha, M. C. E. Simsekler, A. Alshehhi, and M. Atif Omar, "Predicting Hospital No-Shows: Interpretable Machine Learning Models Approach," *IEEE Access*, vol. 12, no. September, pp. 166058–166067, 2024, doi: 10.1109/ACCESS.2024.3490662.

[16] A. Kuiper, M. Mandjes, J. de Mast, and R. Brokkelkamp, "A flexible and optimal approach for appointment scheduling in healthcare," *Decision Sciences*, vol. 54, no. 1, pp. 85–100, Feb. 2023, doi: 10.1111/deci.12517.

[17] H. Feng, Y. Jia, T. Huang, S. Zhou, and H. Chen, "An adaptive decision support system for outpatient appointment scheduling with heterogeneous service times," *Scientific Reports*, vol. 14, no. 1, p. 27731, Nov. 2024, doi: 10.1038/s41598-024-77873-x.

[18] R. Ali, A. Hussain, S. Nazir, S. Khan, and H. U. Khan, "Intelligent Decision Support Systems—An Analysis of Machine Learning and Multicriteria Decision-Making Methods," *Applied Sciences*, vol. 13, no. 22, p. 12426, Nov. 2023, doi: 10.3390/app132212426.

[19] R. Tong, Z. Zhu, and J. Ling, "Comparison of linear and non-linear machine learning models for time-dependent readmission or mortality prediction among hospitalized heart failure patients," *Heliyon*, vol. 9, no. 5, p. e16068, May 2023, doi: 10.1016/j.heliyon.2023.e16068.

[20] A. Alshehri *et al.*, "Machine Learning Approaches to Predict No-Shows in Saudi Arabian Primary and General Healthcare Settings," *Saudi Journal of Health Systems Research*, vol. 5, no. 1, pp. 1–17, Nov. 2024, doi: 10.1159/000542701.

[21] C. Deina, F. S. Fogliatto, G. J. C. da Silveira, and M. J. Anzanello, "Decision analysis framework for predicting no-shows to appointments using machine learning algorithms," *BMC Health Services Research*, vol. 24, no. 1, p. 37, Jan. 2024, doi: 10.1186/s12913-023-10418-6.

[22] A. Soeleman, F. Al Zami, and D. I. Krisnawati, "Comparison Of Arrival Classification Of Outpatient Patients Based On Appointment Using Adaboost And Random Undersampling Methods," *Jurnal Keuangan Dan Perbankan*, vol. 11, no. 1, pp. 28–40, 2023, doi: 10.32832/moneter.v11i1.56.

[23] F. Alzami, Abdussalam, R. A. Megantara, A. Z. Fanani, and Purwanto, "Diabetic Retinopathy

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

Grade Classification based on Fractal Analysis and Random Forest," in *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, IEEE, Sep. 2019, pp. 272–276. doi: 10.1109/ISEMANTIC.2019.8884217.

[24] A. Vafaeinejad, A. Sharifi, and S. N. Khan, "Robust County-Level Corn Yield Estimation Using Ensemble Machine Learning and Multisource Remote Sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 18, pp. 16942–16953, 2025, doi: 10.1109/JSTARS.2025.3585779.

[25] S. Han, B. D. Williamson, and Y. Fong, "Improving random forest predictions in small datasets from two-phase sampling designs," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, p. 322, Dec. 2021, doi: 10.1186/s12911-021-01688-3.

[26] A. Nurhindarto, E. W. Andriansyah, F. Alzami, P. Purwanto, M. A. Soeleman, and D. P. Prabowo, "Employee Attrition and Performance Prediction using Univariate ROC feature selection and Random Forest," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, vol. 4, Nov. 2021, doi: 10.22219/kinetik.v6i4.1345.

[27] A. Jarmakovica, "Machine learning-based strategies for improving healthcare data quality: an evaluation of accuracy, completeness, and reusability," *Frontiers in Artificial Intelligence*, vol. 8, no. July, pp. 1–14, Jul. 2025, doi: 10.3389/frai.2025.1621514.

