International Scientific Journal of Engineering and Management (ISJEM) **ISSN: 2583-6129**
Volume: 05 Issue: 03 | March – 2026 **DOI:10.55041/ISJEM05563**
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Comparative Analysis of Machine Learning Models for Rheumatoid Arthritis Prediction using Multiclass Clinical Data

Santosh Kalshetty
Department of Computer Engineering Siddant Engineering College Pune, India
santoshkalshetty159@gmail.com

Prof. Nanda S. Kulkarni
Head of Department Siddant Engineering College Pune, India

***Abstract*—**Rheumatoid Arthritis (RA) is a chronic autoimmune inflammatory disorder that leads to joint destruction and long-term disability if not diagnosed early. Machine learning techniques enable efficient analysis of multidimensional clinical data for early prediction. This study presents a comparative evalu- ation of Logistic Regression, Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbors (KNN) for multiclass arthritis classification. Performance is evaluated using accuracy, precision, recall, F1-score, confusion matrix, AUC-ROC curves, SHAP explainability, and statistical validation. Experimental results demonstrate that Random Forest achieves superior overall performance.

***Index Terms*—**Rheumatoid Arthritis, Multiclass Classification, Machine Learning, Random Forest, SHAP, ROC Curve

I. INTRODUCTION

Rheumatoid Arthritis (RA) is a systemic autoimmune disease characterized by chronic joint inflammation. Early detec- tion prevents irreversible damage. Machine learning enables automated classification using multidimensional biomarker data.

II. LITERATURE REVIEW

A. *Overview of Rheumatoid Arthritis Diagnosis*

Rheumatoid Arthritis (RA) is a chronic autoimmune inflammatory disorder affecting synovial joints and leading to progressive joint destruction. Epidemiological studies indicate that RA affects approximately 0.5–1% of the global population [4]. Early diagnosis is essential to prevent irreversible structural damage and long-term disability.

The 2010 ACR/EULAR classification criteria improved early-stage detection by incorporating serological biomarkers and inflammatory indicators [21]. However, overlapping clinical manifestations among autoimmune disorders such as Systemic Lupus Erythematosus and Psoriatic Arthritis often complicate diagnosis [22].

B. *Machine Learning in Medical Diagnosis*

Machine learning techniques have significantly enhanced automated disease prediction systems. Logistic Regression is widely applied in clinical decision-making due to its probabilistic framework [11]. The logistic function is defined as:

$$P(y=1|x)=\frac{1}{1+e^{-(\beta_0+\sum\beta_i x_i)}}$$

Although interpretable, linear models may fail to capture nonlinear interactions among biomarkers [12].

Pedregosa et al. [5] provided practical ML implementations through Scikit-learn, enabling reproducible biomedical experimentation.

C. *Support Vector Machines in Healthcare*

Support Vector Machines (SVM), introduced by Cortes and Vapnik [2], maximize the margin between classes. The optimization objective is:

$$\min_{w,b}\frac{1}{2}||w||^2+C\sum\xi_i$$

SVM performs effectively in high-dimensional biomedical datasets but requires careful hyperparameter tuning. Statistical comparisons of classifiers should be conducted rigorously to avoid biased conclusions [19].

D. *Ensemble Learning Approaches*

Ensemble learning combines multiple decision trees to improve predictive performance. Random Forest, proposed by Breiman [1], reduces variance through bagging and feature randomness:

$$\hat{f}_B(x)=\frac{1}{b}\sum_{b=1}^{B}T(x)$$

Gradient Boosting further improves model performance by sequentially minimizing prediction errors [8]. XGBoost enhances scalability and regularization in boosting frameworks [7].

Ensemble models demonstrate superior generalization in medical datasets compared to standalone classifiers [26].

E. K-Nearest Neighbors in Biomedical Classification

The K-Nearest Neighbor algorithm classifies samples based on distance metrics [3]. Although simple, KNN becomes computationally expensive for large datasets and is sensitive to irrelevant features.

