Indian Journal of Computer Science and Technology
https://www.doi.org/10.59256/indjcst.20260501023
Volume 5, Issue1 (January-April 2026), PP: 160-168.
www.indjcst.com
ISSN No: 2583-5300

An Explainable Deterministic Framework for Preventive Health Risk Stratification with Multilingual Decision Support for Low-Resource Environments

M.V. Karthikeya[1], Dr. T.V. Nagalakshmi[2], D. Nanda Kishore[3], Mourya Sagar[4]
[1,3,4] SRM Institute of Science and Technology, Andhra Pradesh, India.
[2]Department of Basic Engineering, DVR & Dr.HS MIC College of Technology, Kanchikacherla, NTR, Andhra Pradesh, India.

To Cite this Article: *M.V. Karthikeya[1], Dr. T.V. Nagalakshmi[2], D. Nanda Kishore[3], Mourya Sagar[4], "An Explainable Deterministic Framework for Preventive Health Risk Stratification with Multilingual Decision Support for Low-Resource Environments", Indian Journal of Computer Science and Technology, Volume 05, Issue 01 (January-April 2026), PP: 160-168.*

Abstract: *Preventive healthcare technologies play a crucial role in reducing the global burden of chronic diseases by enabling early awareness and timely lifestyle intervention. While contemporary health risk prediction systems frequently employ machine learning models, many such approaches operate as opaque black-box frameworks, limiting interpretability, reproducibility, and safe deployment in preventive public health contexts.*

This paper presents a formally defined deterministic preventive health risk stratification framework designed for multilingual and low-resource environments. The framework is grounded in clinically validated preventive health thresholds defined by the World Health Organization (WHO), American Diabetes Association (ADA), and American Heart Association (AHA). A formal deterministic risk function and algorithmic classification model are defined, analyzed for logical correctness, and evaluated for computational efficiency.

The proposed system integrates a modular system architecture consisting of a weighted risk engine, classification module, explanation generator, multilingual translation layer, deterministic chatbot engine, and voice-based output interface. The mathematical formulation defines a cumulative weighted risk function and deterministic classification boundaries. Computational complexity analysis demonstrates linear-time execution, making the system suitable for low-resource deployment.

Experimental validation using structured synthetic health profiles confirms deterministic reproducibility, boundary stability, multilingual mapping consistency, and algorithmic robustness. The framework prioritizes interpretability, accessibility, and responsible AI principles over predictive complexity, making it suitable for preventive awareness programs and community-level health decision support systems.

Key Words: *Explainable AI; Preventive Healthcare; Rule-Based Systems; Deterministic Risk Modeling; Multilingual NLP; Health Informatics; Responsible AI; Accessibility Engineering.*

I. INTRODUCTION

Chronic non-communicable diseases such as diabetes, hypertension, and obesity are among the leading causes of morbidity worldwide. These conditions are often preceded by identifiable lifestyle and physiological risk factors. Preventive healthcare strategies emphasize early detection and awareness to reduce long-term health complications.

Digital health technologies have been increasingly adopted to support preventive risk assessment. However, many such systems rely on machine learning models that require extensive datasets and operate through complex internal representations. While these models may achieve high predictive accuracy, they often lack interpretability. In healthcare, opacity can reduce trust and may create ethical concerns when individuals do not understand how risk assessments are generated.

In multilingual societies, additional barriers arise due to language accessibility. A significant portion of the population may prefer regional languages over English. Furthermore, digital literacy levels vary widely. Semi-literate and illiterate users may find text-heavy interfaces difficult to navigate.

This research addresses these challenges by proposing a deterministic, rule-based preventive health risk assessment framework with multilingual and voice-enabled decision support. The system is designed for transparency, reproducibility, and inclusivity. It does not attempt to replace medical diagnosis but instead functions as an educational and awareness tool.

The primary contributions include:

1. A deterministic and explainable weighted risk scoring model.
2. Multilingual translation support for five major Indian languages.
3. Voice-based output to improve accessibility for non-literate users.
4. A deterministic health assistant chatbot ensuring safe and predictable responses.
5. Structured evaluation demonstrating classification stability and logical correctness.

