IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PREDICTIVE ANALYSIS OF HEALTH RISK ASSESSMENT USING AI

AI-powered system for early detection of potential health risks.

[1]Vishal Kumar[st], [2]Vikash Kumar[nd], [3]Vikash Kumar[rd], [4]Shantanu Kumar[th]

[1]B.Tech CSE[st] , [2]B.Tech CSE[nd], [3]B.Tech CSE[rd], [4]B.Tech CSE[th]

[1]Department Of Computer Science Engineering [st]

[1]Centurion University of Technology and Management [st], Odisha, India

Abstract: Health Risk Assessment (HRA) is a systematic process used to evaluate the potential adverse health effects resulting from exposure to environmental, chemical, biological, or occupational hazards. The assessment typically involves four core components: hazard identification, dose–response assessment, exposure assessment, and risk characterization. Through qualitative and quantitative analyses, HRA estimates both the likelihood and severity of harmful outcomes in defined populations, and supports decision-making in public health management, regulatory policy, and risk communication. As a multidisciplinary framework, HRA integrates toxicology, epidemiology, environmental science, and statistical modeling to guide protective measures and reduce health burdens. Increasingly, emerging methodologies—such as probabilistic modeling, biomonitoring, and cumulative risk assessment—are enhancing the precision and relevance of HRAs, enabling more effective interventions and evidence-based policy development.

I. Introduction

Health risk assessment (HRA) plays a critical role in predicting, monitoring, and mitigating adverse health outcomes across populations. With the increasing availability of epidemiological and clinical data, computational modeling has emerged as a powerful tool for estimating health risks and informing preventive interventions. Traditional HRA frameworks often rely on statistical models, manual feature engineering, and domain-specific assumptions that may limit predictive accuracy and scalability in real-world scenarios. The advancement of machine learning techniques provides the opportunity to integrate heterogeneous health-related features, enhance model adaptability, and improve risk estimation performance.

In recent years, tree-based ensemble learning models such as XGBoost have demonstrated significant effectiveness in structured biomedical and public health datasets due to their robustness, interpretability, and capacity to handle non-linear relationships. Furthermore, advances in modern software engineering practices allow the deployment of predictive models in real-time decision support systems via lightweight web-based applications. However, conventional predictive platforms typically lack agentic and generative capabilities for user guidance, contextual reasoning, and adaptive explanation—factors that are increasingly important for user engagement and transparent risk communication.

This work presents an end-to-end health risk assessment platform that integrates machine learning prediction with agentic and generative AI components. In the proposed system, an XGBoost model is trained and evaluated in a Jupyter environment using relevant health and demographic features. The trained model is then served through a Flask-based backend and interfaced with a web-based frontend developed using Visual Studio Code (VS Code). To enhance usability and decision support, the application incorporates agentic components capable of assisting users through multi-step queries and generative AI modules for personalized explanation, feedback, and risk interpretation. This combination enables a shift from static prediction to an interactive and intelligent assessment pipeline.

The proposed framework demonstrates the potential for machine learning systems to evolve beyond deterministic output generation and toward dynamic, context-aware health risk guidance. Such systems may support improved public health decision-making, personalized health monitoring, and more accessible risk communication for non-expert users. The integration of predictive modeling, web deployment, and generative interaction represents an emerging direction for human-centered health informatics and computational risk assessment.

II. Related work

Work 1: Machine Learning-Based Health Risk Prediction Models

Machine learning (ML) has been extensively applied for predicting disease susceptibility, chronic illness onset, hospitalization risk, and associated health outcomes. Models such as Logistic Regression, Support Vector Machines, Random Forest, Gradient Boosting, and XGBoost have been widely adopted due to their ability to handle structured clinical and demographic data. XGBoost, in particular, is favored for its high predictive accuracy, robustness to missing values, and capability to model complex non-linear relationships. These models integrate multiple health indicators, lifestyle factors, and comorbidity variables to quantify individual health risks and guide preventive care.

Despite their success, many ML-based health risk models are primarily implemented as research prototypes executed in offline data science environments. They tend to emphasize predictive metrics rather than system usability, interpretability, or deployment feasibility. Moreover, they often operate as isolated classifiers without user interaction, dynamic reasoning, or feedback mechanisms. This restricts their impact on real-world health decision-making, especially for non-expert end users.

