IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

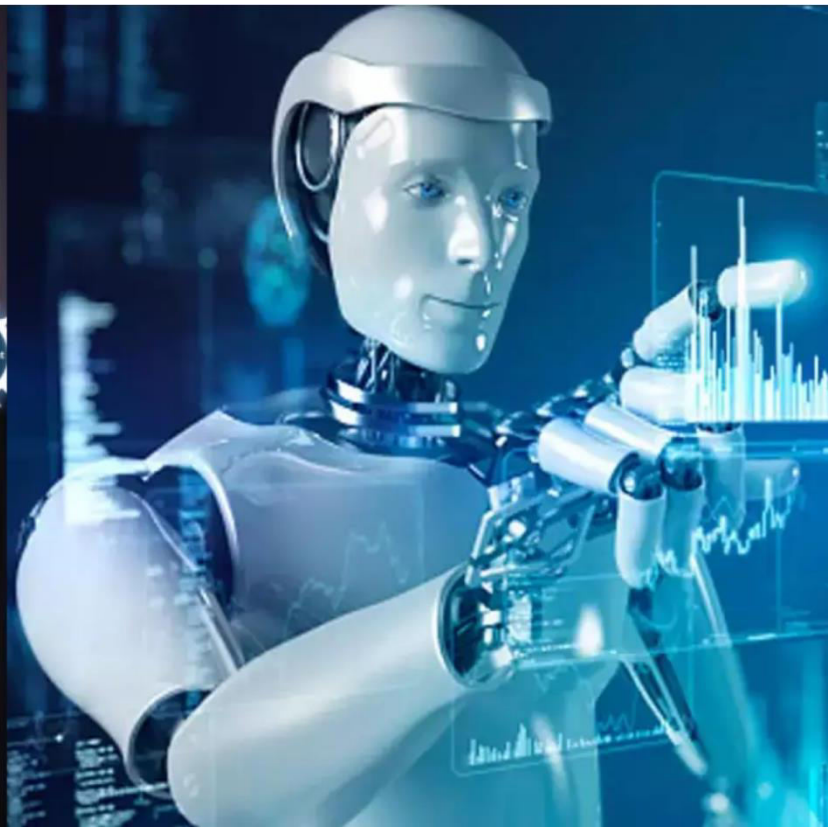
ISSN(O): 2319-8753
ISSN(P): 2347-6710



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 9.935

Volume 15, Issue 10, October 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

Personalized Preventive Healthcare System using AI for Disease-Aware Diet and Lifestyle Recommendations

Shravan Rajendra Harne, Siddhi Gajendra Kakade, Dr. R. A. Kale, Prajakta Shuddhodhan Gade, Sneha Dhanraj Khadse

Department of Computer Science and Engineering, Prof. Ram Meghe Institute of Technology and Research, Badnera–Amravati, India

ABSTRACT: Many wellness applications give every user the same targets and little indication of why a suggestion is made. This paper describes a personalized preventive healthcare system, a web-based prototype that combines a profile assessment, diet and workout guidance, a conversational assistant, experimental risk estimation with SHapley Additive exPlanations (SHAP), and persistent progress records. The system uses React and TypeScript, a Java 21 Spring Boot back end with MySQL, and a Python FastAPI service that serves Random Forest classifiers for diabetes-related and heart-related risk. Each classifier was trained on 5,000 synthetic, rule-generated records. On the repository's seed-42 data the Random Forest reached accuracies of 0.669 and 0.725 and macro ROC-AUC values of 0.856 and 0.881. Across 30 regenerated datasets it beat a logistic regression baseline by about 1.4–1.6 accuracy points but stayed about 3 points below the ceiling set by label noise, and a logistic regression given rule-indicator features did better still. The models therefore recover the synthetic labelling rule; their scores are not clinical accuracy. On the public Pima diabetes data, age and BMI gave a cross validated ROC AUC of about 0.73, with no significant Random Forest advantage. Warm /predict requests had a median latency of 125 ms locally. Seven of nine functional scenarios passed, one was partially verified and one was blocked. The Gemini-based assistant and the generated explanations have not been evaluated, and the wearable values are simulated.

KEYWORDS: Preventive healthcare; explainable machine learning; SHAP; Random Forest; health recommender system; large language model; web application.

I. INTRODUCTION

Unhealthy diet and physical inactivity are recognised behavioural risk factors for noncommunicable disease [1], so everyday lifestyle choices are a reasonable target for preventive technology. Generic wellness tools apply the same calorie target or workout list to every user and ignore differences in age, body composition, reported conditions, food preferences and sleep or stress patterns. The prototype described here conditions its guidance on a stored health profile instead. Recent work suggests that artificial intelligence (AI) can support personalised planning: a national preventive-care pilot in Singapore used an agentic assistant based on a large language model (LLM) to draft and refine health plans, and participants reported generally positive acceptance [2]. At the same time, health-related machine learning raises concerns about opacity [9], and language models can produce fluent but unsupported text [11]. A practical system therefore needs a clear separation between rule-based calculations, model-based estimates and generated text, together with honest communication of uncertainty.

This paper does not claim a new learning algorithm or a clinically validated predictor. It describes how the components of a student-built prototype are connected, which are implemented and which are planned, and what a reproducible evaluation of the risk models shows and does not show. Contributions: (i) a layered architecture with explicit boundaries between implemented and planned parts; (ii) a prediction service whose every response **states** that the models use synthetic data; (iii) a Manual/Tracker mode design that keeps simulated tracker values separate from user-entered data; (iv) a reproducible evaluation consisting of repeated runs, an exact noise ceiling, SHAP checks against the labelling rule, a separate public-data experiment and local latency measurements; and (v) a functional test record that distinguishes PASS, PARTIALLY VERIFIED and BLOCKED outcomes.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

II. RELATED WORK

Goh et al. piloted an agentic LLM assistant in a national preventive programme with 20 residents and 7 clinicians [2]; the pilot is small, so it indicates feasibility rather than effectiveness. A scoping review of LLMs for exercise advice found promise but called for better accuracy, closer alignment with professional standards, transparency and ethical integration [3]. Singar et al. review genetic inputs for personalised nutrition and stress privacy and ethics [4]; The system uses no genetic data. A Random Forest averages many trees trained on resampled data [5] and is a common baseline for tabular classification, implemented in scikit-learn [6]. SHAP attributes a prediction to features through Shapley values [7], and LIME fits a local surrogate model [8]. Rudin argues that explanations of black-box models can mislead in high-stakes settings and that interpretable models should be preferred [9]; a SHAP chart shows how a model behaves, not whether it is medically correct. Retrieval-augmented generation (RAG) conditions a model on retrieved passages [10], and a survey of hallucination explains why ungrounded output can be unreliable [11].

