

PERSONALISED WEATHER INTELLIGENCE SYSTEM: A HEALTH-AWARE MACHINE LEARNING APPROACH TO PERSONALISED WEATHER RECOMMENDATIONS

Abdul Kareem

MSc Artificial Intelligence, London Metropolitan University, London, United Kingdom

abdulkareem55501@gmail.com

Corresponding Author: *

Abdul Kareem

DOI: <https://doi.org/10.5281/zenodo.21131501>

Received	Accepted	Published
25 April 2026	04 June 2026	21 June 2026

ABSTRACT

This paper presents the design, implementation and evaluation of a Personalised Weather Intelligence System for health-aware outdoor decision support. The central problem addressed is that conventional weather applications present the same meteorological information to all users, even though temperature, humidity, air quality and ultraviolet exposure have different practical meanings for older adults, people with asthma, cardiovascular disease or diabetes, and users with high heat or cold sensitivity. The proposed system combines live data from the OpenWeatherMap API with individual health-profile inputs and trained machine learning models to generate four personalised outputs: adjusted perceived temperature, health-risk classification, activity-suitability assessment and forecast-based activity timing guidance. The machine learning pipeline used the ASHRAE Global Thermal Comfort Database II and retained 16,141 complete records from 109,033 original measurements after strict cleaning. Three predictive components were selected through systematic algorithm comparison: a Random Forest Regressor for perceived temperature prediction, a Random Forest Classifier for activity suitability, and a Neural Network Multi-Layer Perceptron for health risk classification. The final prototype was implemented using FastAPI, scikit-learn, joblib and a CDN-based React frontend. Results show strong performance for perceived temperature prediction (MAE = 0.216°C), activity suitability classification (weighted F1 = 0.968) and health risk classification (accuracy = 0.992). However, the health risk model's high-risk recall of 0.719 confirms that the system remains a decision-support prototype rather than a deployable medical or safety-critical product. The paper concludes that personalised weather intelligence is technically feasible and practically meaningful, but future work requires outdoor thermal comfort data, class-balancing for high-risk examples, expert clinical validation and formal user testing. Its contribution is therefore both practical and critical: it demonstrates an integrated system while identifying the data, safety and user-evaluation requirements that must precede real deployment.

Keywords: Personalised weather intelligence; thermal comfort; health-aware decision support; machine learning; FastAPI; ASHRAE Global Thermal Comfort Database II; activity recommendation; environmental health risk.

1. INTRODUCTION

Weather applications are now part of everyday decision-making, but their outputs remain mostly generic. A healthy adult, a seventy-year-old with

heart disease, and a person with asthma standing in the same location typically receive the same temperature, “feels-like” value, air quality warning and activity guidance. The meteorological data

may be accurate, but the interpretation is not personalised. This is a significant gap because environmental exposure does not affect all users equally. Evidence from epidemiology and physiology shows that temperature and humidity affect cardiovascular and respiratory admissions differently across demographic and clinical subgroups (Achebak et al., 2024), that older adults experience reduced thermoregulatory capacity and heightened susceptibility during heat waves (Kenney et al., 2014), and that combined high temperature and humidity can increase mortality risk, especially among vulnerable groups (Fang et al., 2023). For health-sensitive users, the question is therefore not merely whether the forecast is correct, but whether the forecast is meaningful for their own body, risks and planned activity.

The Personalised Weather Intelligence System was developed to address this gap. Rather than altering the raw forecast, it adds an interpretation layer that combines current weather values with user profile information. Its aim is to translate temperature, humidity, wind speed, air quality and UV index into outputs that are easier to act upon: a personalised perceived temperature, a health-risk level, activity-specific suitability ratings, and safer future time windows for outdoor activity. This framing positions the system as decision support, not as medical diagnosis or automated clinical instruction. The project therefore sits at the intersection of thermal comfort research, machine learning, context-aware computing, health-aware recommendation systems and software prototyping.

The study was guided by a practical research question: to what extent can real-time weather data, individual health profile information and trained machine learning models be combined to generate personalised weather intelligence that is more useful than standard meteorological outputs? Eight objectives structured the work: integrating live weather and health profile data; training and comparing machine learning models; predicting personalised perceived temperature; classifying health risk; assessing selected outdoor activities; recommending safer forecast windows; implementing a FastAPI and

React prototype; and evaluating performance through model metrics, functional testing and critical appraisal. The paper integrates the literature review, methodology, system design, implementation, evaluation findings and conclusion while retaining the core citations, tables, figures and alphabetised reference list.

The article is organised as a conventional research manuscript: problem and motivation, literature synthesis, methodology, system design, implementation, results, findings and conclusion. This structure foregrounds the research contribution for an academic reader while preserving the evidence needed to evaluate the system. The retained tables and figures document model comparisons, system architecture, implementation workflow, user interface behaviour and evaluation outcomes that cannot be fully communicated through prose alone.

2. Literature Review and Research Gap

The literature reveals a strong justification for personalised weather interpretation. Mainstream applications such as BBC Weather, AccuWeather, Apple Weather and Google Weather provide highly detailed meteorological data, including temperature, humidity, wind, rain probability, air quality and UV information. However, most of these systems still use uniform interpretations. Traditional apparent-temperature indices, including Steadman's apparent temperature scale and wind-chill calculations, are based primarily on environmental variables rather than the personal health characteristics of the user (Steadman, 1984; Environment and Climate Change Canada, 2001). These models are helpful at population level but cannot represent the difference between a healthy young adult and a user whose asthma, diabetes, cardiovascular disease or cold sensitivity changes the risk meaning of the same forecast.

The personalisation gap is also a usability and equity issue. A generic forecast assumes that users can interpret the health implications of environmental variables themselves. This places the burden of risk translation on precisely the users who may be least able to absorb it quickly: older adults, people managing chronic illnesses, carers planning activities for vulnerable relatives,

and users who must decide whether it is safe to commute, exercise or remain outdoors. Generic air quality or pollen alerts partly address environmental exposure, but they usually apply the same threshold to everyone. A severe asthmatic and a healthy adult may therefore receive identical AQI messaging despite very different risk profiles. This mismatch between raw meteorological accuracy and personalised decision relevance is the core justification for a health-aware interpretation layer.

