International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

HealthGuard AI: A Multi-Stage Machine Learning Framework for Personalized Disease Risk Stratification and Adaptive Health Recommendation

Ashwani Kumar[1], Dr. Sunil Maggu[2]
[1,2]Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi

Abstract- — The intersection of machine learning and preventive healthcare offers transformative potential for earlydisease detection and personalized health guidance. However, most existing systems either producebinary classification outcomes without contextual risk stratification or provide static, non-adaptivehealthrecommendations disconnected from individual prediction confidence scores. This paper introducesHealthGuard AI, a novel multi-stage predictive framework that integrates supervised machinelearningclassification with probability-based risk stratification and a dynamically adaptivehealthrecommendation engine. The system simultaneously addresses three major chronic diseasedomains—Type 2 Diabetes Mellitus, Coronary Heart Disease, and Parkinson's Disease — using clinicallyvalidatedfeature sets drawn from UCI Machine Learning Repository datasets. Beyond binary prediction, HealthGuard AI applies predict_proba() outputs to stratify individual disease risk into Low(≤0.40), Medium (0.41–0.70), and High (> 0.70) categories, each triggering a distinct, evidence-alignedhealthrecommendation profile. An additional Body Mass Index and Basal Metabolic Rate estimationmoduleemploying the Mifflin-St Jeor equation further extends the system's scope into nutritional healthanalytics. Deployed as an interactive web application via Streamlit Community Cloud, HealthGuardAIachieves classification accuracies of 78.5%, 81.3%, and 87.2% for Diabetes, Heart Disease, andParkinson's Disease respectively. The system demonstrates that probability-aware risk stratification, when combined with adaptive, risk-tiered recommendations, produces a meaningfully richer andmoreclinically actionable output than conventional binary prediction pipelines. Experimental results, systemarchitecture, and the clinical relevance of risk-tiered adaptive recommendations are discussedindetail.

Keywords: Disease Risk Stratification, Machine Learning, Predict_proba, Personalized Health Recommendations, Diabetes Prediction, Heart Disease Prediction, Parkinson's Disease, BMI Analysis, Streamlit, Preventive HealthcareAI

I. INTRODUCTION

Chronic non-communicable diseases (NCDs) represent the leading cause of global mortality, accounting for approximately 74% of all deaths worldwide according to WHO estimates. Early detection and timely behavioral intervention are universally recognized as the most cost-effective strategies for reducing the burden of NCDs. Despite significant advancesinclinicaldiagnostics, a critical accessibility gappersists:the majority of at-risk individuals — particularlyin low-income and rural settings —lackaccessto routine screening, specialist consultations, or evidence-based lifestyle guidance.

Artificial Intelligence and Machine Learning havememerged as powerful tools for bridging this gap. Supervised learning models trained on clinical datasets can achieve diagnostic accuracy comparable to specialist-level screening for conditions such as diabetes and cardiovascular disease. However, existing ML-based health screening tools predominantly operate as binary classifiers, producing a positive or negative outcome without communicating the gradient of risk that separates a patient at 55% probability from one at 92% probability. This information asymmetry limits the clinical utility of such systems.

HealthGuard AI addresses this limitation hrough a fundamentally different design philosophy: rather than treating prediction as a terminal step, it treats prediction probability as the primary input to a downstream risk stratification and recommendation pipeline. The core innovation of this work is the systematic application of probability calibration outputs (predict_proba()) to drive personalized, risk- tiered health recommendations — a design that more closely mirrors the graduated clinical decision-making process used by healthcare professionals.

This paper makes the following contributions:

- A multi-disease ML prediction framework covering Diabetes, Heart Disease, and Parkinson's Disease using clinically validated feature sets

- A novel probability-based three-tier risk stratification system (Low / Medium / High) with defined probability thresholds
- An adaptive health recommendation engine that generates risk-level-specific, evidence-aligned preventive guidance
- An integrated BMI and caloric estimation module using the Mifflin-St Jeor BMR formula for nutritional health analysis
- A deployable, web-accessible implementation via Streamlit Community Cloud

II. RELATED WORK

Machine learning applications in healthcarehavebeen extensively explored acrossmultipledisease domains. Kavakiotis et al. (2017)conducted a comprehensive survey of MLanddata mining methods in diabetesresearch, identifying Random Forest and Support VectorMachines as consistently high-performingalgorithms.

