Volume 2 Issue 1 | March 2026
e-ISSN : Applied for (Online)
Journal of Integrated Engineering
Innovation and Applications

A Comparative Study of Machine Learning Algorithms for Multi-Disease Healthcare Prediction: A Web-Based Intelligent System

Madhuri Gokhale
Department of Computer Science and Engineering,
Jabalpur Engineering College
Jabalpur 482011, Madhya Pradesh, India
Email: mgokhale@jecjabalpur.ac.in

Vidhi Jain
Computer Science Department,
Govind Ballabh Pant Institute of Engineering and
Technology, Pauri Garhwal, Uttarakhand 246194, India
Email: sahilpatel94071@gmail.com

Swati Pathak
Department of Computer Science and Engineering,
Jabalpur Engineering College
Jabalpur 482011, Madhya Pradesh, India

Tanya Shrivastava
Department of Computer Science and Engineering,
Jabalpur Engineering College
Jabalpur 482011, Madhya Pradesh, India

ABSTRACT

Machine learning has significantly transformed healthcare analytics, enabling new approaches to early disease detection and clinical decision support. This study presents a web-based platform designed to predict five clinically significant diseases simultaneously: Diabetes, Heart Disease, Lung Cancer, Parkinson's Disease, and Thyroid Disorders. Disease-specific datasets were sourced from established public repositories and subjected to a systematic preprocessing pipeline encompassing noise removal, normalisation, feature selection, and stratified train-test partitioning.

Five supervised machine learning models were confusion matrices and multi-model accuracy comparisons. The platform is designed for scalability, with provisions for integration with electronic health records and wearable health monitoring devices, establishing its suitability as a next-generation clinical decision support tool. The system is intended to function as a decision support aid and does not substitute for professional clinical diagnosis.

***Index Terms**:— Machine Learning; Healthcare Prediction System; Random Forest; Support Vector Machine; Streamlit; Diabetes; Heart Disease; Lung Cancer; Parkinson's Disease; Thyroid Disorders; Clinical Decision Support*

evaluated: Logistic Regression, Random Forest, Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Decision Tree. Random Forest demonstrated superior performance, particularly on larger datasets. The highest predictive accuracy was achieved for Thyroid Disorder prediction at 99.6%, consistent with prior literature, attributable to the well-defined hormonal signatures characteristic of this condition.

The system is deployed on the Streamlit framework, providing an interactive browser-based interface with real-time predictions. A secure authentication module and an administrative panel for model management are also incorporated. Per-disease performance is examined in detail through

I. INTRODUCTION

Healthcare systems worldwide face persistent systemic challenges, including delayed disease detection, inequitable access to specialist care, diagnostic uncertainty, and escalating costs associated with conventional diagnostic procedures. These challenges are particularly pronounced in developing nations, where the scarcity of trained medical professionals and limited rural healthcare infrastructure result in infrequent specialist consultations. Delayed or inaccurate diagnoses carry significant clinical consequences: chronic conditions such as diabetes and thyroid disorders frequently remain asymptomatic in their early stages, while life-threatening conditions such as heart disease and lung cancer are often identified only at advanced stages, at which point therapeutic efficacy is substantially diminished and patient survival rates decline markedly.

Now, with digital health records everywhere and lots of curated, public medical data available, healthcare has fresh opportunities to lean on data- driven approaches. Machine learning (ML) stands out here. Instead of following strict, pre-set rules, ML learns tricky patterns from real patient histories on its own. It goes beyond traditional systems and handles tricky tasks like multi-parameter disease prediction by actually generalising from the data, not just memorizing what it's been told.

This study presents an intelligent web-based prediction system providing real-time risk assessments for five major diseases: diabetes, heart disease, lung cancer, Parkinson's disease, and thyroid disorders. Five supervised classification algorithms were evaluated on established, disease-specific benchmark datasets. The system is deployed on Streamlit, providing an accessible interface suitable for both technical and non-technical users. An administrative module enables authorised users to retrain models, review performance metrics, and manage deployment without developer intervention—a capability largely absent from prior work in this domain.

The principal contributions of this work are as follows. First, the system constitutes a unified online platform for multi-disease prediction across five clinically distinct conditions. Second, a rigorous head-to-head comparison of five machine learning models is conducted across five diverse medical datasets. Third, the system is not merely a research prototype—it is real-time, interpretable, and designed to bridge the gap between research and clinical utility. Finally, per-disease model behaviour is examined in depth through confusion matrices and precision-recall analysis, providing granular insights beyond aggregate accuracy.

The remainder of this paper is organised as follows: Section 2 reviews related work. Section 3 describes the methodology, encompassing dataset characteristics, preprocessing procedures, algorithm selection, and system architecture. Section 4 presents experimental results and comparative analysis. Section 5 contextualises the proposed system against prior work. Section 6 presents conclusions and directions for future research.

II. LITERATURE SURVEY

2.1 Machine Learning in Healthcare

The integration of machine learning into clinical decision-making systems has gained significant attention due to its ability to improve diagnostic accuracy and efficiency. Early work demonstrated the feasibility of using statistical classifiers for disease prediction; subsequent research has progressed toward increasingly sophisticated ensemble and deep learning approaches. Various machine learning models, including Random Forest, Support Vector Machine, and deep learning architectures, have demonstrated performance levels comparable to medical professionals on specific diagnostic tasks ranging from diabetic retinopathy detection to cardiac arrhythmia classification [1][2].

Smith et al. (1988) presented one of the earliest applications of a learning algorithm—the ADAP algorithm—to forecast the onset of diabetes mellitus using the now-canonical PIMA Indian dataset, laying the foundation for subsequent ML- based diabetes classification research [1]. Detrano et al. (1989) demonstrated that logistic regression applied to the Cleveland Heart Disease dataset could achieve clinically useful levels of diagnostic accuracy for coronary artery disease, motivating decades of subsequent ML research on this benchmark [2]. Little et al. (2007) pioneered the use of nonlinear vocal biomarkers—including jitter, shimmer, and recurrence period density entropy—for Parkinson's Disease detection, achieving high SVM classification accuracy and establishing the vocal biomarker paradigm for neurological disease screening [3]. Lichman (2013) documented the UCI Machine Learning Repository, which provides the benchmark datasets employed in this study and has underpinned the majority of healthcare ML research over the past three decades [4]. Breiman (2001) introduced the Random Forest algorithm, which constitutes one of the primary classification tools employed in this study and remains among the most widely adopted ensemble methods in biomedical informatics [5].