Gap. Each source addresses one component: plan generation, explanation or grounding. None of them, as reviewed here, examines how a prototype should keep simulated tracker inputs separate from manual entries, record the mode and data snapshot behind each saved recommendation, and state the synthetic-data status of its risk estimates in the API response. This is a narrow engineering gap identified from a limited set of sources, not a claim of novelty over the field.

III. SYSTEM DESIGN AND IMPLEMENTATION

A. Architecture

A React and TypeScript single-page interface communicates over REST APIs with a Java 21 Spring Boot (4.1) back end that handles registration, e mail one time password (OTP) verification, Google sign-in, cookie based sessions, assessment storage, dashboard metrics and workout personalisation, with MySQL storage (Fig.3.1). A separate Python FastAPI service exposes /predict, which runs scikit-learn Random Forest models and computes SHAP values. The assistant and a back-end explanation service are configured to call the Google Gemini API with the model identifier gemini 3.5 flash lite. In the deployed configuration, the front end runs on Vercel, the back end and ML service on Render and the database on Railway, all on free tiers.

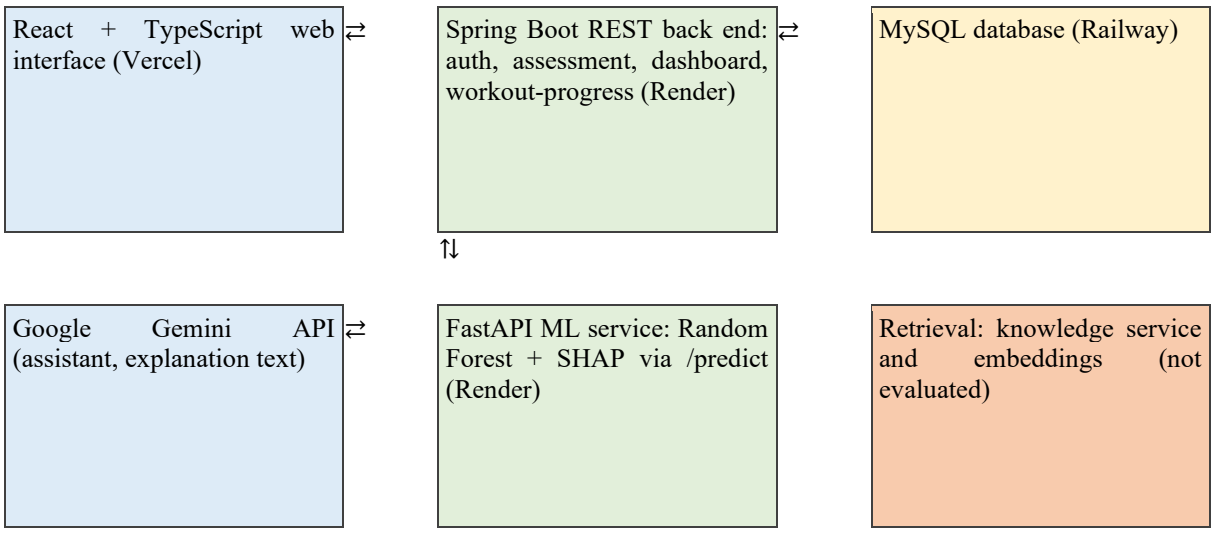


Fig. 3.1 Layered architecture

The browser interface talks only to the Spring Boot back end, which stores data in MySQL and calls the FastAPI service for risk estimates and SHAP values. The shaded retrieval component is implemented in code but not evaluated.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

B. Implementation status

It separates implemented, implemented-in-code and planned features. The notes show where evidence is missing: the assistant, retrieval and explanation text are unevaluated, and Google sign-in is unverified.

Table 3.1 separates implemented features from planned ones and states the main limitation of each.

TABLE 3.1 IMPLEMENTATION STATUS OF THE MAIN FEATURES

Feature	Status	Note
Registration, e-mail OTP, e-mail and Google sign-in, cookie sessions	Implemented	Scenarios 1–6 (Table VIII); Google sign-in not verified locally
Health assessment, dashboard (BMI, calorie and protein targets), diet and workout guidance	Implemented	Rule based; recommendation quality not evaluated
Random Forest risk estimates (diabetes, heart) with SHAP	Implemented	Synthetic labels; not clinically validated
Assistant calling the Gemini API with a safety system instruction	Implemented in code	Answer quality not evaluated; educational use only
Passage retrieval for the assistant (knowledge service, embeddings)	Implemented in code	One 3.2 kB file, four topics; retrieval not evaluated; the project report describes full RAG as incomplete
Plain language explanation of SHAP factors (Gemini)	Implemented in code	No generated text sampled or evaluated
Manual and Tracker modes; simulated wearable values; tracker context sent to the assistant only in Tracker mode	Implemented	Simulated values, not device readings
Diet plan history (time, mode, profile and tracker snapshots); per-user, per date workout completion	Implemented	Persists across refresh
Wearable synchronisation; training on real data; clinical validation; evaluated RAG	Planned	Not started

It separates implemented features from planned ones and states the main limitation of each and shows the implemented in code and planned features.