Their analysis notedthat mostsystems reported binary classificationmetricswithout addressing the clinical significanceofprediction confidence gradients. For cardiovascular disease prediction, multiplestudies including Detrano et al. andsubsequentML-based analyses have demonstratedthat the Cleveland Heart Disease Dataset supportsreliable binary classification withlogisticregression achieving accuracies inthe80–85%range.

A key limitation identified acrossthesestudies is the absence of actionableoutput—models predict but do not recommend. In the domain of Parkinson's Disease, voicebiomarker-based detection systemspioneeredby Little et al. (2009) demonstratedthat MDVP(Multi-Dimensional Voice Program) parameters, particularly MDVP:Fo, jitter, shimmer, andNHR, carry strong discriminative power. Subsequentensemble learning approaches improveduponthese results, with decision tree andrandomforest models consistently outperforminglinearclassifiers on this feature set.

The concept of probability-basedriskstratification in clinical decisionsupportsystems (CDSS) has been exploredinspecialized contexts. Frizzell et al. (2017) applied gradient boosting to stratify heart failurerisk, demonstrating the clinical valueofgraduated risk outputs. However, tothebestofour knowledge, no prior publicly accessible, deployable system has applied predict_proba()outputs as the driver of an adaptive, multi-tieredrecommendation engine spanningmultiplechronic disease domains simultaneously.

Existing web-based health screening tools such as IBM Watson Health and Google's various health initiatives operate on proprietary datasets and infrastructure, limiting their accessibility and reproducibility. HealthGuard AI distinguishes itself through its open-source foundation, deployment accessibility, and the novelty of its probability-to-recommendation pipeline.

III. SYSTEM ARCHITECTURE

Overview

HealthGuard AI is structured as a five-stage processing pipeline, as illustrated in Fig. 1. The pipeline consists of: (1) Structured User Input Acquisition, (2) ML-Based Disease Classification, (3) Probability Score Extraction, (4) Three-Tier Risk Stratification, and (5) Adaptive Recommendation Generation. Each stage operates independently, enabling modular updates without disrupting the overall system.

Fig. 1: Five-Stage Pipeline Architecture of HealthGuard AI

Disease Prediction Module

Three independent supervised classification models handle disease prediction:

Diabetes Module: Random Forest Classifier trained on the Pima Indians Diabetes Dataset (768 samples, 8 features). Features include Glucose Level, Blood Pressure, Skin Thickness, Insulin Level, BMI, Diabetes Pedigree Function, Age, and Number of Pregnancies.

Heart Disease Module: Logistic Regression trained on the Cleveland Heart Disease Dataset (303 samples, 13 features). Features include Age, Gender, Chest Pain Type, Resting Blood Pressure, Serum Cholesterol, Fasting Blood Sugar, Resting ECG, Maximum Heart Rate, ExerciseInduced Angina, ST Depression (oldpeak), ST Slope, Major Vessels, and Thalassemia type.

Parkinson's Disease Module: DecisionTreeClassifier trained on the UCI Parkinson'sDataset (195 samples, 22 voicebiomarkerfeatures) including MDVP:Fo(Hz), MDVP:Fhi(Hz), MDVP:Flo(Hz), MDVP:Jitter(%), MDVP:Jitter(Abs), MDVP:RAP, MDVP:PPQ, Jitter:DDP, MDVP:Shimmer, Shimmer:APQ3, Shimmer:APQ5, MDVP:APQ, Shimmer:DDA, NHR, HNR, RPDE, DFA, spread1, spread2, D2, and PPE.

Probability Extraction and Risk Stratification

The core innovation of HealthGuardAI liesinitsuse of the predict_proba() functionfromScikitlearn, which returns the posterior probabilityofclass membership for each input sample.

Forbinary classifiers, this yields a two-elementprobability vector [P(negative), P(positive)].HealthGuard AI extracts P(positive) —theprobability of disease presence —as theprimaryrisk metric. Based on clinical reasoning andprobabilitycalibration analysis, three risk thresholdsaredefined:

Risk Tier	Probability Range	Clinical Implication
Low Risk	$P \leq 0.40$	Minimal disease likelihood; preventive maintenance recommended
Medium Risk	$0.40 < P \leq 0.70$	Elevated concern; lifestyle intervention and medical review warranted
High Risk	$P > 0.70$	High disease likelihood; immediate medical consultation required

Table 1: HealthGuard AI Risk StratificationThresholds

These thresholds were selected basedonthebalance between sensitivity andspecificity optimization across all three models. The 0.70 upper threshold aligns with the positive predictive value benchmarks typically considered clinically significant in preliminary screening contexts.