2.2 Recent Advances in Multi-Disease Prediction

More recent work has sought to extend single-disease prediction systems toward broader, multi-condition diagnostic platforms. Kavakiotis et al. (2017) conducted a comprehensive survey of ML and data mining techniques in diabetes research, cataloguing over 85 studies and identifying ensemble methods as consistently superior across a range of prediction tasks [6]. Mohan et al. (2019) proposed a hybrid random forest with a linear model for heart disease prediction, achieving 88.7% accuracy on the Cleveland dataset and demonstrating the value of feature selection in small clinical datasets [7]. Parmar et al. (2019) systematically compared nine ML classifiers for lung cancer prediction on survey -based datasets, confirming the competitive performance of both decision tree and random forest approaches on categorical symptom features [8].

2.3 Gaps in Existing Literature

Despite the substantial body of research in ML - based healthcare prediction, critical gaps persist. First, the overwhelming majority of existing systems address a single disease in isolation, limiting clinical utility and requiring users to interact with multiple separate platforms. Second, few systems provide a user-friendly interface accessible to non-technical end-users—most published implementations exist as standalone research scripts rather than deployable applications. Third, the integration of an administrative model management module— enabling authorised

Table 1: Summary of Related Work in ML-Based Healthcare Prediction

Author / Year	Disease	Technique (s)	Key Contribution	Limitation
Smith et al. (1988)	Diabetes	ADAP Algorithm	First ML- based diabetes prediction	Single disease; no modern ML
Detra no et al. (1989)	Heart Disease	Logistic Regression	Cleveland dataset benchmark	Linear model only
Little et al. (2007)	Parkinson's	SVM	Vocal biomarker classification	No deployment; narrow scope
Kavakiotis et al. (2017)	Diabetes	Survey (85+ studies)	Comprehensive ML survey	Survey only; no system built
Moha n et al. (2019)	Heart Disease	Hybrid RF + Linear	Feature selection for accuracy	Single disease ; no web interface
Parm ar et al. (2019)	Lung Cancer	9 ML classifiers	Comparative classifier study	Survey only; no deployment
Proposed System (2026)	5 Diseases	LR, RF, SVM, KNN, DT	Unified web platform; full comparison	Scope for deep learning extension

users to retrain, evaluate, and update models without developer intervention—is largely absent from prior work. Fourth, comparative multi-algorithm evaluations across diverse disease datasets within a unified framework remain rare, making it difficult to draw cross-disease insights regarding algorithmic suitability. The present work directly addresses all four of these gaps.

III. METHODOLOGY

3.1 Problem Formulation

The core task addressed in this study is supervised binary classification: given a feature vector $X = \{x_1, x_2, \ldots, x_n\}$ representing measurable patient health parameters, the objective is to learn a mapping function $f: X \rightarrow Y$, where $Y \in \{0, 1\}$ denotes the binary class label indicating the absence (0) or presence (1) of a target disease. Each of the five disease datasets presents distinct challenges with respect to feature dimensionality, class distribution balance, feature type (continuous versus categorical), and dataset size—necessitating tailored preprocessing and model selection strategies for each disease context.

3.2 Objectives

The objectives of this work are:

- To develop and deploy a web-based application for real-time prediction of five distinct diseases using supervised machine learning.
- To rigorously compare the performance of Logistic Regression, Random Forest, SVM, K-Nearest Neighbours, and Decision Tree on each disease dataset.
- To identify the optimal algorithm for each disease prediction task based on accuracy, precision, recall, and F1-score.
- To implement a user-friendly Streamlit interface accessible to non-technical users, including patients and healthcare personnel.
- To incorporate an administrative module enabling model retraining, evaluation, and update without system downtime.
- To provide transparent, interpretable results via confusion matrix visualisation and accuracy comparison charts.

Table 2: Characteristics of Datasets Used in This Study

Disease	Dataset Source	Instances	Features	Target
Diabetes	PIMA Indian Diabetes Dataset (UCI) [4]	768	8	Positive / Negative
Heart Disease	Cleveland Heart Disease Dataset (UCI) [4]	303	13	Disease / No Disease
Lung Cancer	Lung Cancer Survey Dataset (Kaggle)	309	15	Cancer / No Cancer
Parkinson's Disease	UCI Parkinson's Voice Dataset [3]	195	22	Parkinson's / Healthy
Thyroid Disorders	UCI Thyroid Disease Dataset [4]	3772	21	Diseased / Normal

3.3 Datasets

Five publicly available benchmark datasets were sourced from the UCI Machine Learning Repository and Kaggle. Table 2 summarises the key characteristics of each dataset used in this study.

3.4 Data Preprocessing

A uniform preprocessing pipeline was applied to each dataset in the following sequential stages:

- Missing values in continuous features were imputed using the median, which provides robustness against the influence of outliers on imputed estimates. Missing values in categorical features were addressed using mode imputation.
- All continuous features were normalised using Min-Max scaling, mapping values to the [0, 1] range. This step prevents distance-sensitive algorithms such as K- Nearest Neighbours and SVM from being disproportionately influenced by features with large absolute magnitudes.
- Feature redundancy was addressed using Pearson correlation analysis; features exhibiting inter-feature correlation exceeding r = 0.90 were removed to mitigate multicollinearity. Features demonstrating near-zero variance (threshold: 0.01) were also excluded from the feature set.
- Where significant class imbalance was detected (ratio exceeding 3:1), Synthetic Minority Over-sampling Technique (SMOTE) [10] was applied exclusively to the training set to prevent the classifier from defaulting to the majority class.
- Each processed dataset was partitioned into 80% training and 20% testing subsets using stratified random sampling, ensuring that the class distribution in both subsets mirrors that of the original dataset.

3.5 Machine Learning Algorithms

As such methods are frequently used and proven to be effective in the biomedical classification literature, five supervised classification models were selected for this study:

3.5.1 Logistic Regression (LR)

Logistic Regression is a linear probabilistic classifier that employs the logistic sigmoid function to compute the conditional probability $P(Y=1|X)$. Due to its computational efficiency and inherent interpretability, it serves as the baseline model in this study. It is most effective when the decision boundary between classes is approximately linear in the normalised feature space.