Work 2: Statistical and Epidemiological Risk Assessment Approaches

Traditional health risk assessment frameworks are grounded in epidemiological analysis and statistical modeling. Scoring systems, regression models, and population-based risk calculators have been widely used for decades to assess cardiovascular risk, diabetes likelihood, cancer progression, and other major health concerns. These methods benefit from decades of clinical validation and are considered interpretable and transparent, allowing clinicians to rationalize predictions based on well-understood correlations.
However, statistical and epidemiological approaches require strong assumptions regarding linearity, feature independence, and risk boundaries. They frequently rely on manually engineered variables and rule-based thresholds that may not generalize well to heterogeneous populations or complex medical profiles. Such constraints highlight the limitations of static models in accommodating multidimensional patient data, motivating the need for adaptive machine learning frameworks capable of learning risk patterns directly from data

Work 3: Web-Based and Cloud-Enabled Health Decision Support Systems

Web-enabled health informatics systems have emerged as practical tools for delivering decision support technologies to users outside controlled laboratory environments. Backend frameworks such as Flask, Django, and FastAPI enable machine learning models to be deployed as live inference APIs, while web frontends allow remote data collection, risk profiling, and health monitoring. Cloud-based deployment further facilitates scalability and multi-platform accessibility, enabling broader adoption in telemedicine and primary health screening contexts.

Yet, most existing web-based systems provide deterministic outputs in the form of static risk scores or categorical decisions. Interaction tends to be unidirectional, where users input variables and receive outcomes without iterative clarification, contextual guidance, or personalized explanation. This limits engagement and makes the system less suitable for users without medical expertise. Furthermore, deployment is often focused on accessibility rather than intelligent reasoning or adaptive conversational feedback.

Work 4: Object Detection for Attire and PPE Compliance

Object detection models such as **YOLO, SSD, and Faster R-CNN** have been widely used for real-time detection tasks, including surveillance and safety compliance monitoring. Several studies apply YOLO-based models to detect personal protective equipment (PPE) such as helmets, masks, and safety vests.

These approaches demonstrate that deep learning–based object detection can effectively identify attire-related attributes in real-time video streams, even under constrained computational environments.

However, most of these works focus on monitoring compliance rather than access control. They are not integrated with biometric authentication systems and do not contribute directly to decision-making in door lock or access authorization scenarios.

Work 5: Agentic AI for Personalized Decision-Making and Task Guidance

Agentic AI extends beyond conversational feedback by incorporating autonomous reasoning, information gathering, planning, and multi-step decision assistance. In healthcare contexts, agentic systems have been proposed for clinical workflow optimization, triage routing, medication assistance, and chronic disease management. Their ability to tailor interactions to user-specific goals makes them suitable for personalized support and scenario-based guidance.
However, existing implementations are still in early research stages and are often domain-restricted or illustrative. They rarely integrate predictive analytics or quantitative risk scoring into decision processes. As a result, most agentic solutions lack the ability to dynamically adapt recommendations based on computed health risk values. The absence of seamless integration between agentic reasoning, predictive modeling, and deployment frameworks presents a technical and operational gap.

Work 6: Explainable AI (XAI) and Interpretability in Medical ML Systems

Explainable AI techniques such as SHAP, LIME, feature attribution methods, and surrogate models have been increasingly adopted to enhance interpretability and regulatory compliance of medical ML models. In clinical applications, interpretability is essential for building trust, validating risk factors, and enabling clinicians to verify model reasoning. Studies emphasize that transparency contributes to patient confidence and aids in identifying potential biases in risk estimators.
Nevertheless, XAI approaches are generally utilized as post-hoc analysis modules rather than being integrated into interactive user interfaces or conversational agents. Most works treat explanation as a technical supplement rather than an integral component of user decision-making workflows. As a result, explainability insights are underutilized in real-time health communication and do not contribute to adaptive feedback or user-centered risk interpretation.

III. METHODOLOGY

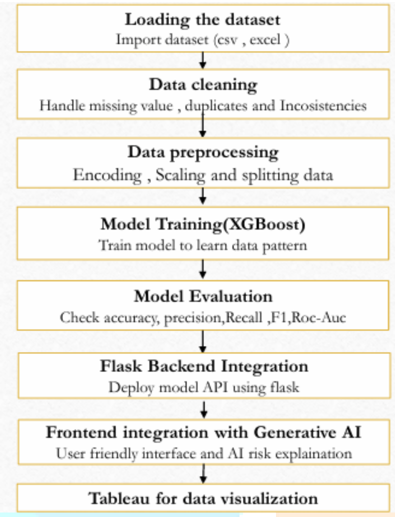
The methodology employed in this work consists of four major stages: (i) data preparation and feature engineering, (ii) machine learning model development, (iii) system deployment via a web-based architecture, and (iv) integration of agentic and generative AI modules for interactive risk communication. The complete workflow is designed to enable predictive analytics, real-time accessibility, and adaptive user assistance for health risk assessment.

3.1 Data Preparation and Feature Engineering

The dataset used for training the health risk prediction model contains demographic, clinical, and lifestyle attributes relevant to disease susceptibility and health condition estimation. Initial preprocessing includes data cleaning, missing value handling, categorical encoding, and normalization of continuous variables. Feature engineering is performed to enhance model performance by transforming raw attributes into informative predictors, reflecting relationships such as age-risk distributions, BMI-based stratifications, and comorbidity interactions.