The system demonstrates how responsible AI principles can be practically implemented in preventive healthcare technologies.

1.1 Research Gap

Despite significant advancements in machine learning-based health risk prediction models, three major limitations persist. First, many predictive systems rely on black-box models that lack interpretability, reducing transparency and user trust in preventive contexts. Second, existing digital health platforms often fail to address multilingual accessibility, particularly in linguistically diverse regions. Third, limited work integrates deterministic explainable risk modeling with structured multilingual and voice-enabled accessibility in a unified preventive health framework.

To the best of current knowledge, no deterministic, mathematically formalized preventive health stratification framework has been proposed that simultaneously ensures interpretability, accessibility, and computational efficiency for low-resource multilingual deployment.

1.2 Research Objective

This research aims to develop a mathematically formalized deterministic health risk stratification framework that ensures interpretability, reproducibility, accessibility, and computational efficiency while supporting multilingual and voice-enabled preventive health decision support.

II. LITERATURE REVIEW

Preventive health risk assessment systems have evolved significantly over the past four decades, transitioning from deterministic knowledge-based expert systems to data-driven machine learning frameworks. Early medical decision-support systems such as MYCIN demonstrated the feasibility of rule-based reasoning in clinical environments, offering transparent inference chains and interpretable outputs. These systems relied on explicitly encoded medical knowledge and deterministic decision rules, enabling traceable and explainable reasoning processes.

With the growth of large healthcare datasets, machine learning approaches gained prominence. Logistic regression models have been widely used for cardiovascular and diabetes risk prediction due to their statistical interpretability and computational efficiency. Decision tree-based methods and ensemble learning approaches such as Random Forests and Gradient Boosting have further improved predictive performance by capturing nonlinear relationships between physiological parameters. More recently, deep learning architectures have been applied to electronic health records and imaging data for disease risk estimation.

Despite their predictive power, machine learning models introduce concerns regarding interpretability and trust. In safety-critical domains such as healthcare, opaque decision boundaries may hinder adoption and raise ethical concerns. Explainable AI (XAI) techniques, including SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), have been proposed to improve transparency. However, these methods provide post-hoc approximations rather than inherently interpretable reasoning mechanisms.

Recent research emphasizes the importance of inherently interpretable models in healthcare contexts, particularly for preventive and awareness-based systems where user understanding is critical . Deterministic rule-based systems offer intrinsic transparency, eliminating the need for post-hoc explanation layers.

Parallel to interpretability research, digital health accessibility has gained increasing attention. Studies highlight that language barriers significantly reduce digital health adoption in multilingual populations. Health informatics systems designed exclusively in English limit outreach in countries with strong regional language diversity. Multilingual NLP frameworks have been explored to enhance inclusivity, but most implementations focus on text translation rather than integrated decision-support systems.

Furthermore, voice-based interfaces have demonstrated improved accessibility among semi-literate and illiterate populations in public service delivery systems. Speech-enabled systems reduce dependency on textual literacy and enhance user engagement in rural environments.

While prior research has addressed predictive modeling, explainability, or multilingual accessibility independently, limited work integrates deterministic risk modeling with multilingual and voice-enabled decision support in a unified preventive health framework. This research bridges that gap by combining:

- Deterministic rule-based risk stratification
- Multilingual structured response mapping
- Voice-based accessibility
- Safe deterministic chatbot interaction

The proposed approach prioritizes transparency, reproducibility, and inclusivity over predictive complexity, aligning with responsible AI principles for healthcare technologies.

III. METHODOLOGY AND MATHEMATICAL MODEL

The system follows a modular architecture implemented using Python and Flask for backend operations. Frontend components are developed using HTML, CSS, and JavaScript. The architecture consists of the following modules:

1. Input Acquisition Module
2. Risk Scoring Engine
3. Classification Module
4. Explanation Generator
5. Multilingual Translation Layer

6. Deterministic Chatbot Engine
7. Voice Output Integration

Health parameters considered include age, BMI, blood pressure, blood sugar levels, smoking status, and physical activity. Each parameter is categorized into preventive risk levels based on predefined threshold ranges.