Thermal comfort research provides the scientific foundation for this project. ASHRAE Standard 55 defines thermal comfort as the condition of mind expressing satisfaction with the thermal environment (ASHRAE, 2020). Earlier thermal comfort models, particularly Fanger's Predicted Mean Vote and Predicted Percentage Dissatisfied framework, shaped decades of heating, ventilation and air-conditioning design (Fanger, 1970). Yet later work challenged the assumption that a single universal model can capture comfort across climates, cultures and adaptive behaviours (Humphreys and Nicol, 2002). This is important because a health-aware weather system must not simply reproduce generic environmental indices. It must allow personal factors to influence interpretation.

The ASHRAE Global Thermal Comfort Database II is a major resource for data-driven thermal comfort modelling. It contains more than 109,000 measurements from field studies across countries, climates and building types, including environmental variables, personal variables and subjective thermal responses (Földváry Ličina et al., 2018). It has been used widely in machine learning research. Luo et al. (2020) found that machine learning methods could outperform traditional thermal comfort models when personal and environmental variables were included. Zhang et al. (2024) similarly showed that personalised thermal comfort models combining environmental and occupant factors can achieve high prediction performance. These findings support the study's use of the ASHRAE database, while also exposing a key limitation: the database is primarily indoor-focused. Outdoor weather includes direct solar radiation, rain, variable gusts, radiant heat from

pavements and the metabolic demands of open-air activity, all of which are not fully represented in indoor thermal comfort data.

Machine learning has become increasingly prominent in thermal comfort and environmental health prediction. Ensemble algorithms such as Random Forest and Gradient Boosting have repeatedly performed well on tabular comfort data, especially when personal demographic features are combined with weather-related variables (Sarkar et al., 2024; Tian et al., 2025). Adaptive approaches can also refine comfort predictions from individual user feedback over time (Cosma and Simha, 2019), although this type of longitudinal feedback was outside the scope of the prototype. For health risk prediction, the literature is newer and less mature. A scoping review by Ssebyala et al. (2024) found that relatively few machine learning studies predict individual health risk from climate-sensitive weather events, which confirms the novelty of the present system. Related studies have linked air quality with emergency department visits for childhood asthma (Rosser et al., 2022), applied Random Forest to environmental classification of cardiovascular and respiratory disease risk (Cappelli et al., 2024), and demonstrated predictive analytics for health management in older adult populations (Wang et al., 2025).

Neural networks are also relevant where risk depends on complex non-linear relationships between personal and environmental variables. The Multi-Layer Perceptron is a feedforward architecture capable of modelling such relationships through layered transformations and non-linear activation functions (Goodfellow et al., 2016). Luo et al. (2020) found that MLP-based approaches performed strongly in thermal sensation prediction, while Togunwa et al. (2023) showed that neural network components can add value in health risk classification tasks where linear models and simpler decision structures may be insufficient. This evidence informed the selection of an MLP for health-risk classification, even though the final evaluation shows that conceptual suitability alone does not remove the need for class-specific safety analysis.

The project also draws on context-aware computing, defined as a system's ability to adapt its behaviour using information about a user's current context (Abowd et al., 1999). In health-aware computing, context may include location, activity, health status and environmental exposure. Haque et al. (2021) developed a personalised neural network model for asthma exacerbation prediction using weather triggers and health profile data, while Ayman et al. (2025) proposed health-aware route planning for asthma patients using live environmental inputs. Kyung et al. (2023) showed that contextually targeted weather-aware messages in mobile health applications can increase engagement with physical activity recommendations. These studies do not fully replicate the present system's broader multi-condition, multi-activity design, but they demonstrate that personalised environmental guidance is technically plausible and behaviourally relevant.

Ethical literature adds a necessary constraint. Health-related machine learning systems must be transparent about limitations and should present outputs as recommendations rather than instructions (Hatherley, 2024). GDPR treats health information as special-category personal data requiring heightened protection and clear consent (European Parliament, 2016). Professional computing guidance also requires developers to consider the public interest and the impact of systems on vulnerable users (BCS, 2022). Shortliffe and Sepúlveda (2018) argue that safety-relevant decision support should not depend solely on data-driven prediction but should combine machine learning with rule-based expert logic. This principle directly shaped the hybrid risk layer in the system. Overall, the literature confirms a clear research gap: existing weather services are informative but insufficiently personalised; thermal comfort models are powerful but mostly indoor-focused; machine learning studies are promising but fragmented; and health-aware weather intelligence requires explicit ethical boundaries.

The review therefore supports both the novelty and the caution of the project. It supports novelty because existing research has usually addressed only one part of the problem: thermal sensation,

asthma prediction, route planning, air quality risk or mobile health messaging. This study combines these strands into a unified prototype that accepts a general user health profile and produces multi-output weather intelligence. It supports caution because every strand also exposes a limitation: thermal comfort data is not the same as outdoor exposure data, weather-associated health outcomes are not identical across diseases, and persuasive health messages must be understandable without overstating certainty. The resulting design challenge is to create useful personalisation without making stronger claims than the evidence permits.

3. Methodology

The project followed a pragmatic, applied research philosophy. Its success was judged by whether a working system could be built, evaluated and critically appraised, rather than by adherence to a single theoretical tradition (Creswell and Creswell, 2018). The dominant approach was quantitative because model performance was assessed using Mean Absolute Error, Root Mean Squared Error, weighted F1 score, accuracy, precision and recall. A qualitative component was also necessary because a health-aware prototype cannot be judged only by headline accuracy; it must be appraised in terms of limitations, deployment risks, ethical responsibilities and user interpretation.

CRISP-DM was selected as the methodological framework because the project is fundamentally a data science and machine learning project rather than a conventional linear software build. CRISP-DM remains one of the most widely referenced process models for data mining and data science, and its iterative phases allow researchers to revisit data preparation, modelling and evaluation decisions as new results appear (Schröer et al., 2021; Martínez-Plumed et al., 2021). Waterfall was unsuitable because it assumes stable requirements and linear progression, while Scrum and similar Agile frameworks do not provide detailed guidance for dataset understanding, feature engineering and model evaluation. The CRISP-DM process used in the project is shown in Figure 3.1.

