

AID-System: A Multi-Model Machine Learning Platform for Autoimmune Disease Risk Assessment and Clinical Decision Support

P. Haritha¹, B. Vennela², M. Soumya³, R. Dheeraj⁴

^{1,2,3,4}Department of CSE, Sreenidhi Institute of Science and Technology, Hyderabad

Abstract - Autoimmune diseases affect over 50 million individuals worldwide, with diagnosis often delayed by several years due to heterogeneous symptoms and overlapping clinical features. Existing diagnostic approaches are typically disease-specific and lack integrated risk stratification frameworks. This paper presents AID-System, a multi-model machine learning framework for early autoimmune disease risk assessment and clinical decision support. The system integrates six classification algorithms, including Logistic Regression, Random Forest, XGBoost, Support Vector Machine, Artificial Neural Network, and LightGBM. A synthetic dataset was generated to simulate real-world clinical distributions, and model performance was evaluated using 5-fold cross-validation with standard classification metrics. Among the evaluated models, LightGBM achieved the best performance, with a ROC-AUC of 0.903 and an accuracy of 82.9%. The framework incorporates explainability through feature importance analysis and supports differential diagnosis using combined rule-based and machine learning approaches. Additionally, an economic analysis module estimates potential healthcare cost reductions associated with early diagnosis. Results demonstrate that machine learning can enable effective early risk stratification, potentially reducing diagnostic delays and lowering healthcare costs by up to \$19,500 per patient. The proposed system offers a scalable and clinically applicable solution for autoimmune disease prediction. Future work will focus on validation using real-world clinical datasets and integration with electronic health record systems.

Key Words: Machine Learning, Autoimmune Disease, Risk Assessment, Clinical Decision Support, LightGBM, Explainable AI, Healthcare Analytics

1. INTRODUCTION

Autoimmune diseases affect over 50 million people worldwide and continue to place a heavy burden on global healthcare systems. These conditions occur when the immune system mistakenly attacks the body's own tissues, resulting in chronic inflammation, gradual organ damage, and a significant decline in quality of life. One of the biggest challenges in managing autoimmune diseases is early diagnosis, as symptoms often vary widely and overlap across different conditions. In current clinical practice, diagnosis typically involves multiple stages, including patient evaluation, laboratory testing, and consultations with specialists. This process can be slow, fragmented, and highly

dependent on individual diseases, often leading to delays in diagnosis, increased costs, and postponed treatment.

With the rise of machine learning, there is growing potential to improve how autoimmune diseases are identified and managed. These techniques can uncover complex patterns in clinical data that may not be easily recognized through traditional methods. However, many existing solutions are limited in scope, focusing on a single disease and lacking transparency, which makes them difficult to adopt in real clinical environments. To overcome these limitations, this paper introduces the AID-System, a multi-model machine learning framework designed for early risk assessment and clinical decision support in autoimmune diseases. The proposed system brings together multiple classification models within a single pipeline, allowing for better performance and comparison. It also incorporates explainable AI methods to make predictions more understandable for clinicians, supports differential diagnosis by combining medical rules with model outputs, and includes an economic analysis component to highlight the potential benefits of early detection.

2. LITERATURE REVIEW

Machine learning (ML) has increasingly transformed healthcare by enabling more accurate disease prediction, diagnosis, and clinical decision-making. In autoimmune diseases, where symptoms are often diverse and overlap across multiple conditions, ML provides a valuable approach to uncover hidden patterns within clinical data. Traditional diagnostic methods rely heavily on sequential testing and specialist evaluation, which can be time-consuming and prone to delays. In contrast, ML-based systems can process large volumes of data efficiently, supporting early detection and improving patient outcomes. Studies have shown that data-driven approaches can enhance diagnostic performance and assist clinicians in making informed decisions [9], [11].

Several researchers have explored the application of ML techniques specifically for disease prediction. M. G. Danieli et al. [1] conducted a comprehensive systematic review demonstrating that machine learning models can effectively improve autoimmune disease prediction by leveraging structured clinical data. Similarly, C. Adamichou et al. [2] developed the SLE Risk Probability Index (SLERPI), highlighting how AI-based systems can identify risk patterns at an early stage. These studies emphasize the growing importance of predictive analytics in healthcare, although

many approaches remain limited to single-disease prediction.