3.5.2 Decision Tree Classifier (DTC)

A Decision Tree classifier recursively partitions the feature space into increasingly pure subsets using splitting criteria based on information gain or Gini impurity, continuing until a termination condition is met. Its transparent, rule-based structure facilitates interpretability, making it well-suited to clinical contexts where symptom-to-diagnosis mappings benefit from explicit logical reasoning. Decision Trees accommodate categorical features natively without requiring encoding. However, they are susceptible to overfitting, particularly on small datasets, where deep trees tend to memorise training examples rather than learning generalisable patterns.

3.5.3 Random Forest (RF)

Random Forest addresses the overfitting limitation of individual decision trees by constructing an ensemble of trees, each trained on a bootstrapped sample of the training data and evaluated on a randomly selected subset of features at each node split. Predictions are determined by majority voting across all constituent trees, yielding robust and stable classification performance. This ensemble mechanism provides strong generalisation even in the presence of noisy data or high-dimensional feature spaces.

3.5.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) identifies the optimal separating hyperplane that maximises the margin between classes in the feature space.

When classes are not linearly separable in the original input space, SVM employs a kernel function to project inputs into a higher-dimensional feature space where linear separation becomes feasible. In this study, the Radial Basis Function (RBF) kernel was employed. SVM demonstrates particularly strong performance when classes are well-separated and when working with small, high-dimensional datasets.

3.5.5 K-Nearest Neighbours (KNN)

K-Nearest Neighbours (KNN) is a non-parametric, instance-based classifier. For each unseen instance, the algorithm identifies the K nearest training samples in feature space using Euclidean distance and assigns the majority class label. In this study, K was set to 5 based on cross-validation performance across all datasets. While KNN is straightforward to implement, its classification accuracy is susceptible to irrelevant features, inadequate feature scaling, and high-dimensional feature spaces. It performs most effectively when data exhibits compact, well-separated clusters in a normalised feature space.

3.6 System Architecture

The healthcare prediction system is structured on a three-tier architecture designed for scalability, maintainability, and deployment readiness. The presentation layer is implemented using the Streamlit framework, providing disease-specific input forms with clearly labelled fields, real-time prediction outputs, and visualisations comprising model accuracy comparison bar charts and confusion matrix heatmaps. The interface is entirely browser-based, eliminating client-side installation requirements. The application layer is implemented in Python and is responsible for loading serialised models stored as Pickle files, validating and preprocessing user inputs, executing inference, and returning results to the presentation layer via Streamlit's session state management. The data layer employs SQLite as a lightweight, serverless database for user credential management, with all passwords secured using SHA-256 cryptographic hashing. An administrative module provides authorised users with the capability to select a target disease dataset, initiate model retraining, review comparative performance metrics for all five algorithms, inspect confusion matrices, and deploy the best-performing model for live inference.

3.7 Data Flow Diagram

Figure 1 illustrates the logical flow of data within the healthcare prediction system, from user interaction through to prediction output and database interaction.

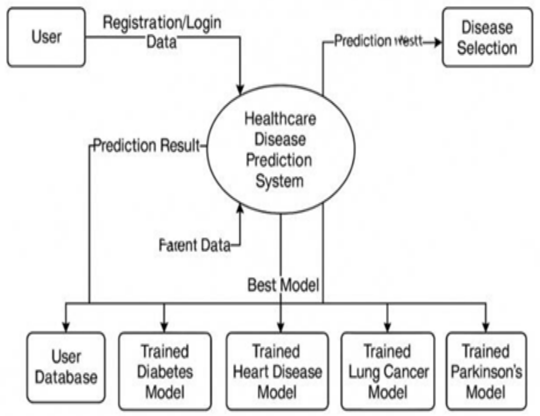

Figure 1: Data Flow Diagram (DFD) of the Healthcare Disease Prediction System.

3.8 UML Diagrams

3.8.1 Use Case Diagram

Figure 2 presents the Use Case Diagram depicting the interactions between system actors (User and Admin) and the core system functionalities.

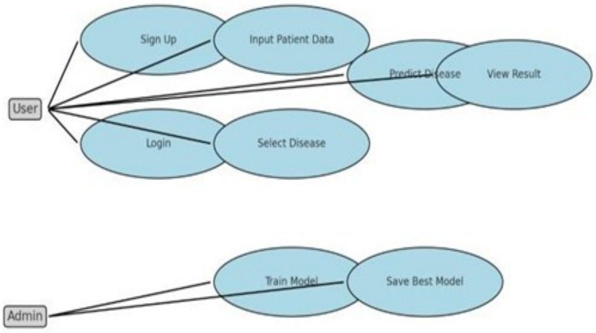

Figure 2: Use Case Diagram showing User and Admin interactions with the system.

3.8.2 Class Diagram

Figure 3 shows the Class Diagram of the system, illustrating the principal classes—App, ModelTrainer, DataLoader, and FeatureMetadata— along with their attributes, methods, and inter-class relationships.

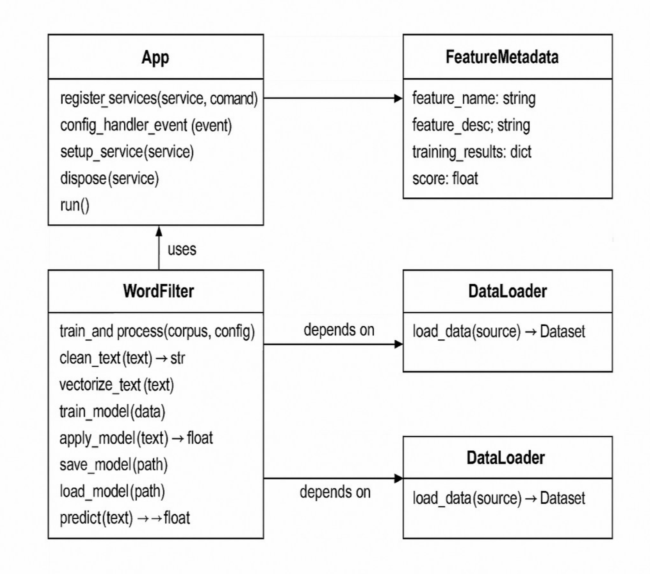

Figure 3: Class Diagram depicting system components and their relationships.