C. Modes, persistence and the assistant

On first use, the user chooses Manual or Tracker mode (kept in the browser's localStorage and changeable later; Fig. 3.2). In Manual mode the tracker page is hidden and no tracker data reach the assistant. In Tracker mode, a demonstration module supplies simulated steps, step goal, heart rate, sleep, active calories and weekly activity; none comes from a device. Each diet plan is saved with its generation time, mode and profile snapshot (plus a tracker snapshot in Tracker mode), so it can be traced to its inputs. Daily workout completion is stored per user and date through an authenticated GET/PUT endpoint (/api/workout-progress/me?date=YYYY-MM-DD). Figs. 3.2-3.7 show the main screens; they show the interface only and do not by themselves demonstrate the underlying code.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

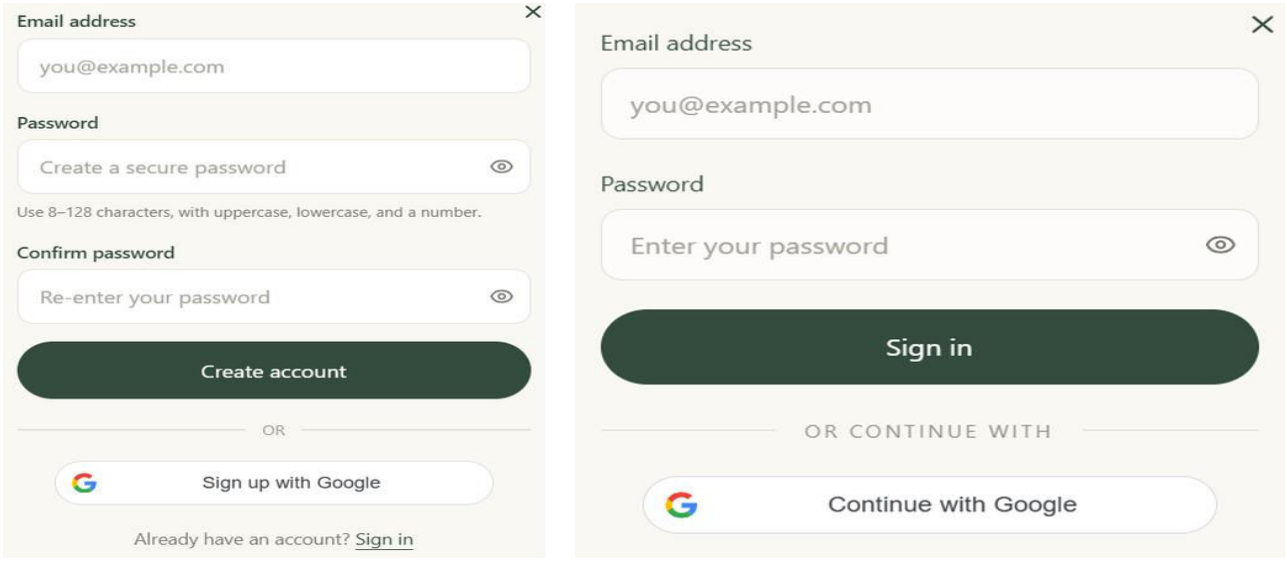


Fig. 3.2 Registration form (left; password rule of 8–128 characters with upper-case, lower-case and a digit; Google sign-up option) and sign-in form (right), cropped to the dialogs

This gives us the five-step assessment (personal information, health profile, lifestyle, diet and goals, review) collects the inputs used by the guidance and risk models in the fig 3.3. The saved Health Profile page shows age, height and weight and links back to editing.



Fig.3.3. Five step health assessment navigation (left) and the saved Health Profile page (right)

The first-use screen lets the user choose Manual Mode or Wearable Tracker Mode in the fig 3.4. Its footer states that the wearable values are simulated sample data, which keeps simulated and user-entered inputs distinct.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

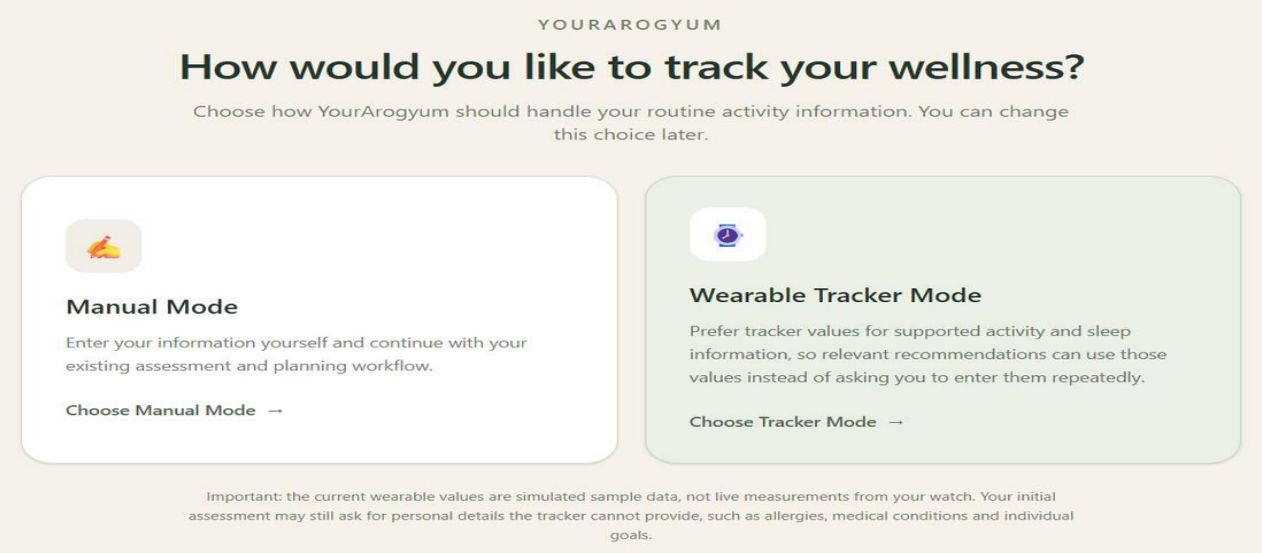


Fig. 3.4 Mode selection screen offering Manual Mode and Wearable Tracker Mode, with a notice that wearable values are simulated sample data

Now, in Tracker mode the assistant shows a banner saying the shared steps, sleep, heart-rate and activity values are simulated, not live in fig 3.5. The greeting limits its scope to nutrition, activity, sleep, hydration and preventive wellness.