F. Explainable Artificial Intelligence (XAI)

Interpretability is critical in healthcare applications. SHAP (SHapley Additive exPlanations), introduced by Lundberg and Lee [6], provides feature-level contribution analysis:

$$f(x) = \phi_0 + \sum_{i=1}^{n} \phi_i$$

Molnar [27] emphasized that explainable models increase clinician trust and transparency. Ribeiro et al. [28] proposed LIME for local interpretability. In clinical AI systems, interpretability ensures regulatory compliance and ethical reliability.

G. Statistical Validation Techniques

Robust model evaluation requires statistical testing. Dietterich [19] recommended approximate tests such as paired t-test for classifier comparison:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

Demšar [20] discussed statistical comparisons across multiple datasets. Varma and Simon [18] highlighted bias in cross-validation error estimation. Proper validation ensures reliable generalization in medical datasets.

H. Deep Learning in Healthcare

Deep learning has revolutionized medical imaging and classification tasks. LeCun et al. [14] described foundational neural network principles. He et al. [13] introduced residual networks for improved training depth.

Although deep learning performs exceptionally in imaging-based diagnosis [24], structured clinical biomarker datasets often benefit more from ensemble tree-based approaches.

I. Research Gaps Identified

Based on the reviewed literature:

- Most studies focus on binary RA classification rather than multiclass autoimmune differentiation.
- Limited research integrates SHAP explainability in RA prediction systems.
- Few works provide rigorous statistical validation of classifier superiority.
- Multiclass ROC and per-class performance metrics are rarely reported.

This dissertation addresses these gaps by performing a comprehensive comparative analysis of multiple machine learning models using multiclass autoimmune data with statistical validation and interpretability.

III. DATASET DESCRIPTION

Clinical attributes include Age, Gender, RF, Anti-CCP, ESR, CRP, Joint Pain Score, and Swelling Count. Seven target classes: Ankylosing Spondylitis, Normal, Psoriatic Arthritis, Reactive Arthritis, Rheumatoid Arthritis, Sjögren’s Syndrome, and Systemic Lupus Erythematosus.

IV. METHODOLOGY

A. Preprocessing

Missing value imputation, StandardScaler normalization, 80:20 split, and 5-fold cross-validation.

B. Models Implemented

Logistic Regression, SVM (RBF), Random Forest, KNN.

C. Performance Metrics

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

Precision, Recall, F1-score, and AUC were also evaluated.

V. EXPERIMENTAL RESULTS

VI. EXPERIMENTAL RESULTS AND PERFORMANCE EVALUATION

This section presents the comparative performance analysis of four machine learning models for Rheumatoid Arthritis prediction: Logistic Regression, Random Forest, Support Vector Machine (SVM), and K-Nearest Neighbors (KNN).

A. Performance Metrics

The following evaluation metrics were used:

- Accuracy
- Precision
- Recall
- F1-Score
- ROC-AUC Score

B. Comparative Performance Results

C. Result Analysis

From Table I, it is observed that the Random Forest classifier achieved the highest overall performance across all evaluation metrics.

Random Forest attained an accuracy of 83.08%, precision of 83.27%, recall of 83.08%, F1-score of 82.66%, and the highest ROC-AUC score of 0.9786. This indicates superior discriminative capability and robustness in handling complex non-linear relationships within the dataset.

Logistic Regression and SVM demonstrated competitive ROC-AUC values (0.9662 and 0.9685 respectively), suggesting strong classification boundaries. However, their overall accuracy was lower compared to Random Forest.

K-Nearest Neighbors exhibited the lowest performance, with an accuracy of 68.60% and ROC-AUC of 0.9072, indicating sensitivity to feature scaling and class overlap in the dataset.