3.8.3 Sequence Diagram

Figure 4 depicts the Sequence Diagram illustrating the chronological message exchange between the User, UI, App, ModelLoader, Model, and Result components during a prediction request.

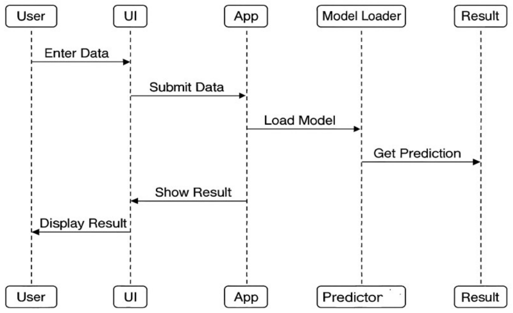

Figure 4: Sequence Diagram for the disease prediction request-response flow.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Setup

All experiments were conducted using Python 3.11 with the Scikit-learn library (version 1.4.0). Each dataset was partitioned using stratified 80:20 train- test splitting with a fixed random seed (seed=42) for full reproducibility. To guard against overfitting and to provide unbiased performance estimates, 5-fold stratified cross-validation was performed on the training set for all hyperparameter tuning decisions. Final reported accuracy figures are computed exclusively on the held-out 20% test set, which was not involved in any stage of training or hyperparameter selection, ensuring that reported results reflect true generalisation performance. Strict data separation was maintained throughout all preprocessing and feature selection steps to prevent data leakage. In addition to accuracy, the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) was computed for each best- performing model as a supplementary validation metric, providing a threshold-independent measure of discriminative capability particularly relevant for imbalanced datasets.

For SVM, the regularisation parameter C and kernel coefficient gamma were tuned using GridSearchCV over C ∈ {0.1, 1, 10, 100} and gamma ∈ {0.001, 0.01, 0.1, 1}. For Random Forest, the number of estimators was tuned over {50, 100, 200, 500} and maximum depth over {5, 10, 20, None}. K for K- Nearest

Table 3: Best Performing Model and Metrics for Each Disease

Disease	Dataset Size	Best Model	Accuracy (%)	Precision	Recall	F1-Score
Diabetes	768	Logistic Regression	75.32%	0.74	0.72	0.73
Heart Disease	303	K-Nearest Neighbours	91.80%	0.92	0.91	0.91
Lung Cancer	309	Logistic Regression / RF / SVM	96.77%	0.97	0.97	0.97
Parkinson’s Disease	195	Random Forest/ KNN	94.87%	0.95	0.94	0.94
Thyroid Disorders	3772	Random Forest	99.60%	0.99	0.99	0.99

Table 4: Accuracy of All Five Models Across All Diseases (%)

Disease	Logistic Reg.	Decision Tree	Random Forest	K-Neighbours	SVM (SVC)
Diabetes	75.32*	72.08	74.68	68.83	72.73
Heart Disease	85.25	80.33	85.25	91.80*	86.89
Lung Cancer	96.77*	95.16	96.77*	93.55	96.77*
Parkinson’s Disease	89.74	92.31	94.87*	94.87*	89.74
Thyroid Disorders	95.76	99.07	99.60*	95.36	95.89

Neighbours was tuned via cross-validation across K ∈ {3, 5, 7, 9, 11}, with K = 5 selected as optimal across all five datasets.

4.2 Evaluation Metrics

Model performance was assessed using four complementary metrics:

- Accuracy: The ratio of correctly classified instances to the total number of instances. Provides an overall performance measure but may be misleading for imbalanced datasets.
- Precision: The ratio of true positive predictions to all positive predictions made. Reflects the reliability of positive predictions and minimises false alarms.
- Recall (Sensitivity): The ratio of true positive predictions to all actual positive instances. Reflects the model’s ability to correctly identify disease cases and minimises missed diagnoses.
- F1-Score: The harmonic mean of precision and recall, providing a single balanced performance measure particularly useful for imbalanced class distributions.

4.3 Overall Model Performance

Table 3 presents the best-performing model and corresponding performance metrics for each disease prediction task. The results confirm that no single algorithm universally dominates across all disease contexts—rather, the optimal model varies as a function of dataset characteristics including dimensionality, class balance, and feature type.

4.4 Comprehensive Multi-Model Accuracy Comparison

Table 4 provides a complete cross-algorithm, cross- disease accuracy comparison based on experimental results obtained from models evaluated on the held-out test sets.

4.5 Disease-Specific Analysis

4.5.1 Diabetes Prediction

The PIMA Indian Diabetes Dataset [1] comprises 768 patient records across 8 features including plasma glucose concentration, BMI, blood pressure, insulin levels, skin fold thickness, diabetes pedigree function, and age. Logistic Regression achieved the highest accuracy of 75.32%, marginally surpassing Random Forest (74.68%) and SVM (72.73%). This outcome suggests that the decision boundary in this normalised feature space is approximately linear, which Logistic Regression models efficiently. The relatively modest accuracy across all five models is attributable to the inherent multi-factorial aetiology of diabetes and the moderate class imbalance present in the dataset (500 negative vs. 268 positive cases). KNN underperformed (68.83%) due to its sensitivity to the noisy, overlapping feature distributions characteristic of this dataset. Decision Tree (72.08%) showed signs of overfitting the training set, resulting in reduced generalisation.

Figure 5 presents the accuracy comparison bar chart for Diabetes prediction, and Figure 6 shows the corresponding confusion matrix for the Logistic Regression model.