Fig.3.5. Assistant page in Tracker mode

The dashboard (top) shows BMI, estimated daily calories, a protein reference and goal progress, and the Diet Plan page (bottom) repeats the calorie, protein and BMI estimates in the fig.3.6. Both are labelled as estimates, not medical prescriptions; the protein value differs (55.2 g vs 55 g) only because the two implementations round differently.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

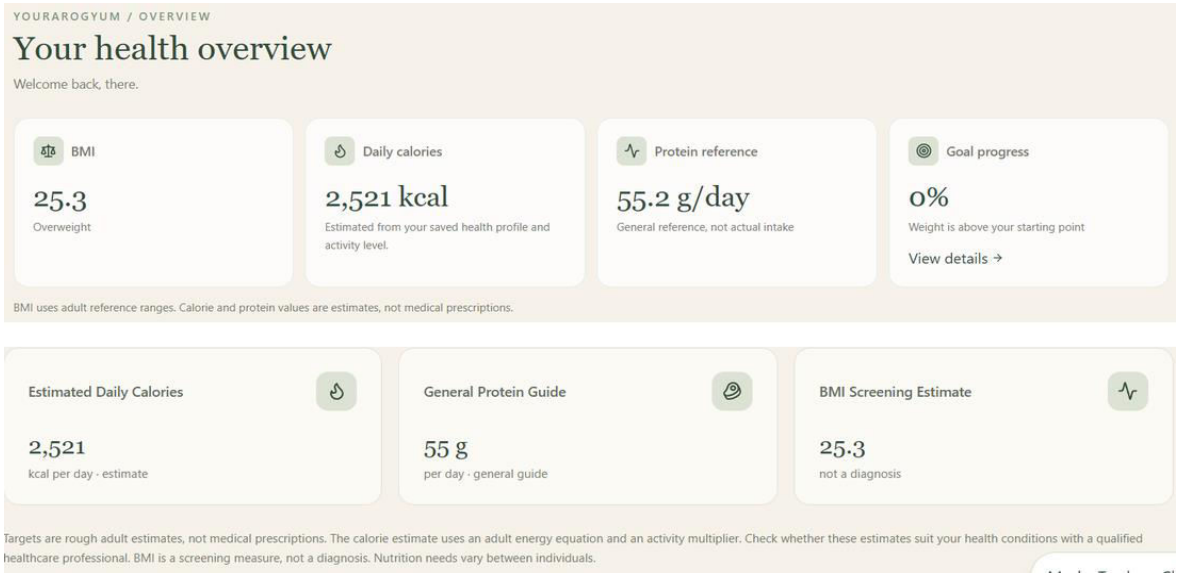


Fig.3.6. Dashboard cards

The workout page (top) shows total duration, exercise count, difficulty and a completion counter with progress bar (0/10 in this capture) Fig 3.7. The My Progress page (bottom) shows latest weight, goal progress and the weight-goal panel; this capture does not itself prove that data persist.

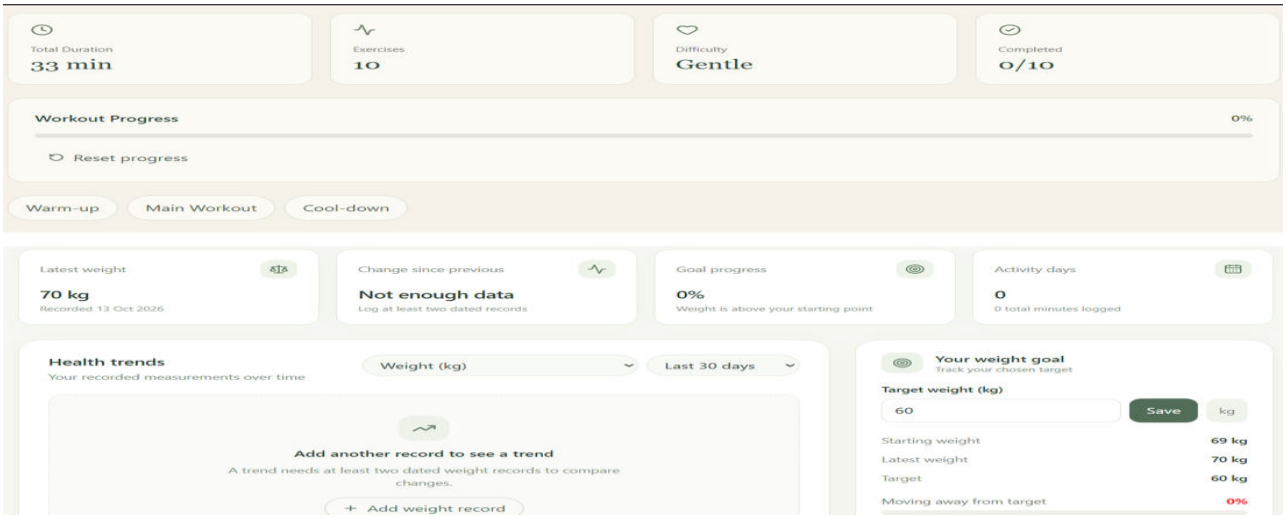


Fig.3.7. Workout plan page (top: session summary, completion counter and progress bar; no exercise completed in this capture) and My Progress page (bottom: latest weight, goal progress, trend placeholder and weight goal)