The business understanding phase translated the overall aim into three predictive tasks: regression for personalised perceived temperature, multi-class classification for health risk, and classification with scoring for activity suitability. The data understanding phase identified the ASHRAE Global Thermal Comfort Database II as the most suitable publicly available dataset because it contains environmental and personal variables that could be mapped to runtime weather and profile inputs (Földváry Ličina et al., 2018). The original dataset contained 109,033 records. Data preparation removed rows with missing values in selected columns rather than imputing thermal comfort variables such as clothing insulation, metabolic rate or age. This decision reduced the usable dataset to 16,141 records but avoided replacing personal physiological measurements with population averages, which would undermine the personalisation objective.

Feature engineering aligned training features with deployment requirements. The runtime system can obtain temperature, humidity, wind speed and air quality from OpenWeatherMap, and it can obtain age, health conditions, heat sensitivity and cold sensitivity from the user. The final feature vector therefore contained eight deployment-compatible features: `feature_temp`, `feature_humidity`, `feature_wind`, `feature_aqi`, `feature_age`, `feature_health_score`, `feature_heat_sens` and `feature_cold_sens`. Figures 3.2, 3.3 and 3.4 retain the original training outputs that document cleaning, algorithm comparison and model selection. A consistent 80/20 train-test split was applied, and five-fold cross-validation was used for selected models. Six algorithms were compared for the regression and activity suitability tasks, while eight were compared for health-risk classification. The project planning Gantt chart retained in Figure 3.5 shows how literature review, data

preparation, design, implementation, modelling, evaluation and writing were sequenced across the project timeframe.

Model evaluation was deliberately designed to look beyond a single headline metric. For regression, MAE was prioritised because it communicates the average absolute prediction error in degrees Celsius, which is directly interpretable by users and assessors. RMSE was retained as a secondary metric because it penalises larger errors more strongly. For classification, weighted F1 was used where class imbalance might affect the interpretation of accuracy, following the broader argument that precision, recall and F-measure reveal different properties of classification systems (Sokolova and Lapalme, 2009). For the health-risk task, per-class precision and recall were essential because a system can achieve high overall accuracy while still failing on the rare but most consequential high-risk class. This choice proved important, because the final MLP accuracy looked excellent until the high-risk recall value exposed a safety-relevant weakness.

Ethical design was embedded in the methodology rather than treated only as a final discussion point. The prototype collects sensitive health inputs, but it does not persist them beyond the session. This data minimisation decision aligns with GDPR principles and reduces the risk associated with an academic demonstration system (European Parliament, 2016; Conduah et al., 2025). The hybrid rule-based layer was also an ethical design decision. It recognises that a model trained on incomplete or imbalanced data may fail to recognise known dangerous conditions, and therefore supplements prediction with threshold logic. The methodology therefore combines empirical machine learning practice with precautionary reasoning appropriate for software that addresses vulnerable users.

Retained Methodology Figures

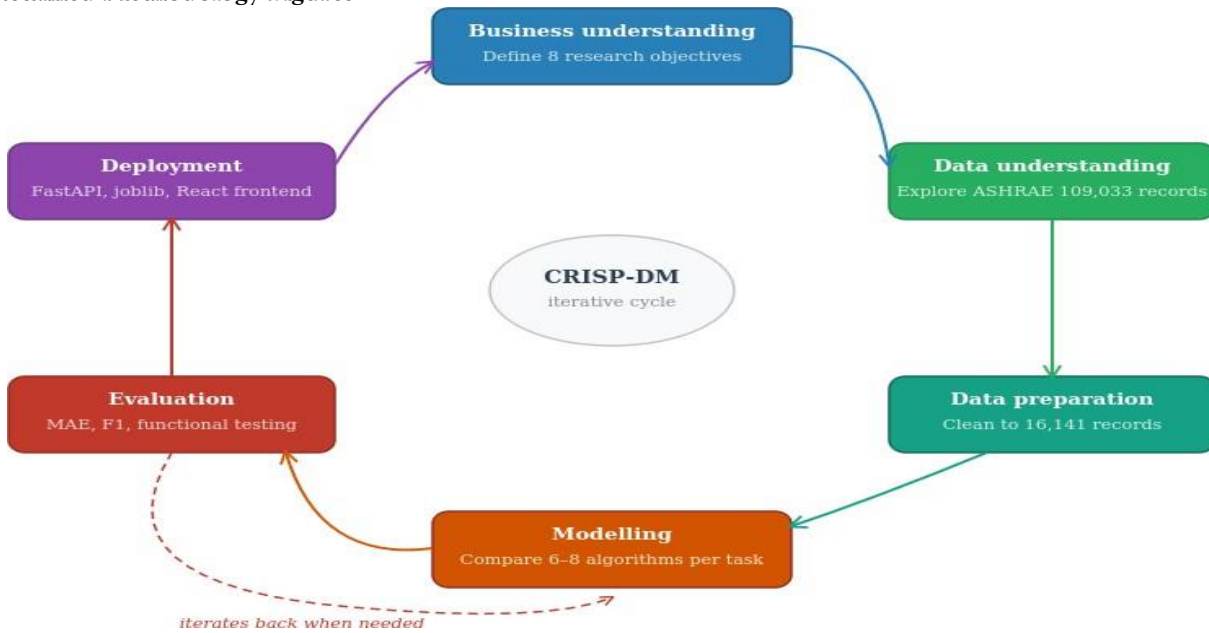


Figure 3.1: The CRISP-DM process model applied to the Personalised Weather Intelligence System

```

Loading ASHRAE dataset...
Total rows loaded: 109033
Rows after removing missing values: 16141

Feature set: ['feature_temp', 'feature_humidity', 'feature_wind', 'feature_aqi',
'feature_age', 'feature_health_score', 'feature_heat_sens', 'feature_cold_sens'
]
These 8 features match exactly what the app provides at runtime.
    
```