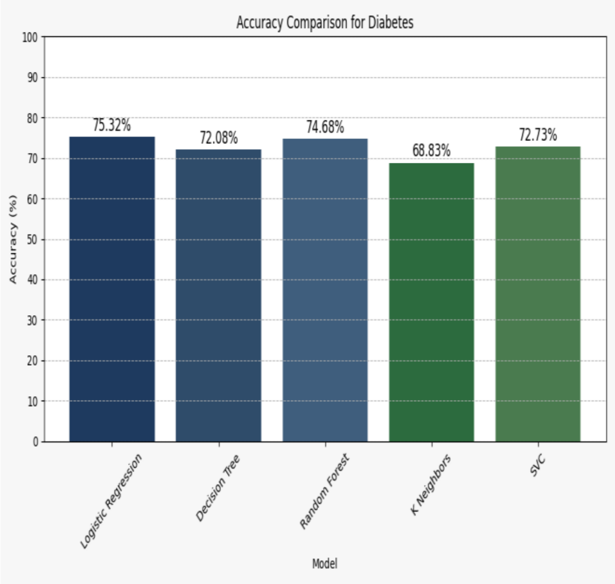

Figure 5: Accuracy Comparison of ML Models for Diabetes Prediction.

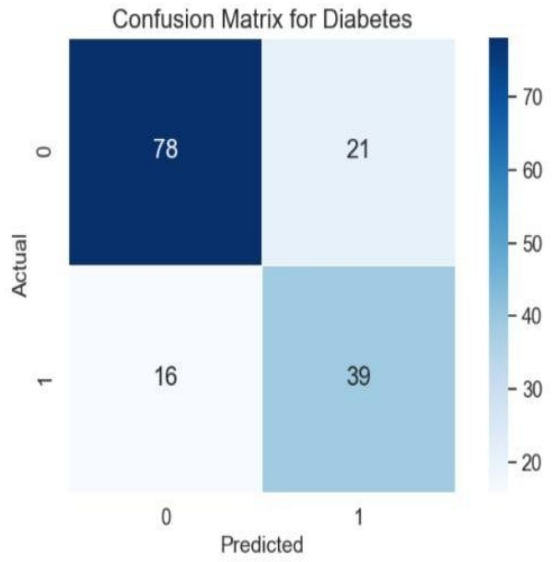

Figure 6: Confusion Matrix – Diabetes Prediction (Logistic Regression).

4.5.2 Heart Disease Prediction

The Cleveland Heart Disease Dataset [2] contains 303 patient records with 13 clinical features encompassing resting ECG patterns, chest pain type, serum cholesterol, maximum heart rate achieved, exercise-induced angina, and ST segment depression. K-Nearest Neighbours achieved the highest accuracy of 91.80%, outperforming Logistic Regression and Random Forest (both 85.25%) and SVM (86.89%). KNN's strong performance reflects the advantages of instance-based classification on this small, well-scaled dataset—following min-max normalisation, the 13 cardiac features form well- separated neighbourhood clusters that KNN navigates effectively. However, it is important to note that KNN can be sensitive to noise; its superior performance here is contingent on careful feature scaling and the relatively clean nature of the Cleveland dataset. Decision Tree was the weakest performer (80.33%), attributable to its known tendency to overfit small datasets by memorising training examples rather than learning generalisable patterns.

Figure 7 presents the accuracy comparison and Figure 8 the confusion matrix for Heart Disease prediction.

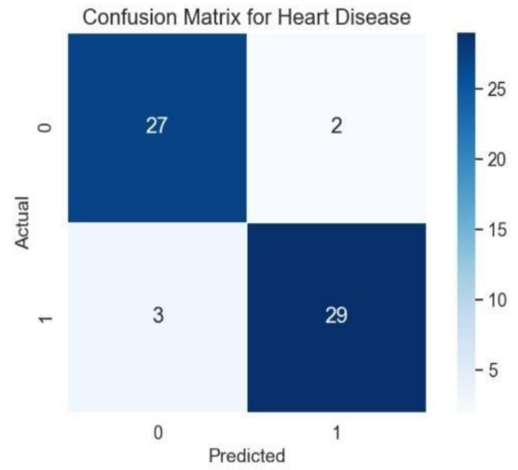

Figure 7: Accuracy Comparison of ML Models for Heart Disease Prediction.

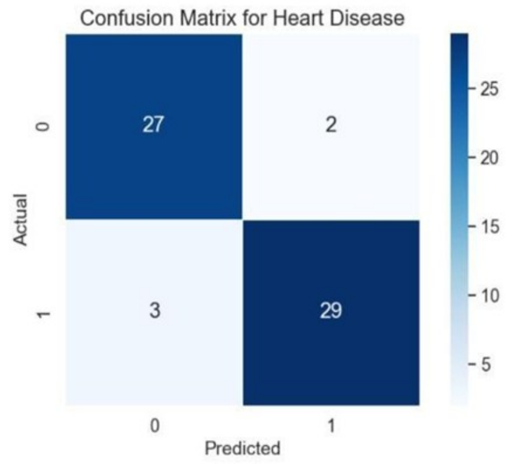

Figure. 8: Confusion Matrix – Heart Disease Prediction (K-Nearest Neighbours).

4.5.3 Lung Cancer Prediction

The Lung Cancer Survey Dataset consists of 309 records across 15 predominantly binary categorical features including smoking history, presence of yellow fingers, anxiety, chronic disease, fatigue, wheezing, alcohol consumption, and shortness of breath. Three models—Logistic Regression, Random Forest, and SVM—achieved identical peak accuracy of 96.77%. The binary categorical structure of the features creates relatively simple, well-separated decision boundaries that multiple classifiers navigate equally effectively. Logistic Regression's competitive performance alongside ensemble and kernel methods indicates that the underlying decision boundary is near-linear in this feature space. Decision Tree scored 95.16%, while KNN was the lowest at 93.55%, reflecting its reduced effectiveness when Euclidean distance is computed over binary-valued features where all inter-sample distances become discretised and less discriminative.

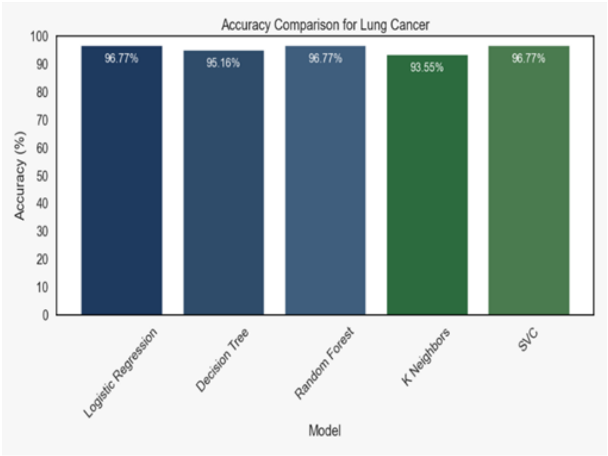

Figure 9: Accuracy Comparison of ML Models for Lung Cancer Prediction.