The assistant sends the user's question and profile context to the back end, which calls Gemini. A system instruction asks the model to act as a wellness assistant; avoid diagnosis, prescribing and claims that risk predictions are medically validated; not invent user data; recommend professional care for serious symptoms; and cite retrieved passages by label while saying when they do not answer the question. For each question, a knowledge service retrieves passages by embedding similarity (minimum cosine similarity 0.55, with a keyword fallback; embedding model gemini-embedding-001), adds them to the instruction and appends source titles and links. The knowledge base is one 3.2 kB Markdown file with four sections (healthy diet, physical activity, sleep, type 2 diabetes prevention), each linked to a WHO or CDC page; it has no content on hypertension, heart disease or thyroid disorders. Chat requests use the default temperature and a 2,500-token limit. A separate service turns the top SHAP factors into 2–3 plain-language sentences (temperature

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

0.4, 300 tokens); its prompt supplies each factor's direction (supports or opposes the predicted class) and absolute size, forbids causal claims and diagnosis, and requires a fixed closing sentence stating that the text is not a medical diagnosis. A failed call returns no explanation. Security controls are e-mail OTP, Google OAuth 2.0, cookie sessions, protected routes and credentials in environment variables (a pattern search of the repository history found no credential-like strings); encryption at rest, consent management, retention and formal security testing have not been evaluated.

IV. METHODS

A. Profile metrics and heuristic estimates

BMI is $W/(H/100)^2$ for weight W (kg) and height H (cm). Basal metabolic rate follows the Mifflin–St Jeor equation [12], $BMR = 10W + 6.25H - 5A + s$, with age A and $s = +5$ for men and -161 for women. Calories are BMR times an activity factor, computed in two separate implementations (the diet-plan page of the front end and the dashboard service of the back end) whose factor tables, handling of unmatched categories and protein rules differ (Table 4.1); the front-end estimate may also be adjusted for a target weight. For one test profile the dashboard showed 55.2 g of protein (Fig. 3.6, top) and the diet page 55 g, consistent with the two rounding rules. These are implementation-level heuristics, not individualised prescriptions.

TABLE.4.1. ACTIVITY FACTORS AND PROTEIN RULES IN THE TWO IMPLEMENTATIONS (IMPLEMENTATION-LEVEL HEURISTICS)

Implementation	Activity factors	Protein reference	Notes
Front end (diet-plan page)	Default 1.2 (also when the activity text is unmatched); light 1.375; moderate 1.55; very or highly active 1.725	0.8 g/kg/day, rounded to whole grams; 1.2 g/kg/day for muscle-gain or strength goals	Target-weight adjustment may change the calorie estimate; precise targets suppressed for some unsupported or medically sensitive profiles
Back end (dashboard service)	Sedentary 1.2; light 1.375; moderate 1.55; very active 1.725; extra active 1.9; unmatched category gives no calorie value	0.8 g/kg/day, rounded to 0.1 g; unavailable if age is missing or under 18, weight is missing, below 2 kg or above 500 kg, or conditions mention kidney, renal, dialysis, CKD, liver disease or pregnancy	Described in the application as a general adult reference, not a treatment target

B. Synthetic data and models

For each model, 5,000 records were generated in Python (seed 42). The diabetes model uses age (18–80), BMI (17–42), activity category (0–3), sleep hours (3–11) and a binary family-diabetes indicator; the heart model uses the same four plus a stress category (0–2) and a binary family-heart indicator. A hand-written points score of these features plus integer noise from -2 to $+2$ gives the class: Low (≤ 3), Moderate (≤ 6) or High. Class counts are 1,624/1,988/1,388 (diabetes) and 725/1,487/2,788 (heart). The labels are a noisy function of the inputs and are not derived from clinical records. Data were split 80/20, stratified (4,000 training, 1,000 test). The Random Forests use 200 trees, depth 10, at least 3 samples per leaf and balanced class weights; the baseline is a standardised multinomial logistic regression (2,000 iterations, balanced weights). SHAP values come from TreeExplainer and are those of the predicted class: a positive value supports it and a negative value opposes it. The service returns the class, probabilities, SHAP impacts and a notice that the models use synthetic data.

C. Evaluation protocol

We report accuracy, macro precision, recall and F1, and macro one-vs-rest ROC-AUC. Variability: data generation was repeated for 30 seeds (0–29, fixed in advance), each with its own stratified split and no tuning; we report means, SD, 95% intervals, and paired Random Forest minus logistic-regression differences with Wilcoxon tests (valid for this synthetic setting only). Noise ceiling: because the score and noise are known, an oracle that predicts the most probable class gives the best achievable accuracy; it uses the generating rule and is not deployable. A logistic regression on threshold

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

indicators of the rule (for example, age ≥ 55) serves as a diagnostic. SHAP checks: a fictional profile was sent to the local service; saved and refitted models were compared, additivity was tested on 300 rows, and mean |SHAP| per feature was compared with the drop in oracle accuracy when the feature is permuted. Public data: because no public dataset has the application's labels or features, new models were trained on the Pima Indians Diabetes Database [13] (768 women, 268 with diabetes by WHO criteria) using age and BMI, then adding the pedigree function, then all eight variables. Zeros were treated as missing and imputed within training folds, and models were compared by 20×5-fold cross-validation [14] and one held-out split. Performance: five start-ups and 300 timed requests after 20 warm-ups on Windows 11 (8 CPUs, Python 3.12.10). Functional tests: nine scenarios were run locally on 22 October 2026 and classed PASS, PARTIALLY VERIFIED or BLOCKED.

V. RESULTS

The model results in Sections V-A and V-B use synthetic, rule-generated labels, and Section V-C adds a separate real dataset (Pima), local latency measurements and functional tests. None of these measures clinical accuracy.

