International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0343

e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2383 - 2385

An Intelligent Multi-Layered Statistical Prediction and Decision-Support Framework for Advanced Healthcare Data Analysis Using Hybrid Artificial Intelligence, Deep Explainable Analytics, Predictive Modeling, And Interpretable Machine Learning Techniques

Adeline Joy[1]
[1]*Department of Mathematics, Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India.*
Email: *adeline@msec.edu.in*[1]

Abstract

The present study examines the impact of Artificial Intelligence (AI) on Human Resource (HR) decision-making and employee experience in modern organisations. The rapid integration of AI technologies into HR functions has significantly transformed traditional practices such as recruitment and selection, employee engagement, performance appraisal, training and development, workforce planning, and talent management. AI-powered tools and analytics enable organisations to process large volumes of employee data efficiently, thereby improving the quality, speed, and accuracy of managerial decision-making. The study further explores how AI contributes to enhanced employee experience by offering personalised learning opportunities, real-time feedback systems, predictive performance analysis, and improved communication mechanisms within organisations. The research is primarily based on secondary data collected from journals, research articles, industry reports, and existing literature related to AI and HR management. Observational analysis has also been incorporated to understand the practical implications of AI adoption in organisational settings. The findings reveal that AI-driven HR systems help reduce repetitive administrative tasks, minimise human error, improve operational efficiency, and support strategic HR planning. In addition, AI enhances employee engagement and satisfaction by creating more responsive and employee-centric work environments. However, the study also identifies several challenges associated with AI implementation, including algorithmic bias, data privacy concerns, ethical issues, lack of transparency, and employee resistance arising from fear of job displacement. The study concludes that a balanced integration of AI technologies with human judgement and ethical considerations can significantly improve HR effectiveness, employee trust, and sustainable organisational growth in the long run.

Keywords: *Artificial Intelligence, HR Decision-Making, Employee Experience, Employee Engagement, HR Analytics, Organizational Effectiveness*

1. Introduction

Healthcare institutions generate large volumes of patient data through medical imaging, electronic health records, wearable devices, and clinical monitoring systems. Traditional methods of healthcare analysis are often insufficient for extracting meaningful insights from such complex datasets. Statistical prediction models combined with artificial intelligence techniques can significantly improve clinical decision-making and disease diagnosis. The integration of machine learning with classical statistical methods has emerged as a promising solution for managing the complexity and heterogeneity of healthcare data. This paper proposes a hybrid analytical framework that leverages the interpretability of statistical models with the predictive power of ensemble learning methods. Explainable AI (XAI) components are incorporated to ensure that the predictions generated are understandable by medical practitioners, thereby supporting trust in automated decision support systems. The remainder of this paper is organized as follows: Section 2 reviews related literature; Section 3 presents the proposed methodology; Section 4 details the mathematical model; Section 5 discusses experimental results; and Section 6 concludes the paper.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2026.0343
e ISSN: 2584-2854
Volume: 04 Issue: 06
June 2026
Page No: 2383 - 2385

2. Literature Survey

A substantial body of research has investigated statistical and machine learning approaches for healthcare data analysis. Han and Kamber [1] established foundational concepts in data mining that underpin many current clinical decision support techniques. Their framework for pattern recognition and predictive modeling has been widely adopted in biomedical research. Hastie et al. [2] provided comprehensive coverage of statistical learning methods including regularized regression, ensemble models, and model selection techniques. These methods form the backbone of the proposed hybrid framework. Their treatment of LASSO regularization in particular is highly relevant to feature selection in high-dimensional clinical datasets. The advent of deep learning has further expanded capabilities for complex medical data. Goodfellow et al. [3] demonstrated the utility of deep neural architectures for feature extraction and classification tasks. However, the interpretability deficit of deep learning models remains a significant barrier to clinical adoption. Lundberg and Lee [4] addressed this gap by introducing SHAP (SHapley Additive exPlanations), a unified framework for interpreting machine learning predictions. This technique assigns feature-level contribution scores to each prediction, enabling clinicians to identify which variables most strongly influenced a given outcome. The integration of SHAP values into clinical prediction workflows has been shown to improve physician trust and decision quality.

3. Proposed Methodology

The proposed framework is a multi-stage pipeline that integrates statistical modeling, machine learning, and explainability components. The methodology is designed to process structured clinical data including patient demographic details, laboratory reports, medical history, and physiological indicators.