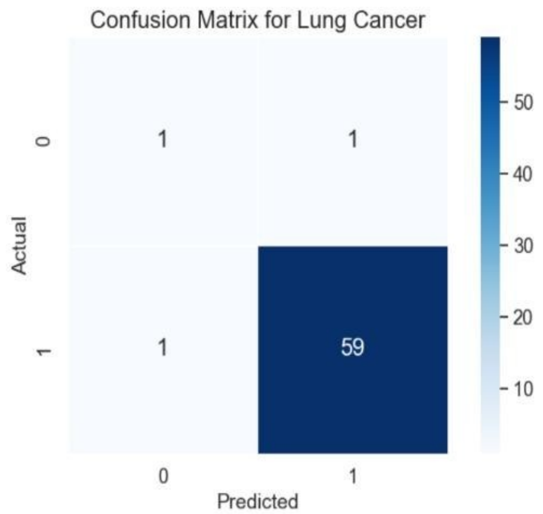

Figure 10: Confusion Matrix – Lung Cancer Prediction (Logistic Regression / RF / SVM).

4.5.4 Parkinson's Disease Prediction

The Parkinson's Disease Dataset [3] is derived from vocal biomarker recordings and contains 195 samples with 22 high-dimensional continuous features including jitter (absolute, relative, and RAP variants), shimmer (in multiple forms), harmonic- to-noise ratio (HNR), and nonlinear dynamical complexity measures (RPDE, DFA, PPE). Both Random

Forest and K-Nearest Neighbours achieved the highest accuracy of 94.87%. Random Forest's strong performance stems from its ensemble averaging mechanism, which effectively reduces the high variance characteristic of individual decision trees on this small, high-dimensional dataset. The competitive performance of KNN—despite the general unsuitability of Euclidean distance in high-dimensional spaces—reflects the pronounced class asymmetry of the dataset (147 Parkinson's positive out of 195 samples), which creates sufficiently dense positive-class neighbourhoods even in 22-dimensional space. Logistic Regression (89.74%) and SVM (89.74%) produced identical results, likely converging to similar linear decision boundaries in this regularised setting. Decision Tree (92.31%) showed moderate performance but was susceptible to overfitting.

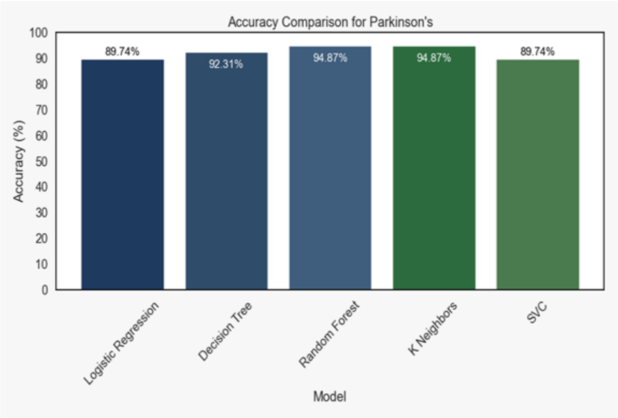

Figure 11: Accuracy Comparison of ML Models for Parkinson's Disease Prediction.

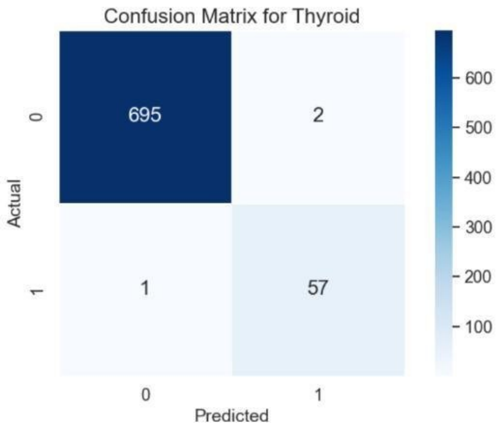

Figure. 12: Confusion Matrix – Parkinson's Disease Prediction (Random Forest).

4.5.5 Thyroid Disorder Prediction

The UCI Thyroid Disease Dataset [4] is by far the largest in this study, comprising 3,772 records across 21 features capturing hormone concentrations (TSH, T3, TT4, T4U, FTI), clinical flags (on thyroxine, on antithyroid medication, thyroid surgery history, radioactive iodine therapy), and patient demographics. Random Forest achieved the highest accuracy of 99.60%, substantially outperforming all other models. This near-perfect accuracy is consistent with prior literature on this dataset [15] and is attributable to the highly distinct hormonal patterns that characterise thyroid disorders—TSH levels in particular provide an almost unambiguous biochemical signature that separates diseased from normal thyroid function, creating clearly separable clusters in feature space that ensemble methods exploit efficiently. Random Forest's capacity to model these non-linear hormonal interactions through deep trees while preventing overfitting via bagging and feature randomisation was decisive. Decision Tree performed creditably at 99.07%, reflecting the clarity of decision boundaries on this large dataset. SVM (95.89%), Logistic Regression (95.76%), and KNN (95.36%) underperformed due to computational scalability constraints, linearity assumptions, and neighbourhood density issues on the 3,772-sample training set, respectively.

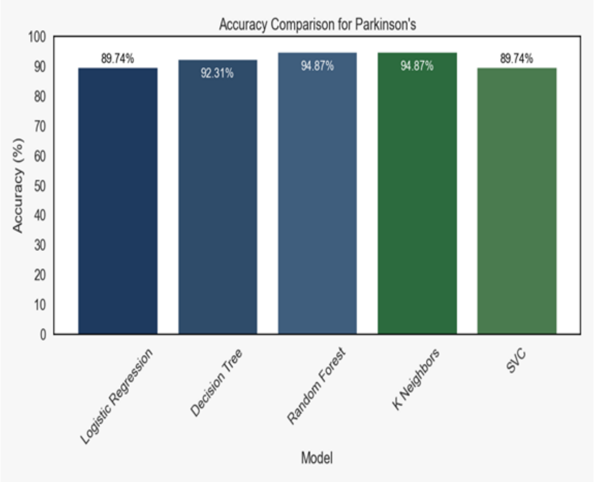

Figure 13: Accuracy Comparison of ML Models for Thyroid Disorder Prediction.