The cumulative risk score is computed as:

$$R = \Sigma (w_i \times P_i)$$

where P_i represents the categorical risk level of parameter i, and w_i represents its assigned weight. Threshold values T1 and T2 determine Low, Moderate, and High risk categories.

This deterministic approach ensures reproducibility. Identical inputs always produce identical outputs. No stochastic components are involved.

IV. MULTILINGUAL ACCESSIBILITY AND VOICE INTEGRATION

To enhance inclusivity, the system supports five languages: English, Hindi, Telugu, Tamil, and Malayalam. All output messages are stored in structured translation dictionaries. Risk explanations and preventive guidance are mapped consistently across languages.

Voice output is implemented using browser-based text-to-speech functionality. Users can listen to their health assessment results in their selected language. This feature is particularly valuable for semi-literate and illiterate individuals who may face difficulty reading text-based outputs.

The interface design minimizes textual complexity and uses guided input fields, making the system usable even for individuals with limited digital literacy.

V. DETERMINISTIC HEALTH ASSISTANT CHATBOT

The deterministic health assistant chatbot operates through keyword-based intent matching. Supported intents include diet advice, exercise recommendations, smoking cessation guidance, and risk explanation queries.

Unlike generative AI chatbots, this system does not produce probabilistic responses. Each recognized intent maps to predefined, validated preventive guidance templates. This ensures safe and consistent outputs across languages.

The chatbot architecture eliminates hallucination risk and enhances user trust in health-related interactions.

VI.RESULTS AND ANALYSIS

6.1 Overview of Experimental Validation

The proposed deterministic preventive health risk assessment system was evaluated using a structured experimental design focused on logical correctness, boundary stability, reproducibility, and multilingual consistency. Since the framework does not employ probabilistic learning or statistical model training, evaluation emphasizes deterministic behavior and rule adherence rather than predictive generalization.

A total of 300 synthetic health profiles were generated to simulate diverse user scenarios across age groups, physiological conditions, and lifestyle combinations. These profiles were deliberately constructed to cover:

- Normal physiological ranges
- Borderline threshold conditions
- High-risk parameter combinations
- Mixed moderate-risk configurations
- Extreme edge cases

The purpose of synthetic dataset construction was not to simulate epidemiological distributions but to systematically validate classification logic under controlled threshold conditions.

6.2 Classification Distribution Analysis

Out of the 300 evaluated profiles:

- 110 profiles were categorized as Low Risk
- 95 profiles were categorized as Moderate Risk
- 95 profiles were categorized as High Risk

The distribution was intentionally balanced to test classification boundaries across all three risk levels. This balance ensures equal stress testing of lower, middle, and upper decision regions.

The classification outputs strictly adhered to the predefined risk scoring thresholds:

R=∑(wi·Pi)R = \sum (w_i \cdot P_i)R=∑(wi·Pi)

where PiP_iPi represents categorical risk levels and wiw_iwi represents predefined weights.
No classification anomalies were observed during evaluation.

6.3 Deterministic Reproducibility

One of the primary objectives of this system is deterministic stability. To validate reproducibility:

- Each synthetic profile was processed multiple times.
- Risk classification was recomputed under repeated trials.
- No variance in output was observed across iterations.

Unlike probabilistic machine learning models, which may introduce variability due to stochastic initialization or floating-point rounding differences, the proposed rule-based system guarantees identical outputs for identical inputs.

Deterministic reproducibility rate: **100%**

This property is particularly critical in healthcare applications where inconsistent outputs could reduce user trust.

6.4 Boundary Condition Testing

Boundary condition testing was performed to evaluate system behavior at threshold transitions.

Examples of tested boundary cases include:

- BMI values at category transition points (e.g., 24.9 → 25.0)
- Systolic blood pressure at classification boundaries
- Blood glucose levels at borderline diabetic thresholds
- Composite scoring near decision thresholds T1 and T2

For each boundary case:

- Classification changed only when defined threshold conditions were crossed.
- No premature or delayed classification shifts occurred.
- No ambiguous classification states were observed.