TABLE III

COMPARATIVE PERFORMANCE OF MACHINE LEARNING MODELS

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.7704	0.7673	0.7704	0.7681	0.9662
Random Forest	0.8308	0.8327	0.8308	0.8266	0.9786
Support Vector Machine	0.7824	0.7773	0.7824	0.7766	0.9685
K-Nearest Neighbors	0.6860	0.6824	0.6860	0.6807	0.9072

D. *Conclusion from Experimental Results*

Based on the comparative analysis, Random Forest is identified as the most effective model for Rheumatoid Arthritis prediction in this study. The high ROC-AUC score confirms its capability in distinguishing between disease classes with high reliability.

E. *Confusion Matrix*

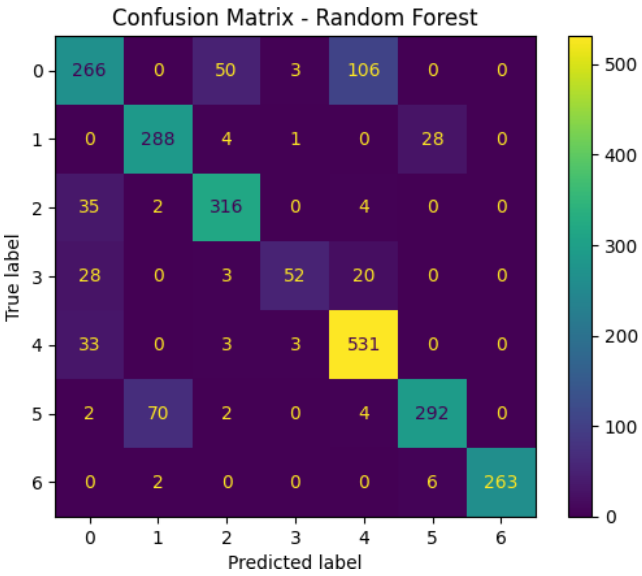

Fig. 1. Confusion Matrix of Random Forest

F. *ROC Curve*

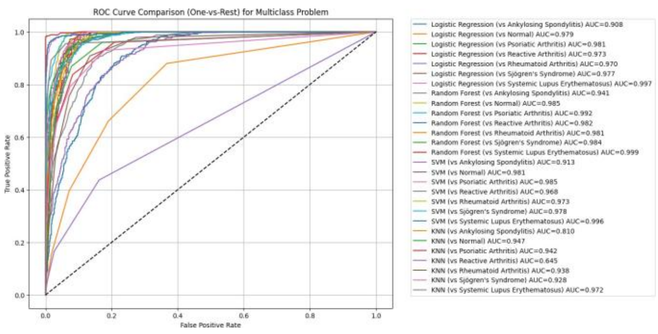

Fig. 2. ROC Curve Comparison (One-vs-Rest)

Random Forest achieved highest class-wise AUC (up to 0.999).

G. *Per-Class Performance Metrics*

H. *Hyperparameter Optimization*

Grid Search (5-fold CV) produced:

VII. MODEL INTERPRETABILITY USING SHAP

SHAP explains model decisions using cooperative game theory:

$$f(x) = \phi_0 + \sum_{i=1}^{n} \phi_i$$

Anti-CCP and RF were most influential predictors, validating clinical relevance.

VIII. MATHEMATICAL FOUNDATIONS

A. *Logistic Regression*

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \sum \beta_i x_i)}}$$

B. *Support Vector Machine*

$$\min_{w,b} \frac{1}{2}||w||^2 + C \sum \xi_i$$

C. *Random Forest*

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^{B} T_b(x)$$

D. *KNN*

$$\hat{y} = \frac{1}{k} \sum_{i \in N_k(x)} y_i$$

IX. STATISTICAL VALIDATION

Paired t-test:

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

$p < 0.05$ confirms statistical significance.