Adaptive Recommendation Engine

The Adaptive Recommendation Engine (ARE) implements a risk-tier-driven evidence mapping system. Rather than applying uniform health advice irrespective of individual risk severity, the ARE selects from three distinct recommendation profiles:

- **High Risk Profile:** Recommendations focus on immediate healthcare system engagement. Specific guidance includes: immediate consultation with a specialist, strict adherence to prescribed medication protocols, complete tobacco and alcohol cessation, adoption of a low-glycemic, low- saturated-fat dietary pattern, and monthly health monitoring with BP and glucose tracking.

- **Medium Risk Profile:** Recommendations target behavioral modification before clinical intervention becomes necessary.Guidance includes: scheduling a physician appointment within 2-4 weeks, adopting a Mediterranean or DASH dietary pattern, achieving 150 minutes of moderate aerobic activity per week, regular home monitoring of BP and glucose levels, and gradual reduction of identified lifestyle risk factors.

- **Low Risk Profile:** Recommendations reinforce health maintenance behaviors. Guidance includes: sustaining current dietary and physical activity patterns, annual preventive health screening, mental wellness activities, adequate sleep hygiene, and health literacy enhancement.

BMI and Caloric Estimation Module

An additional module implements BMI calculation (weight [kg] / height² [m²]) anddailycaloric requirement estimation usingtheMifflin-St Jeor Basal Metabolic Rate (BMR) formula: Male: BMR = 10W+ 6.25H−5A+5Female: BMR = 10W+ 6.25H−5A−161Where W = weight (kg), H = height (cm), A=age(years). Total Daily Energy Expenditure(TDEE)iscalculated by multiplying BMR by avalidatedPhysical Activity Level (PAL) multiplier rangingfrom 1.2 (sedentary) to 1.9 (extremelyactive).Caloric targets are then derived as TDEE±500kcal for weight gain or loss respectively.

IV. DATA PREPROCESSING METHODOLOGY

1. **Missing Value Imputation:** Missing entries in the Diabetes dataset (represented as zero-values in physiologically implausible fields) were replaced using column-wise median imputation, preserving distributional characteristics while minimizing the impact of outliers.

2. **Feature Normalization:** MinMaxScaler normalization was applied to scale all numerical features to the [0, 1] range, ensuring that no single feature disproportionately influences distance- based or gradient-based learning mechanisms.

3. **Train-Test Partitioning:** A stratified 80:20 train-test split was implemented across all datasets to maintain class proportion consistency between training and evaluation sets.

4. **Class Imbalance Handling:** The Diabetes and Heart Disease datasets exhibit moderate class imbalance. SMOTE (Synthetic Minority Oversampling Technique) was applied to the training set only to prevent data leakage, generating synthetic minority class samples to improve model sensitivity for disease- positive cases.

5. **Feature Validation:** Correlation analysis and univariate feature importance scores were computed to validate the clinical relevance of input features and identify potential multicollinearity issues.

V.MODEL SELECTION AND TRAINING

Model selection was guided by three criteria:

1. support for predict_proba() output in Scikit-learn,
2. interpretability suitable for healthcare contexts, and
3. demonstrated performance on tabular clinical data. The following models were selected and trained:

Model	Disease	Key Hyperparameters	Rationale
Random Forest	Diabetes	n_estimators= 100, max_depth=None, random_state =42	Ensemble robustness; handles feature interactions; native predict_proba support
Logistic Regression	Heart Disease	C=1.0, solver=lbfgs, max_iter=1000	Probabilistically calibrated; interpretable coefficients; suitable for linearly separable cardiac features
Decision Tree	Parkinson's	criterion=gini, max_depth=5, random_state =42	Interpretable decision rules; low computational overhead; effective on MDVP voice biomarkers

Table 2: ML Model Selection and Hyperparameter Configuration

All models were trained using Scikit-learn v1.6.1 and serialized using Python's pickle module as
.sav files for persistent deployment within the Streamlit application environment. Model training was conducted in Jupyter Notebook environments with reproducibility ensured through fixed random seeds.