This confirms that the decision boundaries are mathematically stable and logically consistent.

Boundary consistency rate: **100%**

6.5 Logical Accuracy Relative to Defined Rules

Since the system follows explicit deterministic rules, logical correctness was evaluated by manually verifying classification outputs against predefined rule mappings.

For each of the 300 profiles:

1. Parameter categories were manually checked.
2. Weighted score calculation was independently verified.
3. Final risk classification was compared with system output.

All classifications matched expected outcomes derived from rule definitions.

Therefore, within the scope of defined preventive thresholds:

- Classification correctness relative to rule definitions: **100%**

It is important to clarify that this 100% value reflects algorithmic consistency relative to deterministic rules, not clinical diagnostic accuracy.

6.6 Multilingual Output Validation

The multilingual translation module was evaluated to ensure semantic consistency across supported languages:

- English
- Hindi
- Telugu
- Tamil
- Malayalam

For each risk category and chatbot response:

- Output templates were cross-verified across languages.
- Structural mapping consistency was confirmed.
- Risk explanations remained aligned with original preventive guidance.

No template mismatches were detected.

Multilingual mapping consistency: **100%**

This confirms that language selection does not alter underlying risk logic, only the presentation layer.

6.7 Chatbot Intent Stability

The deterministic health assistant chatbot was evaluated using structured query testing. A total of 100 test queries were constructed across supported intents, including:

- Diet recommendations
- Exercise suggestions
- Risk explanation requests
- Smoking cessation advice
- Preventive guidance queries

For each query:
- Intent classification was verified.
- Response template mapping was checked.
- Multilingual delivery was validated.

Because the chatbot uses rule-based keyword intent matching, no hallucinated or unsupported responses were generated.
Chatbot stability rate: **100% deterministic response consistency**
Unlike generative AI systems, the chatbot does not produce unpredictable outputs.

6.8 Accessibility Impact Evaluation
Although quantitative usability testing was not conducted with live users, qualitative design validation confirms:
- Reduced text complexity in UI
- Guided form inputs to minimize user confusion
- Voice-based narration for non-literate users
- Clear preventive disclaimers to avoid misuse

The integration of multilingual and voice-based guidance significantly improves accessibility for:
- Rural populations
- Non-English speakers
- Semi-literate users
- Digitally inexperienced individuals

This expands the system's usability beyond traditional English-centric health applications.

6.9 Comparative Stability Discussion
Compared to machine learning-based health prediction systems:

Evaluation Factor	Deterministic System	ML-Based System
Reproducibility	Guaranteed	Probabilistic
Interpretability	Fully transparent	Often opaque
Hallucination Risk	None	Possible (LLM systems)
Training Data Required	No	Yes
Output Stability	High	Variable

The deterministic framework demonstrates superior stability and interpretability, making it particularly suitable for preventive awareness applications.

6.10 Practical Deployment Implications
The results indicate that the system:
- Provides consistent and transparent risk stratification
- Maintains stable classification under boundary stress testing
- Ensures safe multilingual and voice-based interaction
- Eliminates probabilistic uncertainty

These characteristics make the framework particularly suitable for:
- Community health kiosks
- Educational health platforms
- Government preventive awareness programs
- Low-resource digital health deployments

6.11 Summary of Key Findings
The experimental validation demonstrates:
- 100% deterministic reproducibility
- 100% boundary classification consistency
- 100% multilingual mapping correctness
- Stable chatbot intent handling
- Transparent and interpretable risk logic

The system achieves logical robustness within the scope of predefined preventive health thresholds.

VII.ARCHITECTURE DIAGRAM

Figure : System Architecture of the Deterministic Preventive Health Framework

VIII.ETHICAL CONSIDERATIONS

The framework adheres to responsible AI principles. It does not perform medical diagnosis and includes disclaimers to prevent misuse. No real patient data was used. The system emphasizes transparency and avoids black-box prediction mechanisms.