[28] M.M Jibril *et al.*, "An overview of streamflow prediction using random forest algorithm," *GSC Advanced Research and Reviews*, vol. 13, no. 1, pp. 050–057, Oct. 2022, doi: 10.30574/gscarr.2022.13.1.0112.

[29] M. Dashtban and W. Li, "Predicting non-attendance in hospital outpatient appointments using deep learning approach," *Health Systems*, vol. 11, no. 3, pp. 189–210, Jul. 2022, doi: 10.1080/20476965.2021.1924085.

[30] Y. Yang, S. Madanian, and D. Parry, "Enhancing Health Equity by Predicting Missed Appointments in Health Care: Machine Learning Study," *JMIR Medical Informatics*, vol. 12, p. e48273, Jan. 2024, doi: 10.2196/48273.

[31] F. Ocampo Osorio *et al.*, "Predicting non-shows at outpatient appointments in internal medicine using machine learning models," *PeerJ Computer Science*, vol. 11, p. e2762, Apr. 2025, doi: 10.7717/peerj-cs.2762.

[32] A. Singh, N. Prakash, and A. Jain, "Particle Swarm Optimization-Based Random Forest Framework for the Classification of Chronic Diseases," *IEEE Access*, vol. 11, no. November, pp. 133931–133946, 2023, doi: 10.1109/ACCESS.2023.3335314.

[33] B. M. Porto and F. S. Fogliatto, "Enhanced forecasting of emergency department patient arrivals using feature engineering approach and machine learning," *BMC Medical Informatics and Decision Making*, vol. 24, no. 1, p. 377, Dec. 2024, doi: 10.1186/s12911-024-02788-6.

[34] A. Vallée, "Digital twin for healthcare systems," *Frontiers in Digital Health*, vol. 5, no. September, pp. 1–6, Sep. 2023, doi: 10.3389/fdgth.2023.1253050.

[35] J. R. McColl-Kennedy *et al.*, "Digital twins: a game changer in customer experience," *Journal of Service Management*, pp. 1–31, Dec. 2025, doi: 10.1108/JOSM-12-2024-0540.

[36] V. R. BOPPANA, "Impact of Dynamics CRM Integration on Healthcare Operational Efficiency," *EPH-International Journal of Business & Management Science*, vol. 8, no. 1, pp. 1–9, 2022, doi: 10.53555/eijbms.v8i1.179.

[37] Ashwin Vijaykumar Bajoria, "Cognitive Companion CRM: Proactive Intelligence for Personalized and Efficient Healthcare," *Journal of Computer Science and Technology Studies*, vol. 7, no. 5, pp. 727–736, Jun. 2025, doi: 10.32996/jcsts.2025.7.5.81.

[38] A.-M. Ștefan, N.-R. Rusu, E. Ovreiu, and M. Ciuc, "Empowering Healthcare: A Comprehensive Guide to Implementing a Robust Medical Information System—Components, Benefits, Objectives, Evaluation Criteria, and Seamless Deployment Strategies," *Applied System Innovation*, vol. 7, no. 3, p. 51, Jun. 2024, doi: 10.3390/asi7030051.

[39] Onyekachukwu Victor Unanah and Olu James Mbanugo, "Integration of AI into CRM for Effective U.S. healthcare and pharmaceutical marketing," *World Journal of Advanced Research and Reviews*, vol. 25, no. 2, pp. 609–630, Feb. 2025, doi: 10.30574/wjarr.2025.25.2.0396.

[40] M. Açıkkar and Y. Altunkol, "A novel hybrid PSO- and GS-based hyperparameter optimization algorithm for support vector regression," *Neural Computing and Applications*, vol. 35, no. 27, pp. 19961–19977, Sep. 2023, doi: 10.1007/s00521-023-08805-5.

[41] C. Lartey, J. Liu, R. K. Asamoah, C. Greet, M. Zanin, and W. Skinner, "Effective Outlier Detection for Ensuring Data Quality in Flotation Data Modelling Using Machine Learning (ML) Algorithms," *Minerals*, vol. 14, no. 9, p. 925, Sep. 2024, doi: 10.3390/min14090925.