3.1. Data Preprocessing

Anonymized healthcare datasets are preprocessed to handle missing values through multiple imputation techniques. Categorical variables are encoded using one-hot encoding, and continuous features are normalized using z-score standardization to ensure numerical stability during model training.

3.2. Feature Selection

Feature selection is performed using LASSO (Least Absolute Shrinkage and Selection Operator) regularization. This technique introduces an L1 penalty term into the regression objective, effectively shrinking irrelevant feature coefficients to zero. Correlation-based filtering is additionally applied to remove redundant features, thereby reducing dimensionality and improving model generalizability

3.3. Hybrid Prediction Model

The core prediction framework combines logistic regression for binary classification tasks with decision tree and random forest ensemble models to capture non-linear relationships. The outputs of these models are fused using a weighted voting ensemble strategy that optimizes predictive performance across the target health outcomes: hospital readmission, diabetes progression, and cardiovascular disease risk.

3.4. Explainability with SHAP

SHAP-based explainability is integrated as a post-hoc interpretability layer. For each patient prediction, SHAP values are computed to quantify the marginal contribution of each feature. Visualizations including summary plots and force diagrams are generated to aid clinical interpretation, facilitating transparency and supporting regulatory compliance.

4. Mathematical Model

The logistic regression model used for binary classification is expressed as:

$$P(Y = 1) = 1 / (1 + e\text{-}(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)) \quad (1)$$

where β represents the regression coefficient vector and x represents the input feature vector. The objective function is optimized using maximum likelihood estimation (MLE).

The log-likelihood function L(β) is defined as:

$$L(\beta) = \Sigma \, [y \, i \log Pi + (1 - yi) \log(1 - Pi)] \quad (2)$$

For the LASSO regularization during feature selection, the penalized objective function is given by:

$$\text{minimize } L(\beta) + \lambda \, ||\beta||1 \quad (3)$$

where λ is the regularization hyperparameter controlling the degree of sparsity in the feature coefficient vector. Higher values of λ produce sparser models by driving more coefficients to zero.

5. Results And Discussion

Experimental analysis was conducted on three anonymized healthcare datasets corresponding to hospital readmission, diabetes progression, and cardiovascular disease prediction tasks. All datasets were split into training (80%) and testing (20%) subsets using stratified sampling to preserve class distribution. The proposed hybrid model achieved an overall classification accuracy of 92%, outperforming standalone logistic regression (84%) and decision tree classifiers (87%). The AUC-ROC score was 0.91, indicating strong discriminative performance across all three clinical prediction tasks. Shown as Table 1 Performance Comparison of Prediction Models

Table 1 Performance Comparison of Prediction Models

Model	Accuracy	Precision	Recall	Auc-Roc
Logistic Regression	84%	0.82	0.80	0.83
Decision Tree	87%	0.85	0.84	0.86
Random Forest	89%	0.88	0.87	0.89
Proposed Hybrid	92%	0.91	0.90	0.91

SHAP analysis successfully identified the most influential clinical features for each prediction task. For cardiovascular disease risk, blood pressure, LDL cholesterol level, and age were the top-ranked predictors. For diabetes progression, fasting glucose, BMI, and HbA1c were found to be the most significant indicators. For hospital readmission, prior admission frequency, comorbidity count, and discharge disposition were the dominant variables. The false-positive rate of the proposed model was reduced by 18% compared to conventional logistic regression, a clinically significant improvement that directly reduces unnecessary interventions and associated costs.

Conclusion

This paper presented a hybrid healthcare prediction framework that integrates logistic regression, decision tree analysis, random forest ensemble learning, and SHAP-based explainability for clinical decision support. The proposed model achieves superior predictive accuracy (92%) and AUC-ROC (0.91) compared to conventional statistical baselines, while maintaining transparency through feature-level explanations. The framework is scalable and applicable across multiple disease prediction tasks. Future work will explore deep learning integration for unstructured clinical text and imaging data, as well as real-time deployment via IoT-based platforms.

Acknowledgements

The authors wish to express their gratitude to Meenakshi Sundararajan Engineering College for supporting this research and providing the necessary computational facilities.

References

[1]. J. Han and M. Kamber, Data Mining: Concepts and Techniques, 3rd ed. Burlington, MA: Morgan Kaufmann, 2021.
[2]. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY: Springer, 2019.
[3]. I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2018.
[4]. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Advances in Neural Information Processing Systems (NIPS), vol. 30, pp. 4765–4774, 2017.