Figure 3.2: Terminal output from models/train_models.py confirming dataset loading, cleaning and feature selection

```

PERSONALIZED WEATHER INTELLIGENCE SYSTEM
ML Training Pipeline with CatBoost

Loading ASHRAE dataset...
Total rows loaded: 109033
Rows after removing missing values: 16141

Feature set: ['feature_temp', 'feature_humidity', 'feature_wind', 'feature_aqi', 'feature_age', 'feature_health_score', 'feature_heat_sens', 'feature_cold_sens']
These 8 features match exactly what the app provides at runtime.

MODEL 1: Feels-Like Temperature - 6 Algorithm Comparison
-----
Linear Regression      MAE=0.263  RMSE=0.438
Decision Tree         MAE=0.257  RMSE=0.427
Random Forest         MAE=0.216  RMSE=0.367
Gradient Boosting     MAE=0.234  RMSE=0.379
XGBoost              MAE=0.236  RMSE=0.373
CatBoost             MAE=0.241  RMSE=0.375

BEST: Random Forest (MAE=0.216)

MODEL 2: Activity Suitability - 6 Algorithm Comparison
-----
Decision Tree         F1=0.967  Acc=0.965
Random Forest         F1=0.968  Acc=0.965
Gradient Boosting     F1=0.966  Acc=0.964
XGBoost              F1=0.967  Acc=0.964
LightGBM             F1=0.967  Acc=0.964
CatBoost             F1=0.966  Acc=0.964

BEST: Random Forest (F1=0.968)
    
```

Figure 3.3: Terminal output showing dataset preparation and algorithm comparison results for the perceived temperature and activity suitability models

```

=====
MODEL 3: Health Risk - 8 Algorithm Comparison
=====
Logistic Regression      F1=0.978  Acc=0.978
Decision Tree            F1=0.994  Acc=0.994
Random Forest            F1=0.995  Acc=0.995
Gradient Boosting        F1=0.993  Acc=0.993
XGBoost                  F1=0.994  Acc=0.994
LightGBM                 F1=0.994  Acc=0.993
CatBoost                 F1=0.994  Acc=0.994
Neural Network (MLP)     F1=0.992  Acc=0.992

BEST by accuracy: Neural Network (MLP)
SELECTED: Neural Network (MLP) - clinical justification
Neural Network accuracy 0.992 is within 1% of best 0.995
Neural Networks preferred for health risk - complex non-linear patterns

Running 5-fold cross-validation...
Feels-Like CV MAE: 0.314 (+/- 0.029)
Activity CV F1: 0.950 (+/- 0.009)

=====
FINAL RESULTS SUMMARY
=====
Feels-Like : Best=Random Forest MAE=0.216C
Activity   : Best=Random Forest F1=0.968
Health Risk : Selected=Neural Network (MLP) Acc=0.992

All models saved successfully!
PS G:\Final Year Project\weather_app
    
```

Figure 3.4: Terminal output showing health risk algorithm comparison, cross-validation results and final model selection summary

Project Gantt Chart – Personalized Weather Intelligence System

MSc Artificial Intelligence – London Metropolitan University – Supervisor: Dr Ramzi Djemai

Student: Abdul Kareem | Student ID: 24040813 | Project Duration: February 2025 – May 2025

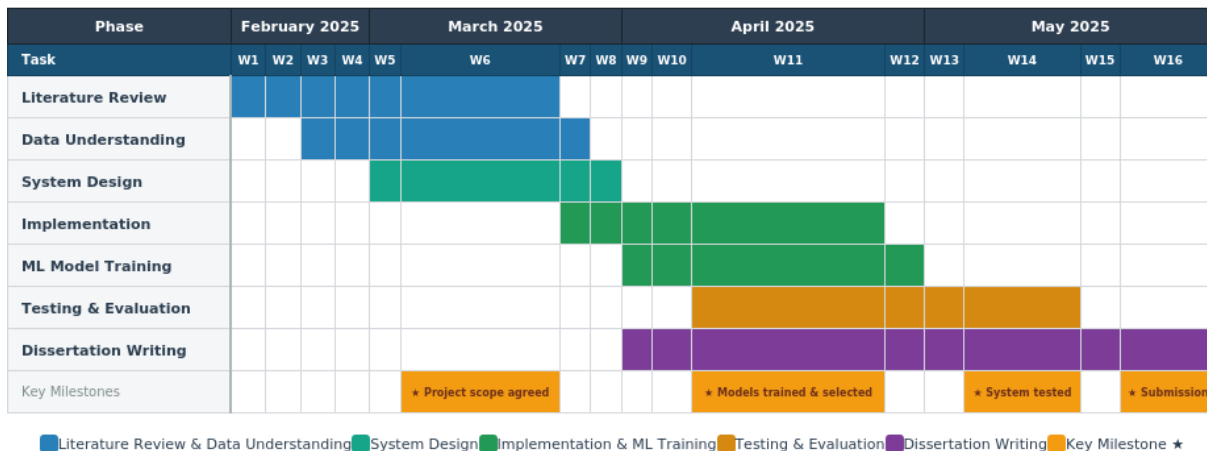


Figure 3.5: Project Gantt chart showing the seven development phases and key milestones across the project timeline

4. System Design

The Personalised Weather Intelligence System was designed as a three-tier architecture with clear separation between presentation, application logic and machine learning. The presentation layer is a React-based frontend served at the /ui endpoint. The application logic layer is a FastAPI backend that validates requests, calls the OpenWeatherMap API, constructs the runtime feature vector and coordinates prediction functions. The machine learning layer contains

three serialized scikit-learn models loaded with joblib. The responsibilities of these layers are summarised in Table 4.1 and the high-level architecture is retained in Figure 4.1. This modular separation is consistent with established software architecture principles because it allows each component to be tested, replaced or improved independently (Bass et al., 2021). FastAPI was selected because it is lightweight, Python-native and well suited to machine learning API deployment. Its Pydantic-based

request validation prevents malformed user profile data from entering the prediction pipeline, while automatic documentation at /docs supports development testing and demonstration (Ramírez, 2021). Django was unnecessary because the prototype did not require an ORM, database-backed administration or large application structure. Flask was possible but offers less built-in validation and documentation. The frontend used React via CDN instead of a Node or Vite build pipeline. This choice deliberately prioritised demonstrability: the full interface can be served directly by FastAPI with a single backend command and no separate frontend build step. The trade-off is that production optimisations such as bundling, tree-shaking and minification were not applied.

The architecture also reflects the need for auditability. A health-aware recommendation should be explainable at the level of data source, feature transformation, model output and rule-based adjustment. For this reason, the system does not hide all logic inside a single model. Weather retrieval, health-score construction, prediction and recommendation generation are separate stages. This enables the developer or evaluator to trace how a city, age, condition set and activity selection produce a final recommendation. The approach is especially important where users may question why a risk was escalated or why one activity is safer than another.