[42] A. Y. Hussein, P. Falcarin, and A. T. Sadiq, "Enhancement performance of random forest algorithm via one hot encoding for IoT IDS," *Periodicals of Engineering and Natural Sciences (PEN)*, vol. 9, no. 3, pp. 579–591, Aug. 2021, doi: 10.21533/pen.v9.i3.872.

[43] F. Pargent, F. Pfisterer, J. Thomas, and B. Bischl, "Regularized target encoding outperforms

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

traditional methods in supervised machine learning with high cardinality features," *Computational Statistics*, vol. 37, no. 5, pp. 2671–2692, Nov. 2022, doi: 10.1007/s00180-022-01207-6.

[44] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5, p. 91, Apr. 2023, doi: 10.3390/computers12050091.

[45] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, Jun. 2024, doi: 10.58496/BJML/2024/007.

[46] M. Mohammed and R. Alsunosi, "Effect of Selecting Validation Dataset on Building Random Forest and Decision Tree Models," *Zenodo*, vol. 5, no. 2, pp. 470–478, 2022, doi: 10.5281/zenodo.7113928.

[47] S. Bishnoi and B. K. Hooda, "Decision Tree Algorithms and their Applicability in Agriculture for Classification," *Journal of Experimental Agriculture International*, vol. 44, no. 7, pp. 20–27, May 2022, doi: 10.9734/jeai/2022/v44i730833.

[48] E. Scornet, "Trees, forests, and impurity-based variable importance in regression," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 59, no. 1, pp. 21–52, Feb. 2023, doi: 10.1214/21-AIHP1240.

[49] K. Gajowniczek and T. Ząbkowski, "ImbTreeEntropy and ImbTreeAUC: Novel R Packages for Decision Tree Learning on the Imbalanced Datasets," *Electronics*, vol. 10, no. 6, p. 657, Mar. 2021, doi: 10.3390/electronics10060657.

[50] M. I. Prasetiyowati, N. U. Maulidevi, and K. Surendro, "Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest," *Journal of Big Data*, vol. 8, no. 1, p. 84, Dec. 2021, doi: 10.1186/s40537-021-00472-4.

[51] M. A. Bouke, A. Abdullah, J. Frnda, K. Cengiz, and B. Salah, "BukaGini: A Stability-Aware Gini Index Feature Selection Algorithm for Robust Model Performance," *IEEE Access*, vol. 11, no. June, pp. 59386–59396, 2023, doi: 10.1109/ACCESS.2023.3284975.

[52] J. Joseph, S. Senith, A. A. Kirubaraj, and S. R. J. Ramson, "Machine Learning for Prediction of Clinical Appointment No-Shows," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 7, no. 4, pp. 558–574, Jul. 2022, doi: 10.33889/IJMEMS.2022.7.4.036.

[53] N. N. Amalina and H. An, "A Multi-Head Attention Soft Random Forest for Interpretable Patient No-Show Prediction," *Systems*, vol. 14, no. 5, p. 576, May 2026, doi: 10.3390/systems14050576.

[54] P. Bharathi, M. Ramachandran, K. Ramu, and S. Chinnasamy, "A Study on Various Particle Swarm Optimization Techniques used in Current Scenario," *Design, Modelling and Fabrication of Advanced Robots*, vol. 1, no. 1, pp. 15–26, Feb. 2022, doi: 10.46632/dmfar/1/1/3.

[55] E. Twumasi, E. A. Frimpong, N. K. Prah, and D. B. Gyasi, "A novel improvement of particle swarm optimization using an improved velocity update function based on local best murmuration particle," *Journal of Electrical Systems and Information Technology*, vol. 11, no. 1, p. 42, Oct. 2024, doi: 10.1186/s43067-024-00168-8.

[56] Z. Liang, "Efficient Representations for High-Cardinality Categorical Variables in Machine Learning," in *2025 International Conference on Advanced Machine Learning and Data Science (AMLDS)*, IEEE, Jul. 2025, pp. 1–11. doi: 10.1109/AMLDS63918.2025.11159409.