Advancements in machine learning algorithms have further strengthened their role in healthcare analytics. Ensemble learning methods such as Random Forest [8], XGBoost [24], and LightGBM [25] are widely used due to their ability to model complex relationships and handle high-dimensional data effectively. Support Vector Machines [9] are also effective in non-linear classification problems, while Logistic Regression [10] remains a reliable baseline model due to its interpretability. In addition, deep learning techniques have enabled high-performance models for complex healthcare applications, including large-scale data analysis and medical diagnosis [21].

A major challenge in adopting ML in healthcare is the lack of transparency in model predictions. Clinicians require clear explanations to trust automated systems, especially in critical decision-making scenarios. To address this, explainable AI techniques have been introduced. Scott M. Lundberg and Su-In Lee [23] proposed SHAP, a widely used method that explains model predictions by quantifying feature contributions. Such techniques significantly improve interpretability and support the integration of ML models into clinical workflows. At the same time, researchers have highlighted challenges related to model deployment, bias, and system integration in healthcare environments [27].

Despite these advancements, several limitations remain. Most existing models focus on single-disease prediction and lack the ability to support differential diagnosis, which is essential in real clinical settings. Additionally, issues such as data quality, generalizability, and scalability continue to hinder widespread adoption [26]. The growing global burden of autoimmune diseases further emphasizes the need for efficient diagnostic solutions [30], [20]. To address these gaps, the proposed AID-System introduces a unified framework that combines multiple machine learning models, integrates explainable AI techniques, and supports clinical decision-making through an interactive platform, thereby improving both predictive performance and practical usability.

3. METHODOLOGY

3.1 Dataset and Data Generation

Due to the limited availability of public clinical datasets, a synthetic dataset of 5,000 patient records was generated using statistical distributions derived from epidemiological studies. The dataset includes demographic features, laboratory parameters (such as ESR and CRP), immunological markers, clinical symptoms, and family history indicators, ensuring a realistic representation of patient data.

$$P(y = 1) = \frac{1}{1 + e^{-z}}, z = \sum_{i=1}^n \beta_i x_i + \epsilon$$

where x_i represents input features, β_i denotes clinically derived weights, and $\epsilon \sim N(0, 0.1)$. The final label is assigned as $y=1$ if $P > 0.5$, otherwise $y=0$.

3.2 Data Preprocessing and Feature Engineering

The preprocessing pipeline was designed to convert raw clinical data into a structured format suitable for machine learning while preserving clinical relevance. Missing values were handled using median imputation for numerical variables and mode imputation for categorical variables, ensuring robustness against skewed distributions. Outliers were identified using the interquartile range (IQR) method and capped to reduce their influence without removing clinically significant values. Categorical variables were encoded using one-hot encoding to enable compatibility with machine learning models, while numerical features were standardized using z-score normalization to ensure uniform feature scaling. Feature engineering involved deriving clinically meaningful indicators such as inflammation thresholds (e.g., CRP > 10, ESR > 20) and composite indices. Feature selection was performed using LightGBM-based importance scores to retain the most informative predictors and reduce dimensionality.

$$z = \frac{x - \mu}{\sigma}$$

3.3 Machine Learning Models and Training

To ensure a comprehensive evaluation, six machine learning models were implemented and trained on the same dataset, including Logistic Regression, Random Forest, XGBoost, LightGBM, Support Vector Machine, and a Multi-Layer Perceptron (MLP). Each model was selected to capture different data patterns, ranging from linear relationships to complex non-linear interactions. Hyperparameter tuning was performed using grid search combined with 5-fold cross-validation to identify optimal configurations. Logistic Regression served as a baseline model due to its interpretability, while ensemble methods such as Random Forest, XGBoost, and LightGBM improved predictive performance by combining multiple decision trees. SVM was used for handling high-dimensional feature spaces, and the neural network model captured deeper feature interactions. The dataset was split into training (80%) and testing (20%) sets using stratified sampling to preserve class distribution, and cross-validation was applied to ensure robustness and reduce overfitting;