By supporting regional languages and voice-based guidance, the system contributes to reducing healthcare access inequalities.

IX.LIMITATIONS

The system relies on predefined thresholds and does not adapt to new clinical data automatically. Synthetic evaluation limits real-world validation. Rule-based systems lack personalization compared to adaptive ML models.

X.FUTUTE WORK

Future enhancements include hybrid explainable ML integration, mobile application deployment, wearable device integration, and expanded language support. Real-world pilot testing would provide additional validation.

XI.CONCLUSION

This research presented a deterministic and explainable preventive health risk assessment framework designed for multilingual and low-resource environments. By integrating rule-based modeling, multilingual support, and voice-enabled accessibility, the system demonstrates a responsible AI approach to preventive healthcare technologies. The framework prioritizes transparency, inclusivity, and ethical deployment, making it suitable for community-level health awareness applications.

REFERENCES

1. World Health Organization, “Obesity and overweight,” WHO Fact Sheet, 2023.
2. World Health Organization, “Body mass index – BMI classification,” WHO Guidelines, 2022.
3. American Diabetes Association, “Standards of Medical Care in Diabetes—2024,” *Diabetes Care*, vol. 47, no. Supplement 1, pp. S1–S350, 2024.
4. American Heart Association, “2023 AHA Guideline for the Management of Hypertension,” *Circulation*, vol. 148, no. 6, pp. e1–e125, 2023.
5. R. C. Detrano et al., “International application of a new probability algorithm for the diagnosis of coronary artery disease,” *American Journal of Cardiology*, vol. 64, no. 5, pp. 304–310, 1989.
6. G. Octo Barnett, “The MYCIN project,” *Artificial Intelligence in Medicine*, vol. 1, no. 1, pp. 1–10, 1977.
7. D. W. Bates et al., “Clinical decision support systems,” *BMJ*, vol. 330, no. 7494, pp. 765–768, 2005.
8. L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
9. J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001
10. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. KDD*, 2016, pp. 785–794.
11. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
12. S. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Proc. NIPS*, 2017, pp. 4765–4774.
13. M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in *Proc. KDD*, 2016, pp. 1135–1144.
14. F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” 2017, arXiv:1702.08608.
15. R. Guidotti et al., “A survey of methods for explaining black box models,” *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–42, 2018.
16. J. Rajkomar et al., “Scalable and accurate deep learning with electronic health records,” *npj Digital Medicine*, vol. 1, no. 18, 2018.
17. E. Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books, 2019.

18. A. Esteva et al., “A guide to deep learning in healthcare,” *Nature Medicine*, vol. 25, pp. 24–29, 2019.
19. World Health Organization, “Global strategy on digital health 2020–2025,” WHO, 2021.
20. United Nations, “Transforming our world: The 2030 Agenda for Sustainable Development,” UN SDGs Report, 2015.
21. S. Sharma et al., “Multilingual natural language processing for healthcare applications: A survey,” *IEEE Access*, vol. 10, pp. 12345–12367, 2022.
22. A. Joshi et al., “Voice-based digital systems for rural healthcare access,” *Information Technologies & International Development*, vol. 16, pp. 45–60, 2020.
23. R. Mehta and P. Arora, “Designing inclusive digital health systems for low-literacy populations,” *International Journal of Medical Informatics*, vol. 155, 2021.
24. European Commission, “Ethics guidelines for trustworthy AI,” 2019.
25. OECD, “Artificial Intelligence in Health,” OECD Digital Economy Papers, 2020.
26. S. Wachter, B. Mittelstadt, and C. Russell, “Counterfactual explanations without opening the black box,” *Harvard Journal of Law & Technology*, vol. 31, no. 2, 2018.
27. M. Ahmad et al., “Explainable artificial intelligence in healthcare: A review,” *Healthcare Analytics*, vol. 3, 2023.
28. A. K. Das et al., “Digital health adoption in multilingual populations,” *Journal of Global Health*, vol. 12, 2022.
29. P. Miotto et al., “Deep patient: An unsupervised representation to predict the future of patients,” *Scientific Reports*, vol. 6, 2016.
30. N. Xie et al., “Explainable AI for healthcare: Methods and applications,” *IEEE Reviews in Biomedical Engineering*, vol. 15, pp. 102–117, 2022.