[57] D. Breskuvienė and G. Dzemyda, "Categorical Feature Encoding Techniques for Improved Classifier Performance when Dealing with Imbalanced Data of Fraudulent Transactions," *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL*, vol. 18, no. 3, pp. 1–17, May 2023, doi: 10.15837/ijccc.2023.3.5433.

[58] M. Alduailij, Q. W. Khan, M. Tahir, M. Sardaraz, M. Alduailij, and F. Malik, "Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method," *Symmetry*, vol. 14, no. 6, p. 1095, May 2022, doi: 10.3390/sym14061095.

[59] M. Alalhareth and S. Hong, "An Improved Mutual Information Feature Selection Technique for Intrusion Detection Systems in the Internet of Medical Things," *Sensors*, vol. 23, no. 10, p. 4971, May 2023, doi: 10.3390/s23104971.

[60] S. Farhadpour, T. A. Warner, and A. E. Maxwell, "Selecting and Interpreting Multiclass Loss and Accuracy Assessment Metrics for Classifications with Class Imbalance: Guidance and Best Practices," *Remote Sensing*, vol. 16, no. 3, p. 533, Jan. 2024, doi: 10.3390/rs16030533.

[61] A. Kumaravel and T. Vijayan, "Comparing cost sensitive classifiers by the false-positive to false-negative ratio in diagnostic studies," *Expert Systems with Applications*, vol. 227, no. May, p. 120303, Oct. 2023, doi: 10.1016/j.eswa.2023.120303.

[62] J. Ircio, A. Lojo, U. Mori, S. Malinowski, and J. A. Lozano, "Minimum Recall-Based Loss Function for Imbalanced Time Series Classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 10, pp. 10024–10034, Oct. 2023, doi: 10.1109/TKDE.2023.3268994.

[63] D. Fourure, M. U. Javaid, N. Posocco, and S. Tihon, "Anomaly Detection: How to Artificially Increase Your F1-Score with a Biased Evaluation Protocol," 2021, pp. 3–18. doi: 10.1007/978-3-030-86514-6_1.

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348

AUTHOR BIOGRAPHY

Azmi Abiyyu Dzaky received the B.S. degree in Computer Science from Universitas Dian Nuswantoro (UDINUS). He is currently pursuing a Master's degree in Computer Science at Dian Nuswantoro University. Professionally, he works as a Data Analyst at a hospital in Indonesia, where he focuses on business and clinical data analysis, as well as developing healthcare research platforms. His current research interests include machine learning applications for clinical decision support systems, healthcare information systems, and data driven hospital operational management.

Farrikh Alzami completed his Doctoral degree (Ph.D.) from South China University of Technology (SCUT), Guangzhou, China. He is currently an Associate Professor in Feature Extraction and Engineering at the Faculty of Computer Science, Universitas Dian Nuswantoro (UDINUS), Semarang, Indonesia. His primary research focuses on feature extraction methods using both deep learning and traditional techniques, with broad applications in Machine Learning, Data Mining, and Computer Vision. He has actively published numerous scientific articles in reputable national and international journals.

Pujiono completed his Doctoral degree (Dr.) from Institut Teknologi Sepuluh Nopember (ITS), Surabaya. He currently serves as a senior lecturer, researcher, and holds the structural position of Vice Dean for Academic and Student Affairs at the Faculty of Computer Science, Universitas Dian Nuswantoro (UDINUS), Semarang, Indonesia. Alongside his academic role, he holds professional engineering credentials as an Insinyur Profesional Madya (IPM) and an ASEAN Engineer (ASEAN Eng.). His primary research interests lie in Computer Science, Computational Mathematics, and Image Processing. He has actively published numerous scientific papers across reputable national and international journal

Corresponding author: Azmi Abiyyu Dzaky, azmiabiyyudzaky456@gmail.com, Department Informatics and Engineering, Faculty of Science, Universitas Dian Nuswantoro, Jl. Imam Bonjol No.207, 50131 Semarang, Jawa Tengah, Indonesia.
DOI: https://doi.org/10.35882/ijeeemi.v8i3.348