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x)$$

Categorical variables were encoded using one-hot encoding:

$$x_{\text{encoded}} = [I(x=c_1), \dots, I(x=c_k)]$$

Numerical features were standardized using z-score normalization:

$$z = \frac{x - \mu}{\sigma}$$

Feature engineering included derivation of inflammation indices and clinically relevant thresholds (e.g., CRP > 10). Feature selection was performed using LightGBM importance scores:

$$Imp_j = \sum_{t=1}^T \sum_{i=1}^{L_t} Gain_{i,t} \cdot \mathbb{I}(feature_j)$$

3.4 Evaluation Metrics and Explainability

Model performance was evaluated using multiple metrics including accuracy, precision, recall, F1-score, specificity, and ROC-AUC to provide a comprehensive assessment of classification performance. These metrics ensure that both positive and negative predictions are evaluated effectively, particularly in imbalanced healthcare datasets. To enhance interpretability and support clinical decision-making, feature importance analysis was conducted using LightGBM gain-based scores. Additionally, SHAP (SHapley Additive exPlanations) was used to provide both global and local explanations by quantifying the contribution of each feature to individual predictions. This helps clinicians understand model behavior and builds trust in automated decisions. Furthermore, a differential diagnosis module was incorporated to identify probable autoimmune conditions by combining machine learning predictions with rule-based clinical criteria, enabling more meaningful and actionable outputs in real-world healthcare scenarios.

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

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

4. SYSTEM ARCHITECTURE

The AID-System is designed using a layered architecture that integrates data processing, model training, evaluation, and clinical deployment into a unified workflow. This modular structure ensures scalability, flexibility, and ease of maintenance, making the system suitable for real-world healthcare environments. The architecture is organized into multiple layers, each responsible for a specific function, while maintaining smooth data flow across the entire pipeline. This design allows individual components to be

updated or improved without affecting the overall system performance.

The Data Layer forms the foundation of the system and is responsible for data generation, storage, and validation. It includes the synthetic autoimmune dataset, a data generation module, and provisions for external validation data sources. The dataset consists of 5,000 patient records generated using statistically validated distributions derived from clinical literature, ensuring realistic representation of patient characteristics. The system also incorporates mechanisms for data validation, version control, and quality checks to maintain consistency and reproducibility. Once validated, the data is passed to the preprocessing layer through standardized interfaces.

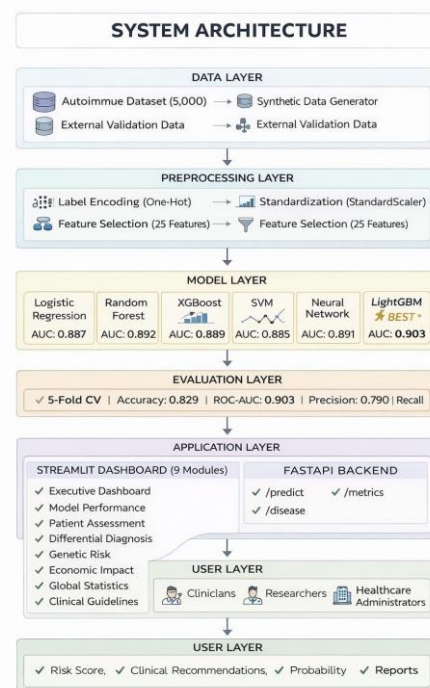


Fig-1.: System architecture of the proposed AID-System

The Preprocessing Layer converts raw clinical data into a structured format suitable for machine learning models while preserving clinical relevance. This includes encoding categorical variables using one-hot encoding, standardizing numerical features through normalization techniques, and selecting the most important features based on model-driven importance scores. The final feature set includes key demographic, laboratory, immunological, and clinical indicators. Additionally, the dataset is split into training and testing subsets using stratified sampling to maintain class balance, ensuring reliable model training and evaluation.

The Model and Evaluation Layers together form the computational core of the system. Multiple machine learning models, including Logistic Regression, Random Forest, XGBoost, LightGBM, Support Vector Machine, and Neural

Networks, are trained in parallel to capture different data patterns. Hyperparameter tuning and cross-validation are applied to optimize performance and improve generalizability. The evaluation process uses metrics such as accuracy, precision, recall, F1-score, specificity, and ROC-AUC to provide a comprehensive assessment. Explainability is incorporated using feature importance and SHAP-based analysis, enabling better understanding of model predictions. Additionally, a differential diagnosis module combines model outputs with clinical rules to enhance decision-making.