VI.EXPERIMENTAL RESULTS

Classification Performance

Table 3 presents the classification performance of each model evaluated on the held-out 20% test set:

Model	Disease	Accuracy	Precision	Recall	F1-Score	Auc-roc
Random Forest	Diabetes	78.5%	0.77	0.76	0.77	0.84
Logistic Regression	Heart Disease	81.3%	0.82	0.80	0.81	0.88
Decision Tree	Parkinson's	87.2%	0.88	0.86	0.87	0.91

Table 3: Classification Performance Metrics (80:20 Train-Test Split)

Parkinson's Disease achieved the highest classification accuracy (87.2%) attributable to the high discriminative power of MDVP voice biomarkers, particularly spread1, PPE, and RPDE, which demonstrate strong separation between Parkinson's-positive and negative cases. Heart Disease achieved 81.3% accuracy with Logistic Regression, consistent with established benchmarks on the Cleveland dataset. Diabetes classification at 78.5% reflects the inherent complexity of metabolic disease prediction from limited clinical parameters.

Risk Stratification Distribution Analysis

To evaluate the clinical utility of the risk stratification system, the distribution of risk tiers across the test sets was analyzed:

Table 4: Risk Tier Distribution Across Test Sets

Disease	Low Risk (%)	Medium Risk (%)	High Risk (%)
Diabetes	52.3	31.4	16.3
Heart Disease	44.7	33.8	21.5
Parkinson's Disease	38.5	27.9	33.6

The Parkinson's Disease module shows a higher proportion of High Risk classifications (33.6%), consistent with the clinical prevalence characteristics of the UCI dataset. The risk tier distributions demonstrate that the probability thresholds produce clinically meaningful stratification rather than arbitrary binning.

System Performance Benchmarking

Metric	Value
Average Prediction Latency	1.1 seconds (end-to-end)
Model Load Time	< 400 ms
Inference Time per Query	< 300 ms
Concurrent User Capacity (tested)	Up to 300 users
System Uptime (Streamlit Cloud)	99.5%
Browser Compatibility	Browser Compatibility

Table 5: System Performance Benchmarks

VI.COMPARATIVE ANALYSIS

Table 6 compares HealthGuard AI against existing systems and prior works on similar disease prediction tasks:

Table 6: Comparative Analysis with Existing Systems

System / Study	Diseases Covered	Risk Stratification	Recommendations	Public Access
IBM Watson Health	Multiple	Limited	Yes (Proprietary)	No
Google Health AI	Multiple	No	No	No
Kavakiotis et al. (2017)	Diabetes only	No	No	No
Little et al. (2009)	Parkinson's only	No	No	No
Outbreak Sense (2024)	3 diseases + Outbreak	No	No	Yes
HealthGuard AI (Proposed)	3 diseases + BMI	Yes (3-tier)	Yes (Adaptive)	Yes

HealthGuard AI is the only publicly deployable system among those reviewed that simultaneously provides multi-disease prediction, calibrated probability-based risk stratification, and adaptive tiered health recommendations across three chronic disease domains.

VII. DISCUSSION

Clinical Significance of Probability- Based Risk Stratification

The shift from binary prediction to probability- based risk stratification represents a clinically meaningful design evolution. In conventional binary systems, a patient with a 42% probability and a patient with a 95% probability both receive identical 'Positive' classifications and identical, if any, recommendations. HealthGuard AI's three- tier stratification system ensures that these fundamentally different risk profiles trigger appropriately differentiated responses — medium-tier lifestyle modification guidance versus high-tier immediate medical intervention.

This approach aligns with established principles in preventive medicine where risk communication is recognized as more effective when graduated rather than binary. Studies in health communication psychology consistently show that individuals are more likely to act on graduated risk information than on binary positive/negative diagnoses, particularly in the absence of immediate symptoms.

Limitations

- **Dataset Generalizability:** All three models are trained on publicly available datasets that may not fully represent the epidemiological characteristics of diverse real-world populations, particularly South Asian patient profiles for whom disease risk factor distributions may differ significantly from the UCI dataset compositions.
- **Static Model Deployment:** Models are not retrained dynamically as new data becomes available. Periodic retraining would be required to maintain calibration accuracy over time.
- **Probability Calibration:** While predict_proba() provides probability estimates, these are not formally calibrated using techniques such as Platt Scaling or Isotonic Regression. Formal calibration would improve the reliability of risk tier boundaries.
- **No Electronic Health Record Integration:** The current system operates independently of existing clinical systems, requiring manual data entry rather than automated parameter import from patient records.
- **Single-Session Statelessness:** The Streamlit deployment does not persist session data, preventing longitudinal risk trend analysis across multiple user interactions.

VIII. FUTURE DIRECTIONS

Based on the limitations identified and the broader research landscape in healthcare AI, the following future directions are proposed:

Probability Calibration Integration: Implement Platt Scaling and Isotonic Regression to formally calibrate model probability outputs, improving the reliability of risk tier assignments.