The data flow begins when the user enters a city and health profile in the frontend. The backend requests current weather data, extracts latitude and longitude, retrieves air quality and UV values, and normalises all environmental values into a single dictionary. User profile inputs are converted into personal features: age is taken directly, selected conditions are converted into a weighted health score, and heat/cold sensitivities are encoded numerically. These are combined with environmental features to create the eight-element feature vector. The system then runs perceived temperature, health risk and activity suitability predictions, and separately retrieves five-day forecast windows for future activity

timing. Figure 4.2 preserves the end-to-end data flow of this pipeline.

The perceived temperature predictor uses a two-stage design. The Random Forest Regressor produces a numerical output, after which a rule-based personalisation offset adjusts the result using the user's health conditions and sensitivity levels. This offset exists because the ASHRAE dataset does not explicitly encode conditions such as asthma, arthritis or heart disease. The health-risk classifier is more safety-sensitive and therefore uses a hybrid two-layer design. First, an MLP returns a Low, Medium or High prediction and probability-based confidence score. Second, a rule-based threshold layer independently checks known risk conditions. The final risk is the higher of the machine learning prediction and the rule-based assessment. This design was selected specifically because the evaluation later showed that high-risk recall from the MLP alone was only 0.719.

Activity suitability is assessed independently for each selected activity, recognising that acceptable conditions for walking may not be acceptable for running, cycling or gardening. A Random Forest Classifier predicts a suitability class, which is then converted to a percentage score through a probability-weighted mapping. Rule-based adjustments account for activity-specific factors such as wind speed for cycling, high UV for gardening and poor air quality for running. The frontend was designed as a step-based single-page interface rather than a dense dashboard. This progressive disclosure approach reduces cognitive load for older or health-vulnerable users by presenting profile entry, summary results, risk explanation and forecast advice in digestible stages (Rogers et al., 2023). Figures 4.3 and 4.4 retain the original landing-page and navigation-flow diagrams.

A further design principle was training-runtime consistency. This study treats this as a deployment concern rather than a minor coding detail. In many academic machine learning projects, models perform well in notebooks but become unreliable when connected to a live application because the inference feature vector does not exactly match the training structure. This system avoids that problem by saving the

feature column order at training time and loading it during runtime prediction. The design also separates machine learning outputs from recommendation text. The model produces numerical or categorical outputs, while a recommendation layer converts those outputs

into human-readable guidance, limiting factors, positive factors and precautions. This makes the prototype easier to inspect and safer to modify because recommendation language can be reviewed without retraining models.

Table 4.1: System architecture layers and their responsibilities

Layer	Technology	Primary Responsibility
Presentation	React via CDN	User input collection, result display, step navigation
Application Logic	FastAPI, Python	Request handling, API orchestration, feature engineering
Machine Learning	scikit-learn, joblib	Perceived temperature, health risk, activity prediction
External Data	OpenWeatherMap API	Live weather, AQI, UV index, forecast data

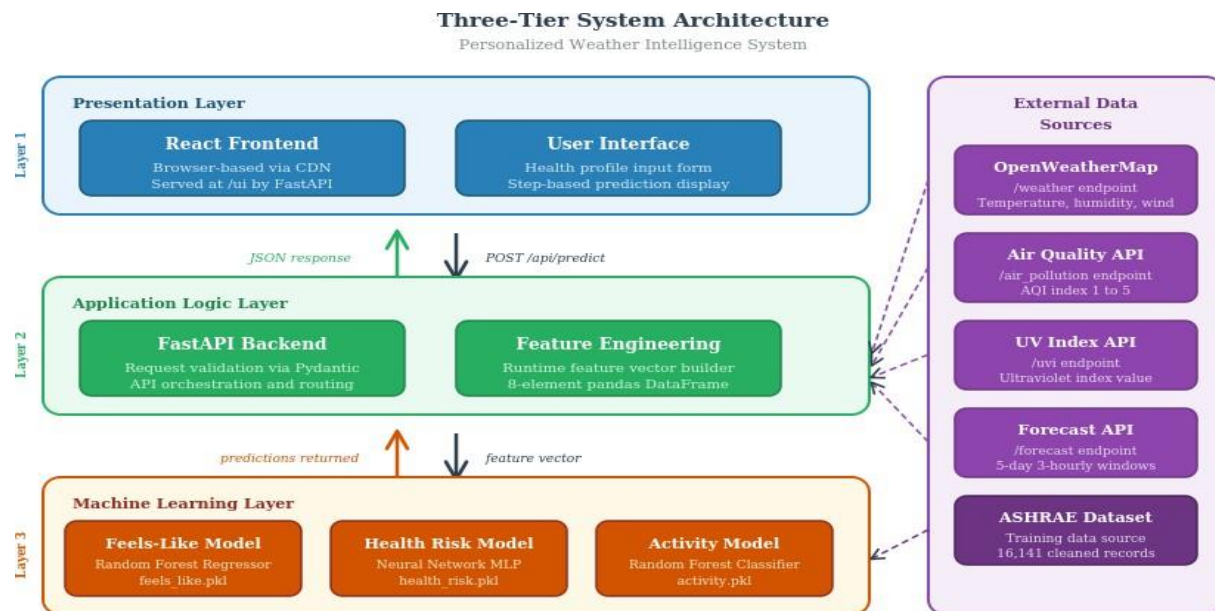


Figure 4.1: High-level architecture of the Personalised Weather Intelligence System