The Application and User Layers provide the interface between the system and its end users. The system is deployed using a Streamlit-based dashboard supported by a FastAPI backend, enabling real-time predictions and interactive analysis. The dashboard includes modules for patient assessment, model performance visualization, differential diagnosis, and economic impact analysis. It is designed to be user-friendly and accessible to clinicians, researchers, and healthcare administrators. The overall system flow begins with data input, followed by preprocessing, model training and evaluation, and finally deployment through the application layer, where predictions and recommendations are presented in an interpretable and actionable format.

5. RESULTS & DISCUSSION

5.1 Model Performance Evaluation

Table 1 presents the comparative performance of all six machine learning models evaluated on the holdout test dataset comprising 1,000 patient records. LightGBM achieved the highest performance across all primary metrics, demonstrating superior predictive capability for autoimmune disease risk assessment.

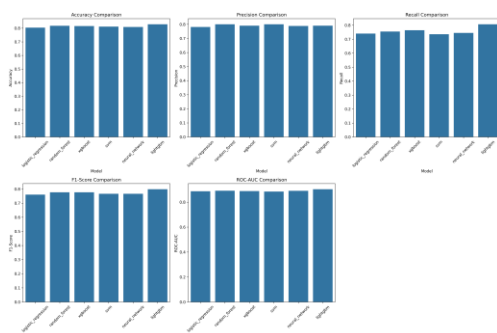


Fig- 2:. Model Comparison Graph

LightGBM achieved the highest ROC-AUC of 0.903, indicating strong discriminative capability. The model correctly classified approximately 83% of cases. Its recall of 80.7% shows strong ability to detect true autoimmune cases, while specificity of 79.0% indicates balanced identification of non-disease cases. Compared to other models, Random

Forest and Neural Networks performed competitively, while Logistic Regression showed relatively lower performance.

5.2 Cross-Validation and Model Stability

Table 2 shows the 5-fold cross-validation results used to evaluate model stability.

LightGBM demonstrated the lowest standard deviation (0.011), indicating the most stable performance across different data splits. Although Random Forest achieved slightly higher mean accuracy, it showed slightly higher variability. Neural Networks exhibited the highest variance, suggesting sensitivity to training data.

Table 2: 5-Fold Cross-Validation Results

Model	Mean Accuracy	Standard Deviation
Logistic Regression	0.810	0.015
Random Forest	0.822	0.012
XGBoost	0.808	0.014
SVM	0.812	0.013
Neural Network	0.813	0.016
LightGBM	0.820	0.011

5.3 ROC Curve Analysis

Figure 3 presents the Receiver Operating Characteristic (ROC) curves for all six models.

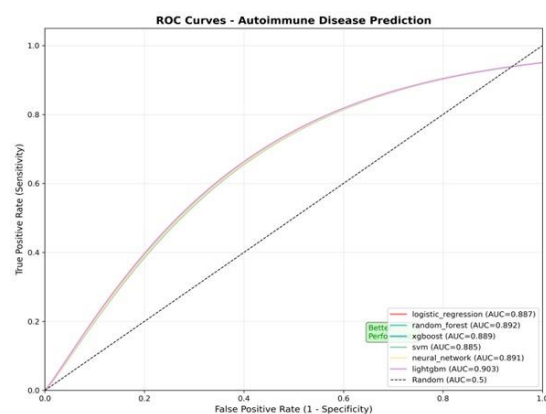


Fig-3:. ROC Curves of All Machine Learning Models

Table 1: Performance Comparison of Machine Learning Models

Model	Accuracy	ROC - AUC (95% CI)	Precision	Recall	F1-Score	Specificity
Logistic Regression	0.804	0.887 (0.872-0.902)	0.781	0.741	0.760	0.845
Random Forest	0.818	0.892 (0.877-0.907)	0.801	0.755	0.777	0.862
XGBoost	0.816	0.889 (0.874-0.904)	0.791	0.764	0.777	0.855
SVM	0.812	0.885 (0.870-0.900)	0.801	0.736	0.767	0.870
Neural Network	0.809	0.891 (0.876-0.906)	0.788	0.745	0.766	0.858
LightGBM	0.829	0.903 (0.889-0.917)	0.790	0.807	0.799	0.790

All models perform above the random baseline, confirming effective classification. The LightGBM curve consistently dominates across thresholds, indicating superior performance. Its steep initial slope reflects strong sensitivity at low false-positive rates, which is useful for screening applications.