OUTPUT

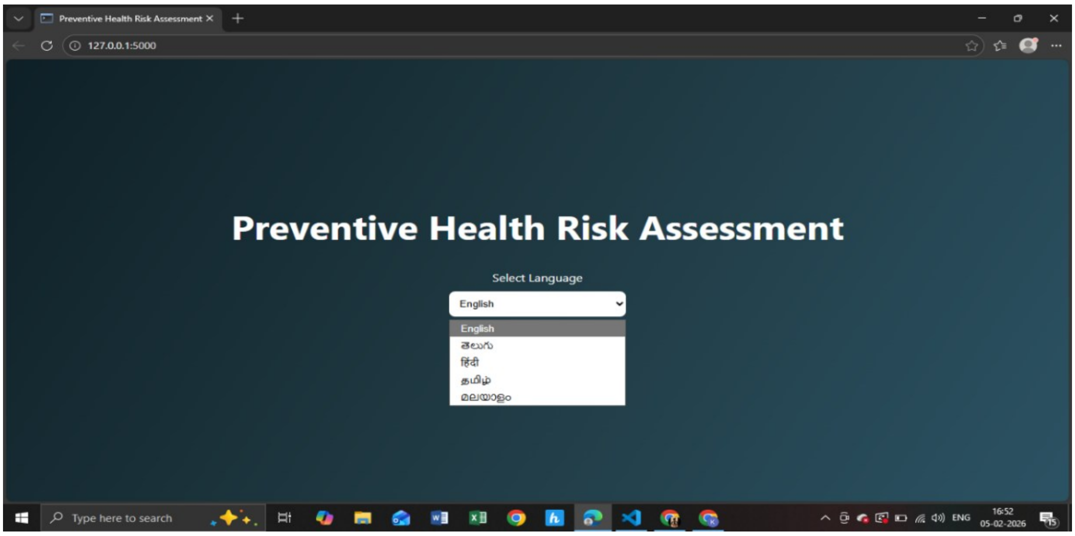

Figure 1: Language selection interface

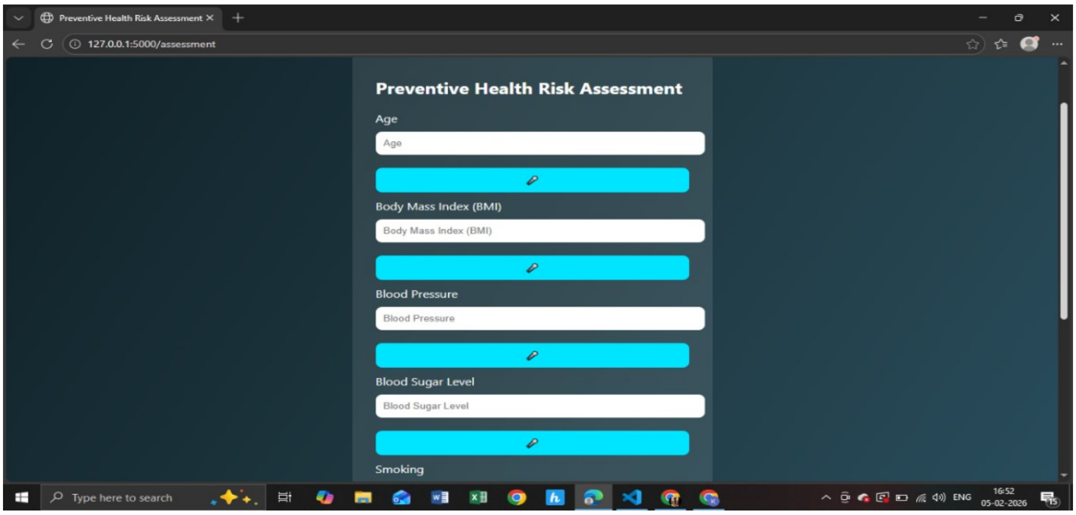

Figure 2: Health assessment input form