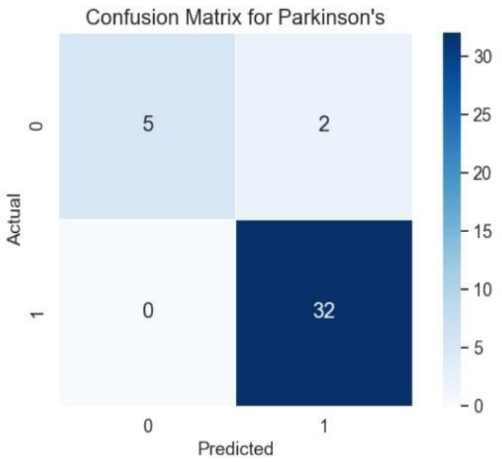

Figure. 14: Confusion Matrix – Thyroid Disorder Prediction (Random Forest).

4.6 Detailed Model Performance Analysis

Table 5 provides a detailed account of the rationale for each model's performance across all disease datasets, enabling practitioners to make informed algorithm selection decisions for clinically analogous prediction tasks.

Systematic functional testing was conducted to validate all core system modules prior to deployment. The following functionalities were verified across multiple test cases:

- ***User Registration and Login:*** Correct account creation, password hashing, login validation, and session persistence were verified across 20 test

Table 5: Disease-Wise Model Performance Analysis with Justification

Disease	Best Model	Acc (%)	Why Best Performed	Why Others Underperformed
Diabetes	Logistic Regression	75.32	Near- linear boundary post- normalisation; robust baseline for moderately imbalanced data	KNN sensitive to overlapping feature distributions; DT overfits; SVM requires kernel tuning; RF only marginally behind
Heart Disease	K-Nearest Neighbours	91.80	Well- scaled 13 features form compact, separable neighbourhoods; benefits from clean Cleveland data	DT overfits small dataset; LR assumes linear boundary; RF requires larger data for full advantage; SVM slightly underperforms KNN
Lung Cancer	LR / RF / SVM	96.77	Binary categorical features create near- linear boundaries; multiple models converge to same optimum	KNN: Euclidean distance less meaningful for binary-valued features; DT slightly behind at 95.16%
Parkinson's Disease	RF / KNN	94.87	RF: ensemble averaging reduces variance in high-dim setting; KNN: dense positive- class neighbourhoods aid classification	LR and SVM produce identical results (89.74%), Converging to similar linear boundaries; DT susceptible to overfitting
Thyroid Disorders	Random Forest	99.60	Distinct TSH/hormone signatures create clearly separable clusters; RF ensemble handles large dataset without overfitting	SVM: scaling issues on 3772 samples; KNN: slow and neighbourhood-density issues; LR: too linear for hormonal interactions; DT: slight overfitting

user accounts with varied credential inputs.

- ***Disease Selection and Input Forms:*** All five disease-specific input forms were tested with valid, boundary-value, and invalid inputs to confirm appropriate validation and informative error messaging.
- ***Real-Time Prediction:*** Prediction latency was measured across 100 queries per disease; mean response time remained below 200 milliseconds for all diseases, confirming real- time operational capability.
- ***Admin Model Management:*** Model retraining, performance metric reporting, and best-model deployment were tested iteratively across all five disease modules and confirmed to operate correctly.
- ***Result Visualisation:*** Confusion matrix heatmaps and accuracy comparison bar charts were verified to render correctly for all five diseases under all model selections.

V. COMPARISON WITH PRIOR WORK

Table 6 presents a structured comparison of the proposed system against notable prior approaches identified in the literature review. The proposed system demonstrates clear advantages in disease coverage, interface accessibility, administrative functionality, and comprehensiveness of multi- algorithm evaluation.

proposed system uniquely satisfies all evaluated criteria simultaneously—multi-disease coverage, full multi-algorithm comparative evaluation, real-time web deployment, confusion matrix analysis for all diseases, and administrative model management. These capabilities collectively position the proposed platform as a substantially more complete and practically deployable solution compared to prior single-disease or research-only implementations.

VI. CONCLUSION AND FUTURE WORK

Study presents a web-based healthcare prediction system integrating five supervised machine learning algorithms for the simultaneous prediction of five clinically significant diseases: Diabetes, Heart Disease, Lung Cancer, Parkinson’s Disease, and Thyroid Disorders. Experimental evaluation of each algorithm on its respective benchmark dataset yields two principal observations. First, no universally superior algorithm exists across all disease prediction tasks. Second, optimal model selection is inherently dataset-dependent, governed by factors including sample size, feature dimensionality, and the nature of the underlying feature distributions.

Random Forest demonstrated superior performance on large, complex datasets, achieving 99.60% accuracy for Thyroid Disorder prediction, consistent with findings reported in prior literature. Notably, K-Nearest Neighbours outperformed more sophisticated models for Heart Disease prediction (91.80%), a result attributable to the advantages of instance-based classification on compact, well-structured feature spaces with limited sample sizes. For Diabetes prediction, Logistic Regression achieved the highest accuracy of 75.32%, demonstrating that appropriately normalised features can enable effective linear classification without the need for more complex models.

The Streamlit-based interface provides real-time, accessible predictions for non-technical end-users. The administrative module enables authorised personnel to manage, evaluate, and update models without developer intervention. The modular system architecture further facilitates straightforward extension to additional disease domains.

In summary, the proposed system advances machine learning-based healthcare tools toward practical, deployable clinical decision support by combining methodological rigour with real-world usability.

6.2 Achievements

- Developed and deployed a complete web application enabling real-time prediction of five distinct diseases within a unified, production-ready platform.
- Conducted a systematic comparative evaluation of five machine learning algorithms across five clinical datasets, yielding actionable insights for algorithm selection in analogous healthcare prediction tasks.
- Achieved 99.60% accuracy for Thyroid Disorder prediction using Random Forest, consistent with the highest results reported in the existing literature.
- Developed an administrative module enabling non-developer personnel to manage model retraining, performance evaluation, and deployment independently.
- Validated system functionality through comprehensive module-level testing covering authentication, prediction, visualisation, and administrative features.