A. Synthetic labels

The repository code regenerates the manuscript's seed-42 results exactly, including all confusion matrices (Tables 5.2 and 5.3). The Random Forest is slightly ahead of logistic regression on every headline metric; the Moderate class is hardest (Table 5.1), and only 6 of 331 diabetes and 1 of 275 heart errors confuse Low with High. A majority-class predictor would score 0.397 and 0.558 accuracy. Across 30 regenerated datasets (Table 5.4, Fig. 5.1) the Random Forest averaged 0.656 (95% interval 0.650–0.662) and 0.728 (0.722–0.733) accuracy, so seed 42 is typical. It exceeded logistic regression by 1.4 accuracy points for diabetes (95% interval 0.8–2.0; better in 22 of 30 datasets; $p = 3 \times 10^{-4}$) and 1.6 points for heart (1.1–2.1; 27 of 30; $p = 3 \times 10^{-5}$), and by 1.7 and 1.2 macro ROC-AUC points. Repeated cross-validation gave similar values (diabetes 0.663 ± 0.018 vs 0.637 ± 0.015; heart 0.727 ± 0.021 vs 0.712 ± 0.015).

TABLE5.1. RANDOM FOREST PER-CLASS RESULTS ON THE SEED-42 TEST SPLIT

Model	Class	Prec.	Recall	F1	n
Diabetes	Low	0.741	0.723	0.732	325
Diabetes	Moderate	0.596	0.564	0.580	397
Diabetes	High	0.684	0.755	0.718	278
Heart	Low	0.651	0.669	0.660	145
Heart	Moderate	0.536	0.576	0.555	297
Heart	High	0.859	0.819	0.839	558

Per-class precision, recall and F1 for the Random Forest. The Moderate class is hardest in both models, and the heart model's High class is easiest.

TABLE 5.2. SEED 42 TEST RESULTS ON SYNTHETIC LABELS (N = 1,000 PER MODEL)

Classifier	Diabetes acc.	Diabetes macro F1	Diabetes AUC	Heart acc.	Heart macro F1	Heart AUC
Random Forest	0.669	0.677	0.856	0.725	0.685	0.881
Logistic regression	0.646	0.651	0.836	0.711	0.672	0.872

The Random Forest and logistic regression compared on the seed-42 test split. The Random Forest is slightly higher on accuracy, macro F1 and ROC-AUC for both models.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

TABLE 5.3. SEED 42 CONFUSION MATRICES (ROWS: ACTUAL; COLUMNS: PREDICTED; L = LOW, M = MODERATE, H = HIGH)

Model	Actual	RF: L	RF: M	RF: H	LR: L	LR: M	LR: H
Diabetes	Low	235	87	3	246	72	7
Diabetes	Moderate	79	224	94	90	188	119
Diabetes	High	3	65	210	1	65	212
Heart	Low	97	48	0	103	42	0
Heart	Moderate	51	171	75	67	168	62
Heart	High	1	100	457	4	114	440

Shows where the errors occur. Most are between neighbouring classes; only 6 of 331 diabetes and 1 of 275 heart Random Forest errors confuse Low with High.

TABLE 5.4. TEST PERFORMANCE OVER 30 REGENERATED SYNTHETIC DATASETS (MEAN ± SD)

Classifier	Diabetes acc.	Diabetes macro F1	Diabetes AUC	Heart acc.	Heart macro F1	Heart AUC
Random Forest	0.656 ± 0.016	0.664 ± 0.015	0.849 ± 0.009	0.728 ± 0.015	0.693 ± 0.017	0.892 ± 0.008
Logistic regression	0.642 ± 0.018	0.648 ± 0.018	0.832 ± 0.010	0.712 ± 0.015	0.679 ± 0.016	0.879 ± 0.009
LR on rule indicators (diagnostic)	0.681 ± 0.014	0.690 ± 0.013	0.859 ± 0.008	0.739 ± 0.013	0.702 ± 0.015	0.900 ± 0.008
Oracle ceiling (uses the rule)	0.688 ± 0.013	0.696 ± 0.013	0.859 ± 0.008	0.761 ± 0.013	0.719 ± 0.016	0.906 ± 0.006

Repeats the comparison over 30 regenerated datasets and adds the rule-indicator model and the oracle ceiling as references. The Random Forest is about 3 accuracy points below the ceiling, so the scores mainly reflect label noise.

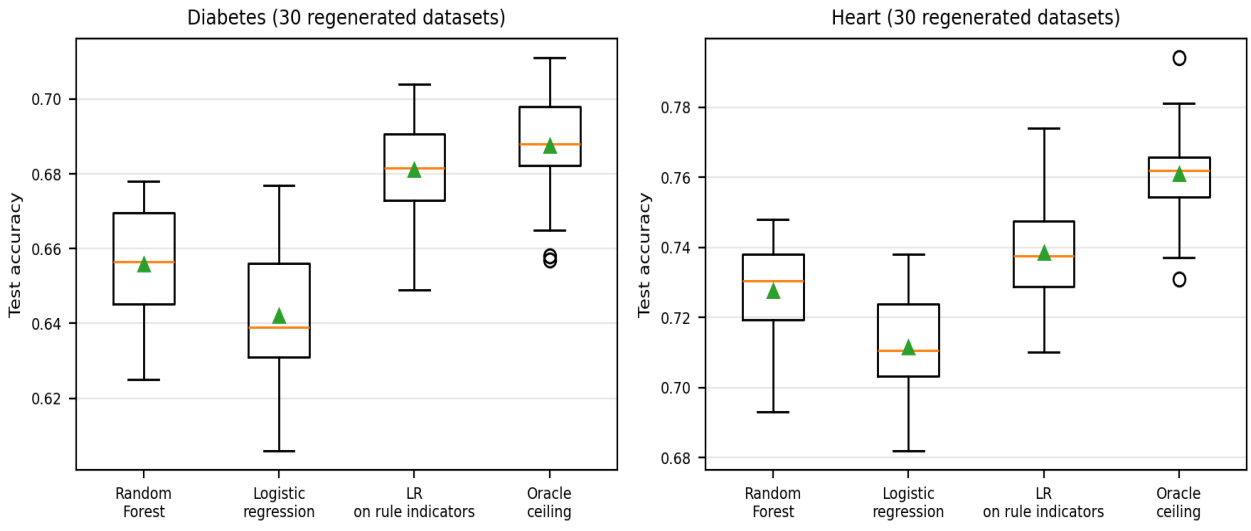


Fig.5.1. Test accuracy over 30 regenerated synthetic datasets (box: quartiles; line: median; triangle: mean)