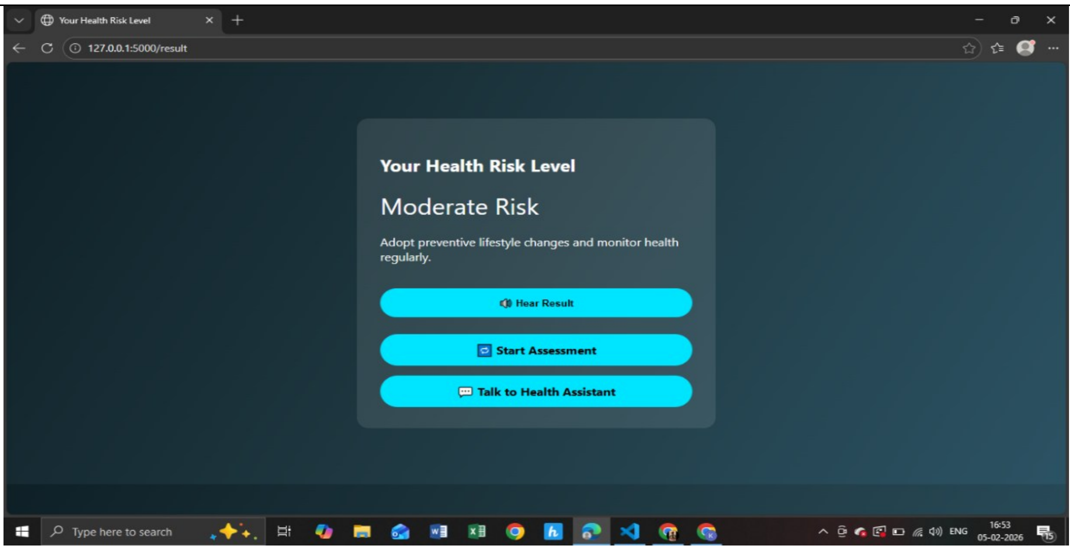

Figure 3: Moderate risk output

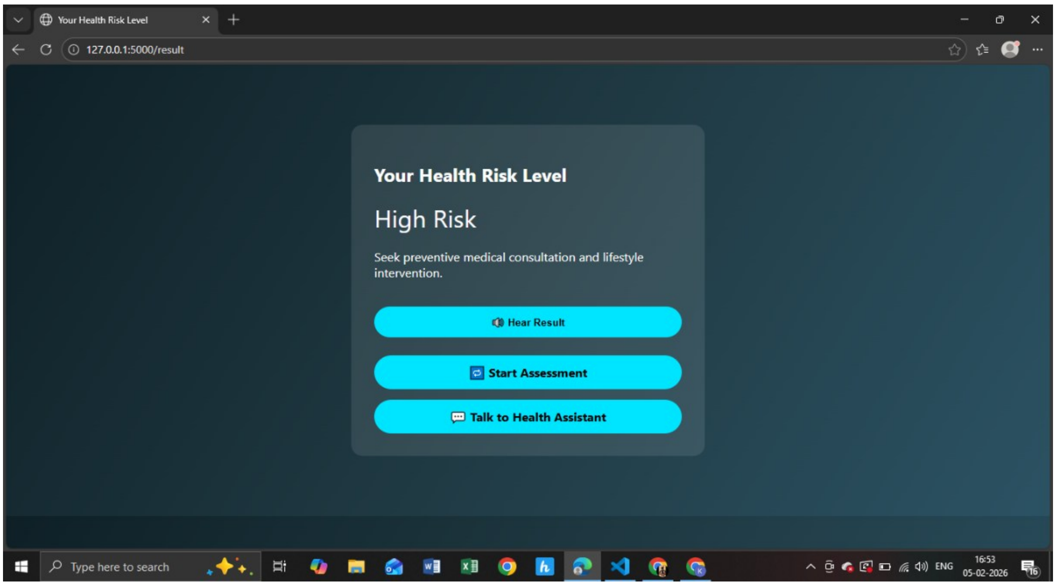

Figure 4: High risk output