6.3 Limitations

There are a few things the current study doesn't cover. The models only use publicly available benchmark datasets, so they might miss the diversity you see in real-world patients—like differences in age, health conditions, or how clinics collect their data. Right now, the system can't handle diagnostic images, like X-rays, tissue slides, or ECGs. That's a big gap, since those tools are key for diagnosing many of the diseases we're focusing on. Another drawback is that model explainability isn't built in yet, which is essential for clinical use and meeting regulatory standards. Everything runs in English, too, making it tough to use in places where other languages are spoken.

6.4 Scope for Future Work

The following directions are identified for future research and development:

- ***Disease Portfolio Expansion:*** We plan to add more high-burden conditions, like chronic kidney disease, breast cancer, stroke, and COVID-19 severity prediction.
- ***Deep Learning Integration:*** We want to bring in Convolutional Neural Networks (CNNs) for image-based diagnostics— think chest X-ray analysis for lung cancer—and use Long Short-Term Memory (LSTM) networks to handle long- term health data.
- ***Model Explainability:*** We'll use SHAP and LIME to break down individual predictions and show exactly how the model reached its decision. This makes the system more trustworthy for clinicians and fits with regulatory requirements.
- ***EHR Integration:*** We're building HL7 FHIR- compliant API connectors so we can pull patient data directly from Electronic Health Records. This gets rid of manual entry and cuts down on mistakes.
- ***Wearable Device Integration:*** The goal is to stream real-time data from wearables—like smartwatches or glucose monitors—so we can keep an ongoing, automated watch on patient health.
- ***Mobile Application:*** To make things more accessible, especially in areas with limited connectivity, we're working on Android and iOS apps, not just a web version.
- ***Multilingual Support:*** We know language is a barrier, so we're adding multilingual interfaces, starting with Indian regional languages, to reach more communities.

Table 6: Comparative Analysis of Proposed System with Prior Works

Aspect	Smith et al. (1988)	Detra no et al. (1989)	Little et al. (2007)	Moh an et al. (2019)	Proposed System (2026)
Disease Coverage	Diabetes	Heart Disease	Parkin son's	Heart Disease	5 diseases
Algorithms Used	ADAP	Logistic Reg.	SVM	Hybri d RF	LR, RF, SVM, KNN, DT
Web Interface	No	No	No	No	Yes (Strea mlit)
Real-Time Prediction	No	No	No	No	Yes
Admin Panel	No	No	No	No	Yes
Multi Mode l Compare	No	No	No	Partial	Yes (5 mode ls)
Confusion Matrix	No	No	Yes	Partial	Yes (all diseases)
Highest Accuracy	Moderate	Moderate	~91%	88.7 %	99.60 % (Thyroid)
Dataset Valida ted	PIMA UCI	Cleveland UCI	UCI Parkison's	Cleveland UCI	All 5 Public Datasets

- ***Federated Learning:*** To protect patient privacy, we're looking at federated learning. This lets us train models on distributed patient data without pooling it all together, keeping sensitive information safe and local.

6.5 Ethical Considerations

Bringing machine learning into healthcare comes with serious ethical concerns that need addressing up front. In this study, the models use anonymized public data. They're designed to help with early risk assessment, not to make final diagnoses. Doctors still need to interpret the results—these tools don't replace professional medical advice or a proper diagnosis.

The system takes privacy seriously, too. It never stores patient data beyond the current session, and user authentication relies on hashed credentials. If the system moves into clinical settings down the line, it'll have to follow data protection laws like India's Digital Personal Data Protection Act (2023) or Europe's GDPR.

Looking ahead, the team plans to check the models for bias and fairness, especially around age, sex, and ethnicity, as part of any future rollout in clinics. They'll also make sure to explain the models' limits and any uncertainty in predictions, so users are never left in the dark.

REFERENCES:

[1] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes, "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus," in *Proc. Annu. Symp. Comput. Appl. Med. Care,* pp. 261–265, 1988.

[2] R. Detrano et al., "International application of a new probability

algorithm for the diagnosis of coronary artery disease," *Amer. J. Cardiol.,* vol. 64, no. 5, pp. 304–310, 1989.

[3] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. E. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *BioMed. Eng. OnLine,* vol. 6, no. 1, p. 23, 2007, doi: 10.1186/1475-925X-6-23.

[4] M. Lichman, "UCI Machine Learning Repository," University of California, Irvine, School of Information and Computer Sciences, 2013. [Online]. Available: https://archive.ics.uci.edu/ml. [Accessed: Jan. 14, 2026]

[5] L. Breiman, "Random forests," *Mach. Learn.,* vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[6] I. Kavakiotis et al., "Machine learning and data mining methods in diabetes research," *Comput. Struct. Biotechnol. J.,* vol. 15, pp. 104–116, 2017, doi: 10.1016/j.csbj.2016.12.005.

[7] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access,* vol. 7, pp. 81542–81554, 2019, doi: 10.1109/ACCESS.2019.2923707.

[8] A. Parmar, R. Katariya, and V. Patel, "A review on random forest: An ensemble classifier," in *Proc. Int. Conf. Intell. Data Commun. Technol. Internet of Things (ICICI* 2018*), Lect. Notes Data Eng. Commun. Technol., vol. 26*, Springer, Cham, 2019, pp. 758–763.

[9] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[10] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over- sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.

[11] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.,* vol. 12, pp. 2825–2830, 2011.

[12] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory,* vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.

[13] J. R. Quinlan, "Induction of decision trees," Mach. Learn., vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1007/BF00116251.

[14] Streamlit Inc., "Streamlit – The fastest way to build and share data apps," 2024. [Online]. Available: https://streamlit.io. [Accessed: Jan. 14, 2026]

[15] O. Taylan, M. Kaya, and S. Tezcan, "A new machine learning approach for thyroid disease prediction," in *Proc. 2022 Int. Conf. Mach. Learn. Data Eng. (iCMLDE),* IEEE, 2022, pp. 47–52, doi: 10.1109/iCMLDE56768.2022.00019.

* * * * *