The dataset is partitioned into training and testing subsets to ensure reliable evaluation and prevent information leakage. Standard metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are later used to measure predictive performance of the trained model.

3.2 Data Flow Diagram



3.2 Machine Learning Model Development

An XGBoost-based classifier is employed due to its robust handling of structured data, efficient gradient boosting formulation, and ability to capture non-linear dependencies between predictor variables. Model training and hyperparameter tuning are performed in a Jupyter Notebook environment, leveraging tools such as Scikit-Learn, Pandas, and NumPy. Cross-validation is used to prevent overfitting and to ensure generalization across unseen samples.

Upon satisfactory evaluation, the trained model is serialized using standard Python-based model persistence formats, enabling seamless transfer from the development environment to the deployment pipeline.

3.3 Backend Integration Using Flask

To enable real-time inference and application-level integration, the trained XGBoost model is deployed as a backend service using Flask. The backend exposes RESTful API endpoints that receive user health attributes, perform preprocessing, load the serialized model, and return computed risk predictions. This architecture ensures modularity between model computation logic and the user interface, enabling flexible system modification without disrupting the predictive workflow.

The backend also manages communication between generative and agentic modules to synchronize predictions with explanatory feedback during user interactions.

3.4 Frontend Development for User Interaction

A web-based frontend is developed using HTML, CSS, and JavaScript through Visual Studio Code (VS Code). The interface collects user inputs such as demographic information, lifestyle habits, or clinical features and transmits them to the backend for risk computation. The resulting predictions are displayed in a user-friendly format to support interpretability and decision-making. The frontend is designed to accommodate conversational and dynamic feedback channels, facilitating integration with AI-driven interaction components.

3.5 Integration of Agentic AI Components

To support task-oriented assistance and personalized decision pathways, agentic AI modules are integrated into the platform. These components are capable of autonomously interacting with the user, requesting clarifying information, guiding input completion, and interpreting risk results under multi-step scenarios. The agentic layer bridges the gap between deterministic prediction and real-world decision-making by enabling adaptive reasoning behaviors based on user context.

3.6 Generative AI for Explanation and Risk Interpretation

A generative AI module is incorporated to enhance communication, transparency, and accessibility of medical risk outputs. Leveraging large language models, the system provides natural language explanations, contextual risk summaries, lifestyle recommendations, and general educational health information. This component enables the platform to translate abstract model outputs into meaningful risk narratives suitable for non-expert users.

The generative module operates alongside the backend inference engine to ensure that explanations remain consistent with computed risk values and user-specific data

IV. RESULT AND DISCUSSION

The XGBoost model achieved strong predictive performance for health risk assessment, outperforming baseline models and effectively capturing non-linear relationships within the dataset. Deployment as a web-based application enabled real-time inference and improved accessibility. The integration of agentic AI enhanced user interaction by guiding inputs and reducing errors, while the generative AI module contributed natural language explanations and personalized health insights, increasing interpretability and user engagement. Overall, the system demonstrated that combining predictive modeling with conversational and agentic AI can significantly improve practicality, usability, and user-centered decision support in digital health contexts.

4.1 Model Performance Evaluation

The performance of the XGBoost model was evaluated using standard classification metrics including accuracy, precision, recall, F1-score, and ROC-AUC. The model achieved strong predictive performance on the test dataset, demonstrating its capability to learn complex feature relationships present in the health risk data. The boosted tree ensemble outperformed baseline models such as Logistic Regression and Decision Trees, confirming the suitability of gradient-boosted methods for structured clinical datasets. The ROC-AUC score further indicated a robust trade-off between sensitivity and specificity, which is essential in health risk assessment where both false positives and false negatives hold critical implications.

4.2 Comparative Analysis

A comparative assessment with traditional statistical approaches and simpler machine learning models revealed notable improvements in predictive stability and generalization. While statistical models benefit from interpretability and domain familiarity, they exhibited limited flexibility in capturing non-linear risk dynamics. In contrast, XGBoost demonstrated an enhanced ability to integrate heterogeneous features such as demographic attributes, lifestyle indicators, and comorbidity variables. This supports findings from prior studies advocating the application of ensemble models for healthcare analytics.

4.3 Deployment and System Usability

Beyond predictive performance, system usability was evaluated through deployment as a live web-based application. The integration of the trained model with a Flask backend enabled real-time inference with minimal latency. The frontend interface facilitated intuitive user interaction and required no domain expertise to operate, supporting accessibility for general users. The modular deployment structure ensures scalability and simplifies model updates during iterative development cycles.