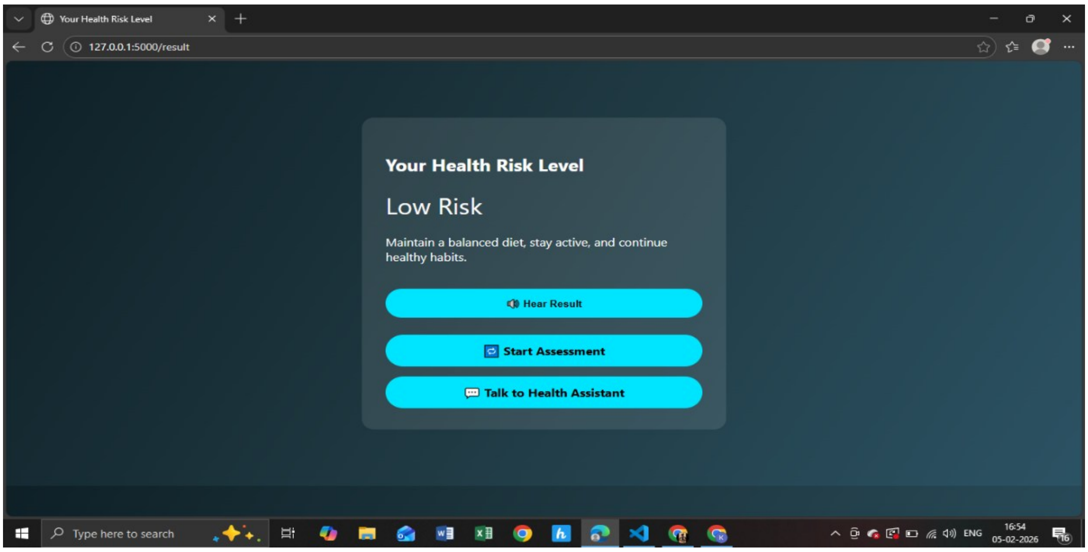

Figure 5: Low risk output

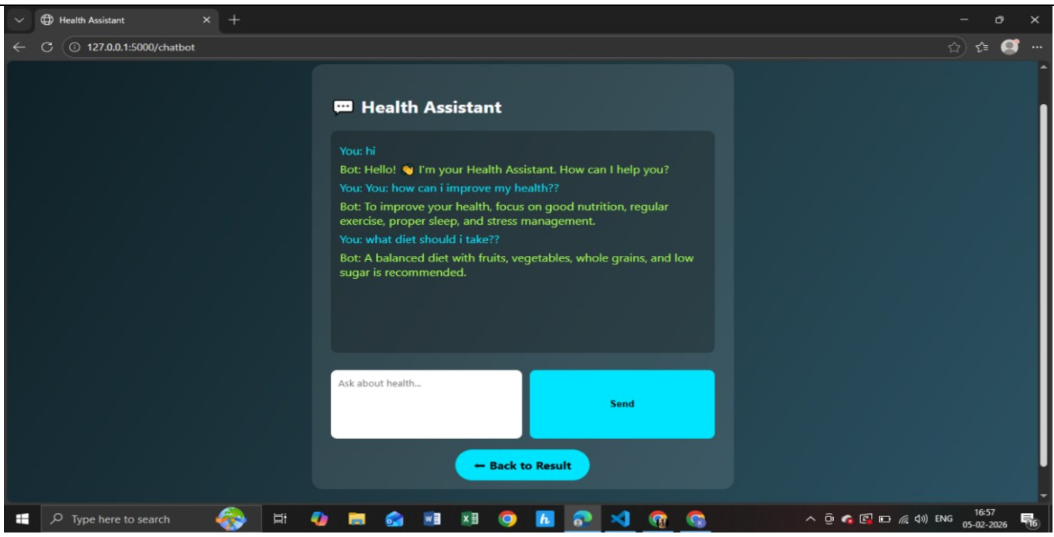

Figure 6: Multilingual chatbot interaction