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

On average the Random Forest is 3.2 accuracy points (95% interval 2.8–3.6) below the oracle ceiling for diabetes and 3.4 points (2.9–3.9) below it for heart (macro ROC-AUC gaps 1.0 and 1.4 points); on seed 42 the ceilings are 0.698 and 0.740 against 0.669 and 0.725. The indicator-based logistic regression comes within 0.7 and about 2 points of the ceiling and outperforms the Random Forest, so most of what the rule makes recoverable is recovered, and the accuracy of 0.66–0.73 mainly reflects the injected noise. These results show how well the models reproduce an invented rule, not that the rule or the models relate to real disease.

B. SHAP and explanations

For the fictional profile, the diabetes model predicted Moderate (probabilities 64.67% Moderate, 30.82% High, 4.51% Low) and the heart model predicted High (70.48%, 28.57% Moderate, 0.95% Low). Fig. 5.2 shows the five largest impacts from the saved service response. The saved models reproduced every probability and impact exactly, matched refitted models on all 1,000 test rows, and SHAP additivity held to 1.3×10^{-14} on 300 rows. For the diabetes model the ranking of features by mean |SHAP| equalled the ranking by oracle permutation importance (family history, age, BMI, sleep, activity); for the heart model the top two (stress, family history) agreed, but the rest differed (Spearman 0.66 over six features). The explanations therefore track the labelling rule, which is not evidence that they are medically meaningful. No generated explanation text was available, so its faithfulness to the SHAP input could not be assessed; a screening script is in the repository.

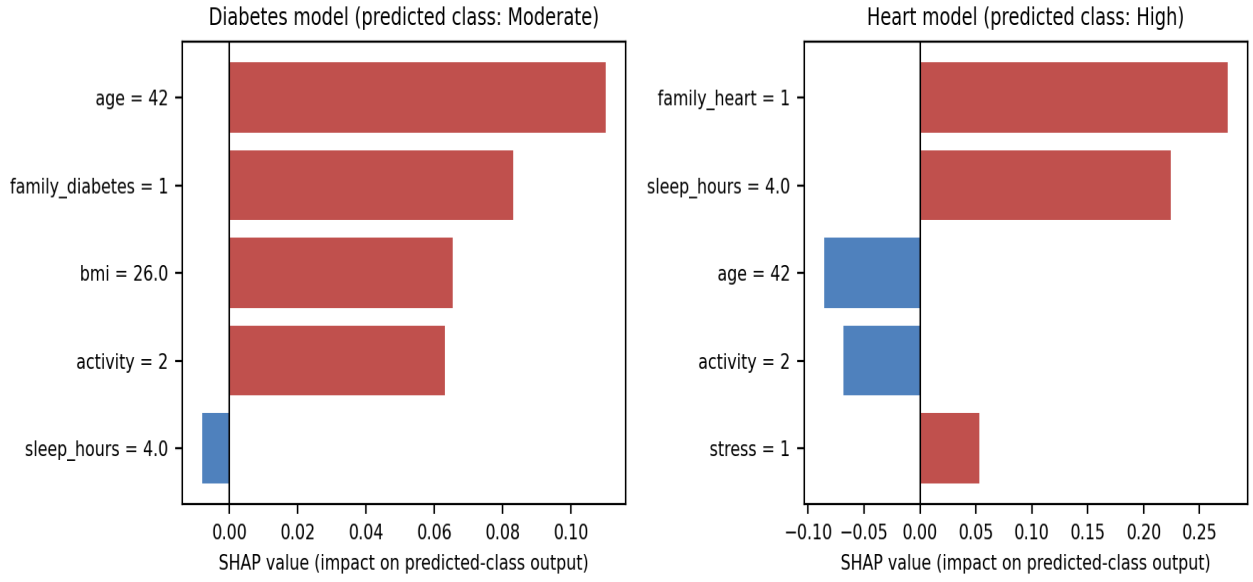


Fig.5.2. Largest SHAP impacts for the predicted class (one fictional profile; synthetic models)

C. Public data, performance and functional tests

On the Pima data (Table 5.5), age and BMI gave a cross validated ROC AUC of about 0.73 for both classifiers (majority class: 0.50) and the Random Forest was not better than logistic regression (difference −0.007, p = 0.61); the pedigree function raised AUC to 0.75–0.76 and all eight variables, including glucose, to 0.84. The application's age and BMI points alone gave an AUC of 0.64 (757 rows), a descriptive check of two of five rule components and not a validation. Discrimination is lower than on the synthetic labels, but the target, population and features are not comparable, and the held-out intervals are wide (Random Forest on age and BMI: AUC 0.745, 95% interval 0.655–0.822, n = 154).

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

TABLE.5.5. PIMA INDIANS DIABETES DATABASE (20×5-FOLD CV, MEAN ± SD; HELD-OUT AUC WITH 95% BOOTSTRAP INTERVAL)

Features	Classifier	CV ROC AUC	Held out AUC
Age, BMI	Logistic regression	0.734 ± 0.035	0.707 (0.626–0.785)
Age, BMI	Random Forest	0.727 ± 0.035	0.745 (0.655–0.822)
Age, BMI, pedigree	Logistic regression	0.748 ± 0.036	0.718 (0.634–0.794)
Age, BMI, pedigree	Random Forest	0.757 ± 0.038	0.777 (0.701–0.848)
All eight (reference)	Logistic regression	0.837 ± 0.027	0.813 (0.743–0.877)
All eight (reference)	Random Forest	0.838 ± 0.028	0.821 (0.751–0.884)

The prediction service became ready 2.64–3.28 s after start (median 3.22 s); the first /predict took 131.8–144.1 ms; and 300 warm requests had a median of 125.1 ms, a 95th percentile of 147.5 ms and no failures. These figures exclude the network, the Spring Boot back end, retrieval and Gemini calls and are not user-facing or free tier cold start times. The front-end production build produced a 941.8 kB script (256.6 kB compressed) and a 57.7 kB stylesheet; browser page-load times were not measured. In the functional tests (Table 5.6), seven scenarios passed, one was partially verified and one was blocked; scenarios 7 and 8 are developer-reported without captured numbers.