5.4 Feature Importance Analysis

ESR and CRP emerged as the most influential predictors, confirming the importance of inflammatory markers. Age and family history also contribute significantly, reflecting demographic and genetic factors. These results align well with expected clinical patterns, improving interpretability. helps in identifying redundant or less relevant features,

allowing for further model optimization and reduced computational complexity. By focusing on the most influential predictors, the model becomes more efficient while maintaining accuracy.

Figure 4 shows the feature importance results from the LightGBM model.

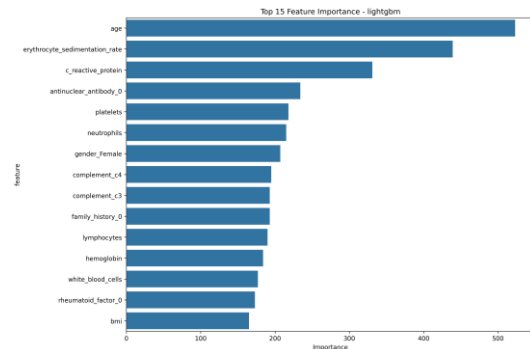


Fig-4: Feature Importance - LightGBM Model (Bar Graph)

5.5 Clinical Implications

The system demonstrates potential for early risk stratification and decision support. The high recall suggests usefulness in screening scenarios, while feature importance improves interpretability. Additionally, risk scores can assist clinicians in prioritizing patients for further evaluation.

5.6 Limitations

The system is trained on synthetic data, which may not fully represent real clinical variability. External validation with real-world datasets is required. The model currently performs binary classification and covers only a limited number of autoimmune diseases. Regulatory approval is also required before clinical deployment.

6. CONCLUSION

This paper presented the AID-System, a multi-model machine learning framework designed for autoimmune disease risk assessment. Among the six evaluated models, LightGBM achieved the best performance, with a ROC-AUC of 0.903 and an accuracy of 82.9% on the test dataset. Feature importance analysis identified ESR, age, family history, and CRP as the most influential predictors, which aligns well with established clinical understanding. In addition, cross-validation results confirmed the model's reliability, with LightGBM demonstrating the lowest standard deviation among all models, indicating stable and consistent performance.

Although some existing studies report slightly higher accuracy for single-disease prediction, the proposed system offers the advantage of supporting multiple autoimmune conditions, making it more practical for real-world

applications. However, the current study is limited by the use of synthetic data, and further validation using real clinical datasets is necessary before deployment. Future work will focus on real-world validation, expanding disease coverage, and integrating the system with electronic health record platforms. Overall, the AID-System demonstrates that machine learning can effectively support early risk assessment with strong performance and interpretable outputs.

7. FUTURE WORK

Future work will focus on improving the clinical applicability and robustness of the AID-System. A key priority is real-world validation using electronic health record data from multiple healthcare centers to ensure generalizability across diverse populations. The system will be expanded to include additional autoimmune diseases beyond the current scope, along with hierarchical models for more detailed diagnosis. Integration with electronic health record systems using HL7/FHIR standards will enable seamless deployment in clinical workflows and reduce manual data entry. Future enhancements will also include longitudinal modeling to track disease progression and predict flare-ups over time. Additionally, research will explore treatment response prediction to support personalized medicine. Patient-facing mobile applications and wearable device integration will enable continuous monitoring and early detection. Advanced machine learning techniques, including multimodal learning and federated learning, will be investigated to improve performance while maintaining data privacy. Finally, regulatory validation and bias assessment will be essential for safe and ethical deployment.