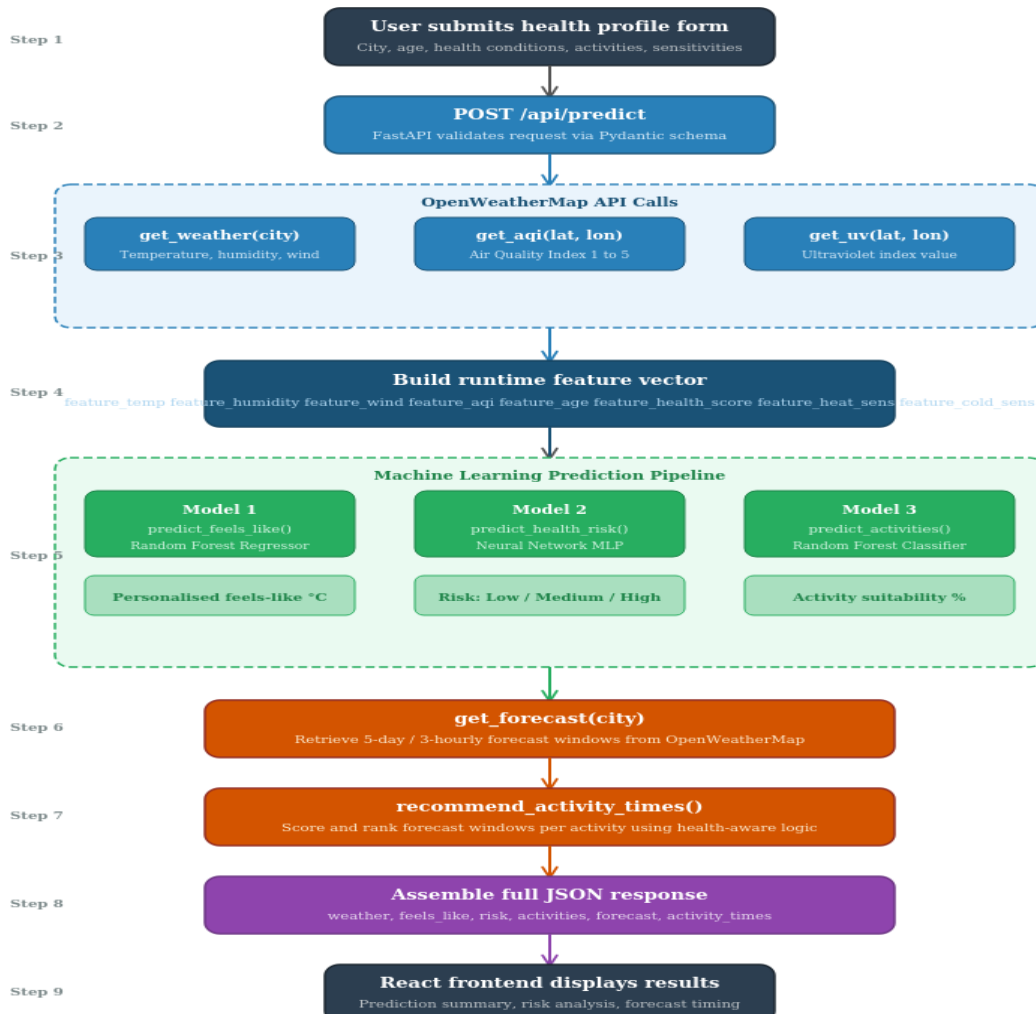


Figure 4.2: End-to-end data flow of the prediction pipeline

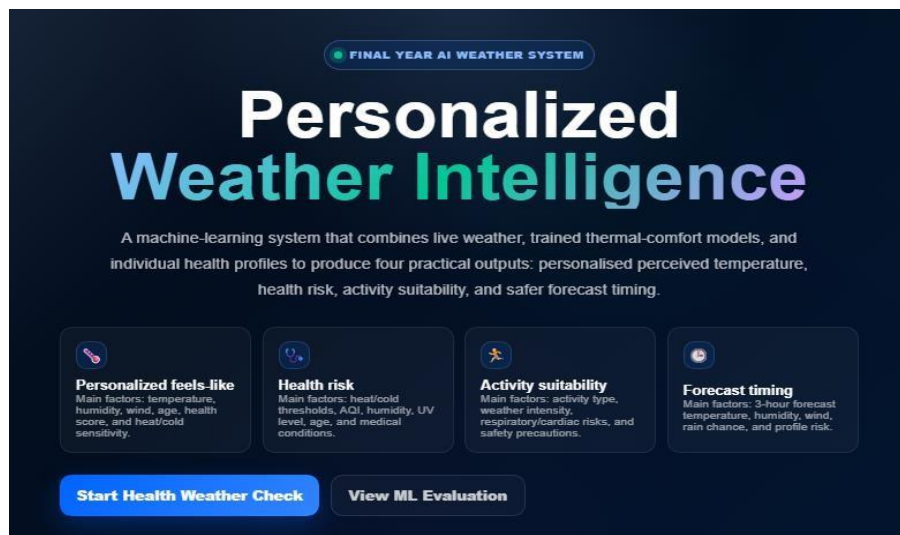


Figure 4.3: Landing page of the Personalised Weather Intelligence System showing the application introduction and entry point

Frontend Step Navigation Flow
 Personalized Weather Intelligence System — User Journey

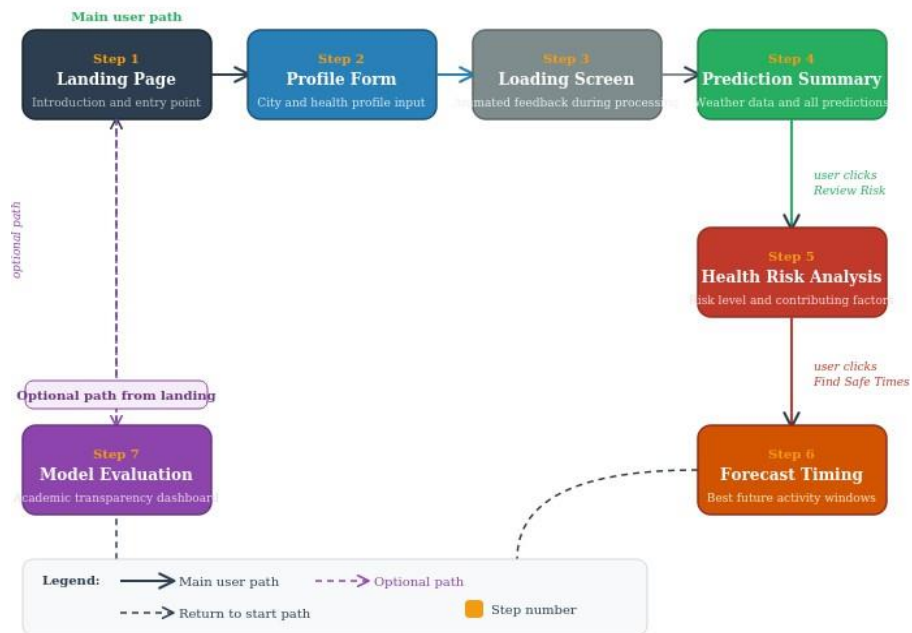


Figure 4.4: Frontend step navigation flow of the Personalised Weather Intelligence System showing the main user path, optional model evaluation route and return navigation

5. Implementation

The implementation follows the architecture described above. The project is organised into modules for API routing, configuration, weather retrieval, prediction functions, model training, serialized model artefacts and frontend delivery. Table 5.1 retains the file-level structure. Sensitive configuration, including the OpenWeatherMap API key, is loaded through environment variables using python-dotenv, following common API security practice and reducing the risk of accidental key exposure (OWASP, 2021).