TABLE.5.6. FUNCTIONAL TEST OUTCOMES (LOCAL, 22 OCTOBER 2026, DEVELOPER-REPORTED)

No.	Scenario	Outcome	Evidence
1	Register and receive OTP by e mail	PASS	OTP received
2	Enter an incorrect OTP	PARTIALLY VERIFIED	UI showed “Bad Request”; HTTP 400 and message in source; account state not checked
3	Enter the correct OTP	PASS	Account verified, user logged in
4	E mail and password login	PASS	Login succeeded
5	Google sign-in	BLOCKED	Client identifier not configured locally
6	Dashboard without login	PASS	Redirected to login (private window)
7	Save and reload assessment	PASS	Developer reported; data retained
8	Dashboard calculations	PASS	Developer reported; no numbers captured
9	POST /predict with a valid profile	PASS	Class, probabilities, SHAP impacts and notice returned

Table 5.6 lists the nine functional scenarios run locally: seven passed, one (incorrect OTP) was only partly verified because the account state was not checked afterwards, and one (Google sign-in) was blocked because the client identifier was not configured. All outcomes are developer-reported, and scenarios 7 and 8 have no captured numbers.

VI. CONCLUSION

The proposed system connects a web interface, an application server, a database and a model service to deliver profile-aware guidance, explainable but unvalidated risk estimates, a general-purpose assistant and persistent progress records. The pipeline from profile to dashboard runs end to end with consistent outputs: the saved models match the training code, SHAP values add up and the sample response is reproducible. On synthetic labels the results are stable across 30 regenerated datasets. The Random Forest is modestly ahead of the standard logistic-regression baseline but behind a baseline given the rule’s indicators, and all stay within about three accuracy points of the ceiling set by label noise, so the models recover an invented rule. On the public Pima data the shared features give only modest discrimination and the Random Forest has no advantage. These results do not show that the risk estimates are valid for real people, that the assistant is accurate or safe, that generated explanations are faithful, or that users benefit.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 9.935

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 10, October 2026

|DOI: 10.15680/IJIRSET.2026.1510094|

The main limitations are: synthetic training data without clinical validation or fairness analysis, and inputs outside the generation ranges (age 18–80, BMI 17–42); an unevaluated assistant whose four-topic knowledge base omits conditions named in the system's aims; two separate, unreviewed calorie and protein heuristics; developer-reported functional tests, with Google sign-in unverified; latency measured only for the prediction service; simulated tracker values; and privacy measures (consent, retention, encryption, breach handling) that still need formal design. All model, SHAP, latency and public-data results can be regenerated from the public repository with fixed seeds and pinned requirements. A pattern search of the history found no credentials, and no licence file existed at evaluation time.

Before any claim of effectiveness, the work needs training and testing on a real dataset that contains the application's features, with calibration and clinical review. It also needs a two-reviewer, rubric-based evaluation of recorded assistant answers and screening of generated explanations against SHAP factors, an expanded and cited knowledge base, measurement of the deployed stack (including page loads and cold starts), security testing, and an ethics-approved usability study.

REFERENCES

[1] World Health Organization, "Noncommunicable diseases," fact sheet, 25 Sep. 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

[2] H. L. Goh, V. Sancenon, B. M. X. Chu et al., "Personalised health plan development using agentic AI in Singapore's national preventive care programme: a pilot study," npj Digit. Med., vol. 9, art. 332, 2026, doi: 10.1038/s41746-026-02514-8.

[3] X. Lai et al., "Using large language models to enhance exercise recommendations and physical activity in clinical and healthy populations: scoping review," JMIR Med. Inform., vol. 13, e59309, 2025, doi: 10.2196/59309.

[4] S. Singar, R. Nagpal, B. H. Arjmandi, and N. S. Akhavan, "Personalized nutrition: tailoring dietary recommendations through genetic insights," Nutrients, vol. 16, no. 16, 2024, doi: 10.3390/nu16162673.

[5] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[6] F. Pedregosa et al., "Scikit-learn: machine learning in Python," J. Mach. Learn. Res., vol. 12, pp. 2825–2830, 2011.

[7] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2017, pp. 4765–4774.

[8] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[9] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nat. Mach. Intell., vol. 1, no. 5, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

[10] P. Lewis et al., "Retrieval-augmented generation for knowledge-intensive NLP tasks," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, 2020, pp. 9459–9474.

[11] Z. Ji et al., "Survey of hallucination in natural language generation," ACM Comput. Surv., vol. 55, no. 12, art. 248, 2023, doi: 10.1145/3571730.

[12] M. D. Mifflin et al., "A new predictive equation for resting energy expenditure in healthy individuals," Am. J. Clin. Nutr., vol. 51, no. 2, pp. 241–247, 1990, doi: 10.1093/ajcn/51.2.241.

[13] J. W. Smith, J. E. Everhart, W. C. Dickson, W. C. Knowler, and R. S. Johannes, "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus," in Proc. Symp. Computer Applications in Medical Care, IEEE Computer Society Press, 1988, pp. 261–265.

[14] C. Nadeau and Y. Bengio, "Inference for the generalization error," Mach. Learn., vol. 52, no. 3, pp. 239–281, 2003, doi: 10.1023/A:1024068626366.



INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details