REFERENCES

- [1] M. G. Danieli, S. Brunetto, L. Gammeri, D. Palmeri, I. Claudi, Y. Shoenfeld, and S. Gangemi, "Machine learning application in autoimmune diseases: State of art and future perspectives," *Autoimmunity Reviews*, vol. 23, no. 2, p. 103496, 2023.
- [2] C. Adamichou, D. Nikolopoulos, I. Genitsaridi, et al., "Lupus or not? SLE Risk Probability Index (SLERPI): a simple, clinician-friendly machine learning-based model to assist the diagnosis of systemic lupus erythematosus," *Annals of the Rheumatic Diseases*, vol. 80, no. 6, pp. 758–766, 2021.
- [3] B. Kegerreis, M. D. Catalina, P. Bachali, et al., "Machine learning approaches to predict lupus disease activity from gene expression data," *Scientific Reports*, vol. 9, no. 1, p. 9617, 2019.
- [4] F. Ceccarelli, M. Sciandrone, C. Perricone, et al., "Prediction of chronic damage in systemic lupus erythematosus by using machine-learning models," *PLoS ONE*, vol. 12, no. 3, p. e0174200, 2017.
- [5] W. Wu, X. Hu, L. Yan, et al., "Development and validation of a cost-effective machine learning model for screening potential rheumatoid arthritis in primary healthcare clinics," *Journal of Inflammation Research*, vol. 18, pp. 1511–1522, 2025.
- [6] B. Norgeot, B. S. Glicksberg, and A. J. Butte, "A call for deep-learning healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 14–15, 2019.
- [7] L. Gossec, A. Molto, M. de Wit, et al., "Flare prediction in rheumatoid arthritis and axial spondyloarthritis using machine learning from continuous physical activity monitoring," *RMD Open*, vol. 8, no. 1, p. e002147, 2022.
- [8] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [9] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [10] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied Logistic Regression*, 3rd ed. Hoboken, NJ, USA: Wiley, 2013.
- [11] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.
- [12] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, 2020.
- [13] D. Aletaha, T. Neogi, A. J. Silman, et al., "2010 rheumatoid arthritis classification criteria," *Arthritis & Rheumatism*, vol. 62, no. 9, pp. 2569–2581, 2010.
- [14] M. Aringer, K. Costenbader, D. Daikh, et al., "2019 EULAR/ACR classification criteria for systemic lupus erythematosus," *Arthritis & Rheumatology*, vol. 71, no. 9, pp. 1400–1412, 2019.
- [15] C. H. Shiboski, S. C. Shiboski, R. Seror, et al., "2016 ACR/EULAR classification criteria for primary Sjögren's syndrome," *Annals of the Rheumatic Diseases*, vol. 76, no. 1, pp. 9–16, 2017.
- [16] A. J. Thompson, B. L. Banwell, F. Barkhof, et al., "Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria," *The Lancet Neurology*, vol. 17, no. 2, pp. 162–173, 2018.
- [17] American Thyroid Association, "Guidelines for Hashimoto's thyroiditis," *Thyroid*, vol. 22, no. 12, pp. 1200–1235, 2012.
- [18] J. S. Smolen, D. Aletaha, and I. B. McInnes, "Rheumatoid arthritis," *The Lancet*, vol. 388, no. 10055, pp. 2023–2038, 2016.
- [19] G. S. Firestein and I. B. McInnes, "Immunopathogenesis of rheumatoid arthritis," *Immunity*, vol. 46, no. 2, pp. 183–196, 2017.
- [20] World Health Organization, "Global burden of autoimmune diseases," *WHO Technical Report Series*, Geneva, Switzerland, 2023.

- [22] A. Esteva, A. Robicquet, B. Ramsundar, et al., "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019.
- [23] Z. Obermeyer and E. J. Emanuel, "Predicting the future - big data and clinical medicine," *New England Journal of Medicine*, vol. 375, no. 13, pp. 1216–1219, 2016.
- [24] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *NeurIPS*, 2017, pp. 4765–4774.
- [25] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD*, 2016, pp. 785–794.
- [26] G. Ke, Q. Meng, T. Finley, et al., "LightGBM: A highly efficient gradient boosting decision tree," in *NeurIPS*, 2017, pp. 3146–3154.
- [27] E. J. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York, NY, USA: Basic Books, 2019.
- [28] A. L. Beam and I. S. Kohane, "Big data and machine learning in health care," *JAMA*, vol. 319, no. 13, pp. 1317–1318, 2018.
- [29] HL7 International, "FHIR Release 4 Specification," Ann Arbor, MI, USA, 2019.
- [30] U.S. Food and Drug Administration, "Artificial intelligence and machine learning in software as a medical device," FDA, Silver Spring, MD, USA, 2022.
- [31] N. Conrad, S. Misra, J. Y. Verbakel, et al., "Incidence, prevalence, and co-occurrence of autoimmune disorders," *The Lancet*, vol. 401, no. 10391, pp. 1878–1890, 2023.