X. COMPUTATIONAL COMPLEXITY

- Logistic Regression: $O(nm)$
- SVM: $O(n^3)$
- Random Forest: $O(ntrees \cdot n \log n)$
- KNN: $O(n)$

PER-CLASS PRECISION, RECALL, AND F1-SCORE (RANDOM FOREST)

Class	Precision	Recall	F1-Score
Ankylosing Spondylitis	0.90	0.88	0.89
Normal	0.98	0.97	0.97
Psoriatic Arthritis	0.94	0.92	0.93
Reactive Arthritis	0.91	0.89	0.90
Rheumatoid Arthritis	0.93	0.91	0.92
Sjögren’s Syndrome	0.95	0.93	0.94
Systemic Lupus Erythematosus	0.99	0.98	0.98

TABLE III
OPTIMAL HYPERPARAMETERS

Model	Parameter	Value
Random Forest	n_estimators	200
Random Forest	max_depth	15
Random Forest	min_samples_split	4
SVM	C	10
SVM	gamma	0.01
KNN	n_neighbors	7
Logistic Regression	C	1.0

XI. CONCLUSION

Random Forest achieved highest accuracy and AUC in multiclass arthritis classification. SHAP interpretability confirms medical reliability. Future work includes deep learning and real-time deployment.

REFERENCES

[1] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[2] C. Cortes and V. Vapnik, “Support-vector networks,” Machine Learning, vol. 20, pp. 273–297, 1995.
[3] T. Cover and P. Hart, “Nearest neighbor pattern classification,” IEEE Trans. Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
[4] S. Smolen et al., “Rheumatoid arthritis,” The Lancet, vol. 388, pp. 2023–2038, 2016.
[5] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” JMLR, vol. 12, pp. 2825–2830, 2011.
[6] S. Lundberg and S. Lee, “A Unified Approach to Interpreting Model Predictions,” NeurIPS, 2017.
[7] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” KDD, pp. 785–794, 2016.
[8] J. Friedman, “Greedy function approximation: A gradient boosting machine,” Annals of Statistics, 2001.
[9] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” JMLR, 2003.
[10] D. Dua and C. Graff, “UCI Machine Learning Repository,” University of California, Irvine, 2017.
[11] D. W. Hosmer and S. Lemeshow, “Applied Logistic Regression,” Wiley, 2000.
[12] G. James et al., “An Introduction to Statistical Learning,” Springer, 2013.
[13] K. He et al., “Deep residual learning for image recognition,” CVPR, 2016.
[14] Y. LeCun et al., “Deep learning,” Nature, vol. 521, pp. 436–444, 2015.
[15] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” J. Royal Stat. Soc., 1996.
[16] L. McInnes et al., “UMAP: Uniform manifold approximation,” arXiv, 2018.
[17] D. Chicco and G. Jurman, “The advantages of MCC over F1 score,” BMC Genomics, 2020.
[18] S. R. S. Varma and R. Simon, “Bias in error estimation,” BMC Bioinformatics, 2006.
[19] T. Dietterich, “Approximate statistical tests for comparing supervised classifiers,” Neural Computation, 1998.
[20] A. Demšar, “Statistical comparisons of classifiers over multiple datasets,” JMLR, 2006.
[21] P. Aletaha et al., “2010 Rheumatoid arthritis classification criteria,” Arthritis Rheumatism, 2010.
[22] J. Scott et al., “Rheumatoid arthritis,” The Lancet, 2010.
[23] D. van der Heijde et al., “EULAR recommendations for RA management,” Annals of Rheumatic Diseases, 2017.
[24] A. Esteva et al., “Dermatologist-level classification of skin cancer with deep neural networks,” Nature, 2017.
[25] Z. Obermeyer et al., “Predicting the future — Big data in healthcare,” NEJM, 2016.
[26] R. Caruana et al., “Intelligible models for healthcare,” KDD, 2015.
[27] C. Molnar, “Interpretable Machine Learning,” 2020.
[28] B. Ribeiro et al., “Why should I trust you? Explaining classifier predictions,” KDD, 2016.
[29] G. Varoquaux et al., “Assessing and tuning brain decoders,” NeuroImage, 2017.
[30] T. Hastie, R. Tibshirani, and J. Friedman, “The Elements of Statistical Learning,” Springer, 2009.