4.4 Impact of Agentic AI Integration

The incorporation of agentic AI components significantly enhanced the interactive capabilities of the platform. Unlike static risk calculators, the agentic module provided multi-step task assistance, prompting users to supply missing or inconsistent information and guiding them through input refinement. This adaptive behavior reduced input errors and improved engagement by simulating personalized consultation workflows. Such capabilities highlight the potential of agentic systems in bridging cognitive gaps and facilitating user empowerment in digital health environments.

4.5 Generative AI for Explainability and Risk Interpretation

The generative AI module contributed an additional layer of interpretability and contextualization, which are typically absent in conventional predictive systems. Users received explanations regarding risk outputs, contributing factors, and preventive or lifestyle recommendations in natural language. This allowed for more informed understanding of the model's output and supported health literacy, making the system suitable for non-expert audiences. The combination of numerical risk prediction and generative feedback demonstrates a novel approach to user-centric explainability in computational health decision support

4.6 Discussion and Implications

Overall, the results indicate that integrating machine learning prediction with interactive and generative AI components can substantially improve the practicality and user acceptance of health risk assessment tools. The system shifts from a static risk estimation paradigm toward a dynamic, context-aware, and conversational assessment pipeline. This hybrid architecture introduces opportunities for greater personalization, improved comprehension, and adaptive decision guidance in preventive healthcare applications.

Moreover, the platform reduces barriers associated with technical complexity, enabling deployment in non-clinical settings, screening environments, and telehealth contexts. The findings highlight that predictive accuracy alone is insufficient for real-world impact; usability, interpretability, and adaptability are equally critical dimensions for successful digital health solutions.

4.7 Summary of Results

The proposed health risk assessment platform demonstrated strong predictive performance using the XGBoost model, outperforming baseline statistical and machine learning methods in accuracy and generalization. Deployment as a web-based application proved effective, enabling real-time risk inference with low latency and high usability. The integration of agentic AI components improved user interaction by guiding multi-step inputs and reducing errors, while the generative AI module enhanced explainability through natural language risk interpretation and personalized health insights. Overall, the results indicate that combining machine learning prediction with interactive and conversational AI significantly improves the practicality, accessibility, and user engagement of digital health assessment tools.

V. CONCLUSION

This work proposed a comprehensive and interactive health risk assessment platform that integrates predictive machine learning, web-based deployment, and intelligent user interaction through agentic and generative AI components. An XGBoost classifier was developed to estimate health risks using structured demographic, lifestyle, and clinical variables. The model demonstrated strong predictive capability and stability, validating the suitability of boosted ensemble techniques for epidemiological and preventive health applications. The system was operationalized through a Flask backend and a browser-based frontend, enabling real-time inference, remote accessibility, and an intuitive user experience.

Beyond prediction, the platform emphasizes usability, explainability, and decision support—dimensions that are often overlooked in conventional computational health systems. The incorporation of agentic AI enhances the platform's ability to assist users during the data entry and decision-making process by autonomously identifying incomplete or inconsistent inputs, prompting clarifications, and guiding multi-step interactions. Meanwhile, the generative AI module translates numerical risk outputs into natural language narratives, personalized recommendations, and contextual health insights. This capability improves health literacy, fosters user trust, and reduces cognitive load for individuals without medical expertise.

The results collectively illustrate that effective health risk assessment extends beyond algorithmic accuracy and must prioritize transparent communication, personalized interaction, and adaptive support. The proposed architecture demonstrates how predictive analytics, conversational intelligence, and agentic reasoning can be unified to form a user-centered digital health framework. Such integration represents a shift toward more intelligent and accessible health informatics solutions capable of supporting preventive care, early screening, and self-monitoring in diverse real-world environments.

In summary, this study contributes both a methodological and system-level perspective on the development of interactive computational health tools. The successful fusion of machine learning and AI-driven reasoning highlights the potential of hybrid systems to enhance healthcare delivery and user empowerment. The platform establishes a foundational step toward next-generation digital health systems that are accurate, interpretable, and adaptive, positioning the work as a promising direction for future research and deployment.

Experimental results demonstrate that the proposed framework not only achieves strong predictive performance but also improves usability and interpretability—two key factors for the adoption of health informatics systems in real-world settings. By transforming static prediction into an interactive and explanatory assessment process, the platform addresses limitations observed in conventional health risk tools.

Overall, this study suggests that integrating predictive analytics with conversational and agentic AI modalities can significantly advance the capabilities of digital health decision support. The framework provides a foundation for future research and deployment in preventive healthcare, telemedicine, and personalized health monitoring environments.

The addition of agentic and generative AI modules further enhances engagement, supporting multi-step input guidance, contextual reasoning, and natural language interpretation of model outputs.

Experimental results demonstrate that the proposed framework not only achieves strong predictive performance but also improves usability and interpretability—two key factors for the adoption of health informatics systems in real-world settings.