The backend is implemented in api.py. The main POST /api/predict endpoint accepts a city and user profile, validates the body with a Pydantic schema, and orchestrates the full prediction workflow. Additional endpoints expose current weather data and model metrics. The /ui endpoint serves the React HTML file, while /docs provides FastAPI’s interactive documentation, shown in Figure 5.1. Error handling was implemented throughout the pipeline. If the weather API fails, the frontend receives a meaningful error instead of an unhandled exception. If forecast retrieval fails,

current predictions can still be returned with a forecast error field. If UV data is unavailable, the value defaults to zero so that the rest of the pipeline continues to operate.

Weather API integration is isolated in modules/weather_api.py. The get_weather function requests current conditions, extracts meteorological values and coordinates secondary air quality and UV retrievals. The get_forecast function retrieves three-hourly five-day forecast windows and converts each window into a simplified dictionary containing temperature, humidity, wind speed, weather condition and rain probability. Table 5.2 summarises the weather fields and their role in the prediction pipeline. This design keeps external API logic separate from prediction code, which improves testability and maintainability.

Runtime feature engineering is implemented in modules/predictions.py. The environmental features are obtained from the weather dictionary, while personal features are computed from the user profile. The health score is produced by summing weighted health-condition values, and sensitivity options are encoded as

Low = 0, Medium = 1 and High = 2. A key deployment detail is that the feature vector is constructed as a pandas DataFrame using the same column names and order saved during model training in `feature_cols.pkl`. This avoids inference-time feature-name mismatch warnings and ensures that scikit-learn receives the same input structure used during training (scikit-learn developers, 2024). Figure 5.2 preserves the runtime feature-engineering pipeline.

Three prediction functions implement the model logic. `predict_feels_like()` loads the Random Forest Regressor, obtains a raw prediction and applies the personalisation offset. `predict_health_risk()` loads the MLP and scaler, applies the same standardisation used during training, calculates confidence from class probabilities and applies the rule-based threshold validation layer. `predict_activities()` evaluates each selected activity, converts class probabilities into a suitability percentage and returns status, recommendation text, limiting factors, positive factors and safety precautions. Forecast recommendations are generated by `recommend_activity_times()`, which scores future forecast windows using the same health-aware logic and returns ranked time slots with explanatory notes. Figures 5.3 and 5.4 retain the forecast recommendation flow and profile input form.

Three implementation challenges were notable. First, scikit-learn warnings appeared when runtime data was provided as unnamed arrays; this was resolved by saving and reusing feature column names. Second, some locations or API tiers did not reliably return UV values; graceful fallback prevented the full request from failing. Third, public datasets do not contain condition-specific training labels for all target health conditions. The transparent rule-based offset and risk validation layers were therefore adopted as practical safeguards rather than pretending that the model had learned relationships not present in the training data.

The frontend implementation also reflects a research-prototype constraint. A production system would likely use a full build pipeline, asset bundling, automated tests, accessibility audits and offline handling. The prototype instead used a single HTML file with React, ReactDOM and Babel loaded from a CDN to reduce deployment friction. This choice made the system easier to demonstrate and evaluate because the whole application could run from the FastAPI server. The limitation is that first load depends on CDN availability and performance optimisations are minimal. In the context of a research prototype, the simplicity benefit outweighed the production disadvantages, but a future deployment would require migration to a conventional build and hosting workflow.

Table 5.1: Key project files and their roles

File or Directory	Role
<code>api.py</code>	FastAPI application, all endpoints and request orchestration
<code>config.py</code>	Environment variable loading and API URL configuration
<code>modules/weather_api.py</code>	All OpenWeatherMap API calls
<code>modules/predictions.py</code>	Runtime feature engineering and all prediction functions
<code>models/train_models.py</code>	Full CRISP-DM training pipeline
<code>models/feels_like.pkl</code>	Serialised Random Forest Regressor
<code>models/health_risk.pkl</code>	Serialised Neural Network MLP

models/activity.pkl	Serialised Random Forest Classifier
models/scaler.pkl	Standard scaler for the neural network path
models/feature_cols.pkl	Saved feature column names for runtime consistency
models/metrics.json	Saved evaluation metrics for the dashboard endpoint
frontend/simple/index.html	Complete React frontend served by FastAPI
data/db_measurements_v2.1.0.csv	ASHRAE training dataset

Table 5.2: Weather data fields retrieved and their role in the pipeline

Field	Source Endpoint	Role in System
Temperature in Celsius	/weather	feature_temp, offset input, risk threshold
Humidity as percentage	/weather	feature_humidity, risk threshold validation
Wind speed in metres per second	/weather	feature_wind, activity limiting factor
Latitude and Longitude	/weather	Secondary AQI and UV API calls
Air Quality Index 1 to 5	/air_pollution	feature_aqi, health risk threshold
UV Index	/uvi	Health risk threshold validation
Forecast windows	/forecast	Activity timing recommendation scoring

Weather Intelligence API 1.0.0 OAS 3.1

[/openapi.json](#)

default ^

- GET / Root v
- GET /ui React UI v
- GET /api/info Get Info v
- GET /api/weather/{city} Weather Endpoint v
- POST /api/predict Predict Endpoint v
- GET /api/metrics Get Metrics Endpoint v

Schemas ^

- HTTPValidationError > Expand all object
- PredictRequest > Expand all object
- ProfileModel > Expand all object
- ValidationError > Expand all object