Federated Learning Architecture: Redesign the training pipeline to support federated learning, enabling model improvement from distributed patient data without centralizing sensitive health information, addressing both privacy concerns and dataset diversity limitations.

Explainable AI (XAI) Integration: Incorporate SHAP (SHapley Additive exPlanations) value computation to

provide per-prediction feature importance explanations, increasing transparency and clinician trust in model outputs.

Longitudinal Risk Tracking: Develop a user authentication and session persistence layer enabling tracking of individual risk trajectories over time, transforming HealthGuard AI from a single-prediction tool into a continuous health monitoring platform.

Expanded Disease Coverage: Extend the framework to cover chronic kidney disease, fatty liver disease, and breast cancer risk assessment, leveraging additional UCI and SEER datasets.

Indian Population Dataset Development: Collaborate with healthcare institutions to curate and publish a labeled dataset representative of Indian patient demographics, addressing the dataset generalizability limitation.

Real-Time Wearable Integration: Develop API interfaces for smartwatch and IoT health device data ingestion, enabling automated, continuous risk assessment from real-time physiological data streams.

IX.CONCLUSION

This paper presented HealthGuard AI, a multi- stage machine learning framework for personalized disease risk stratification and adaptive health recommendation. By extending beyond binary disease prediction to incorporate calibrated probability extraction, three-tier risk stratification, and risk-tiered adaptive recommendations, HealthGuard AI delivers a meaningfully richer and more clinically actionable output than conventional ML-based health screening systems.

The system achieves classification accuracies of 78.5%, 81.3%, and 87.2% for Diabetes, Heart Disease, and Parkinson's Disease respectively, with end-to-end prediction latency under 1.1 seconds and demonstrated scalability to 300 concurrent users. The integrated BMI and caloric estimation module further extends the system's scope into nutritional health analytics.

The fundamental contribution of this work lies not in algorithmic novelty but in the design philosophy of probability-aware, risk-stratified health recommendation — a design that more faithfully mirrors the graduated clinical decision- making process and produces outputs that are demonstrably more actionable than binary classifiers. The open-source, publicly deployable nature of HealthGuard AI ensures that its benefits are accessible regardless of geographic or economic barriers.

HealthGuard AI demonstrates that the combination of accessible machine learning tools, probability calibration, and evidence- aligned recommendation design can produce intelligent healthcare systems that are simultaneously accurate, interpretable, clinically meaningful, and publicly accessible.

Acknowledgements
The authors gratefully acknowledge the Department of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, for providing the academic infrastructure and guidance that supported this research. The UCI Machine Learning Repository is acknowledged for maintaining and providing public access to the healthcare datasets utilized in this study.

REFERENCES

1. Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine Learning and Data Mining Methods in Diabetes Research. Computational and Structural Biotechnology Journal, 15, 104–116. https: //doi.org/10.1016/j.csbj.2016.12.005
2. Little, M. A., McSharry, P. E., Roberts, S. J., Costello, D. A., & Moroz, I. M. (2009). Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection. BioMedical Engineering OnLine, 6(1), 23. https: //doi.org/10.1186/1475-925X-6-23
3. Detrano, R., Jánosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J.-J., Sandhu, S., Guppy, K. H., Lee, S., & Froelicher, V. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. American Journal of Cardiology, 64(5), 304–310.
4. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research, 16, 321–357.
5. Frizzell, J. D., Liang, L., Schulte, P. J., Yancy, C. W., Bhatt, D. L., Fonarow, G. C., ... & Laskey, W. K. (2017). Prediction of 30-day all-cause readmissions in patients hospitalized for heart failure. JAMA Cardiology, 2(2), 204–209.
6. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
7. Mifflin, M. D., St Jeor, S. T., Hill, L. A., Scott, B. J., Daugherty, S. A., & Koh, Y. O. (1990). A new predictive equation for resting energy expenditure in healthy individuals. The American Journal of Clinical Nutrition, 51(2), 241–247.

8. Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.
9. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why Should I Trust You?: Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
10. Dua, D., & Graff, C. (2019). UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences. http ://archive.ics.uci.edu/ml
11. WHO. (2023). Noncommunicable diseases: Key facts. World Health Organization. https ://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases
12. Streamlit Inc. (2023). Streamlit Documentation. https ://docs.streamlit.io
13. Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94–98.
14. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318.
15. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.