Figure 5.1: FastAPI interactive documentation at /docs showing the available API endpoints including POST /api/predict

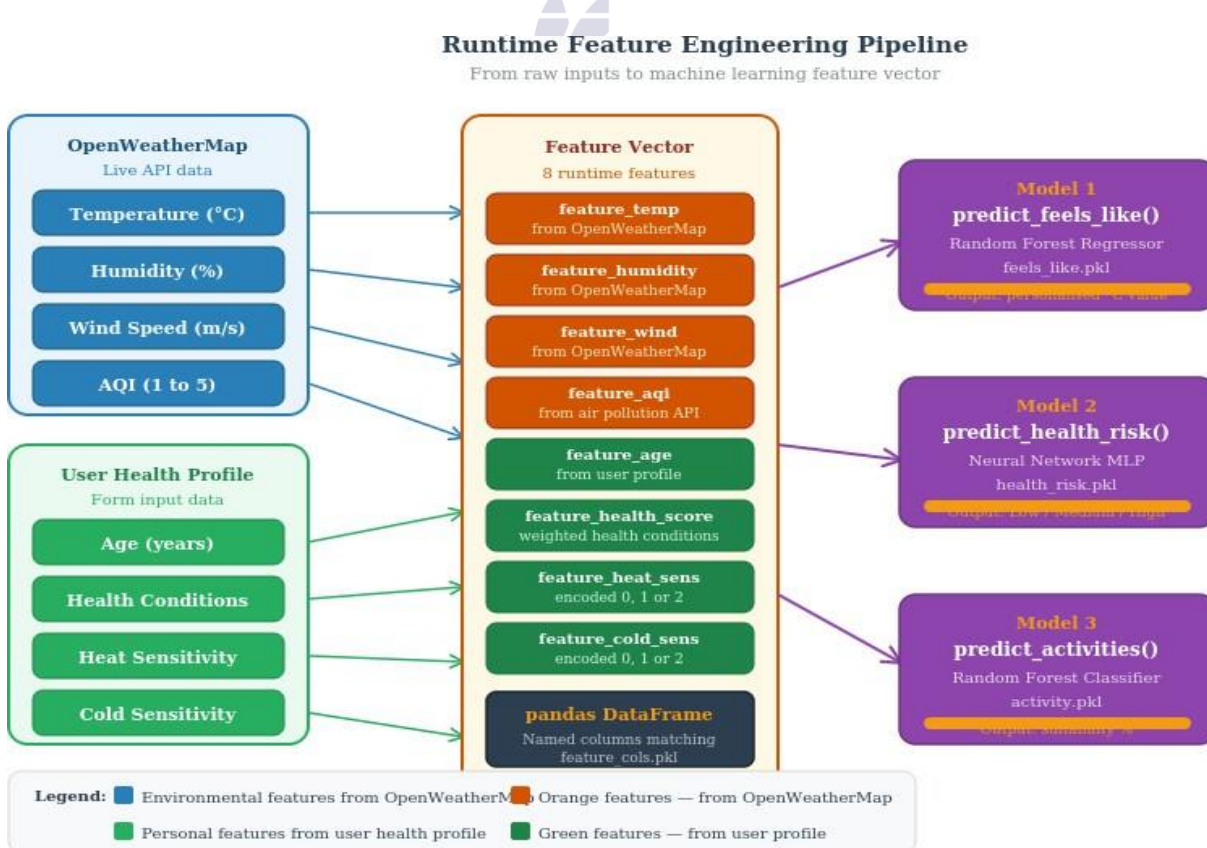


Figure 5.2: Runtime feature engineering pipeline from raw inputs to the machine learning feature vector



Figure 5.3: Forecast activity recommendation flow

Figure 5.4: Profile input form showing city field, age input and multi-select health conditions and activity preferences

6. Results, Findings and Evaluation

The ASHRAE dataset was loaded with 109,033 records and reduced to 16,141 complete records after removing rows with missing selected features. Although this was a substantial reduction, the cleaned data remained large

enough for stable training and cross-validation across the three tasks. The strict non-imputation policy was important because personal variables are central to the system's purpose. Imputing missing age, clothing insulation or metabolic rate values would risk smoothing away precisely the

individual variation that the system aims to recognise.

For perceived temperature prediction, six regression algorithms were evaluated using Mean Absolute Error and Root Mean Squared Error. The comparison is shown in Table 6.1 and Figure 6.1. Random Forest achieved the strongest performance with MAE = 0.216°C and RMSE = 0.367°C. The result is consistent with Breiman's

(2001) explanation of Random Forest's strength as an ensemble of decorrelated decision trees and with studies showing strong ensemble performance on tabular thermal comfort data (Luo et al., 2020; Sarkar et al., 2024). Five-fold cross-validation produced MAE = 0.314°C with standard deviation 0.029, indicating reasonable generalisation.

Table 6.1: Algorithm comparison for perceived temperature prediction

Algorithm	MAE (°C)	RMSE (°C)
Linear Regression	0.263	0.438
Decision Tree	0.257	0.427
Gradient Boosting	0.234	0.379
XGBoost	0.236	0.373
CatBoost	0.241	0.375
Random Forest	0.216	0.367



Figure 6.1: Algorithm comparison for perceived temperature prediction MAE by algorithm

For activity suitability classification, six algorithms produced tightly clustered results, as shown in Table 6.2 and Figure 6.2. Weighted F1 scores ranged from 0.966 to 0.968, suggesting that the task was well structured and the available features captured the main relationships needed

for prediction. Random Forest was selected with weighted F1 = 0.968 and accuracy = 0.965. Cross-validation produced F1 = 0.950 with standard deviation 0.009, which supports the stability of the model. The output interface, retained in Figure 6.3, presents suitability percentages,

recommendation status and safety precautions rather than only a categorical label.

Table 6.2: Algorithm comparison for activity suitability classification

Algorithm	Weighted F1	Accuracy
Decision Tree	0.967	0.965
Gradient Boosting	0.966	0.964
XGBoost	0.967	0.964
LightGBM	0.967	0.964
CatBoost	0.966	0.964
Random Forest	0.968	0.965



Figure 6.2: Algorithm comparison for activity suitability classification weighted F1 score by algorithm

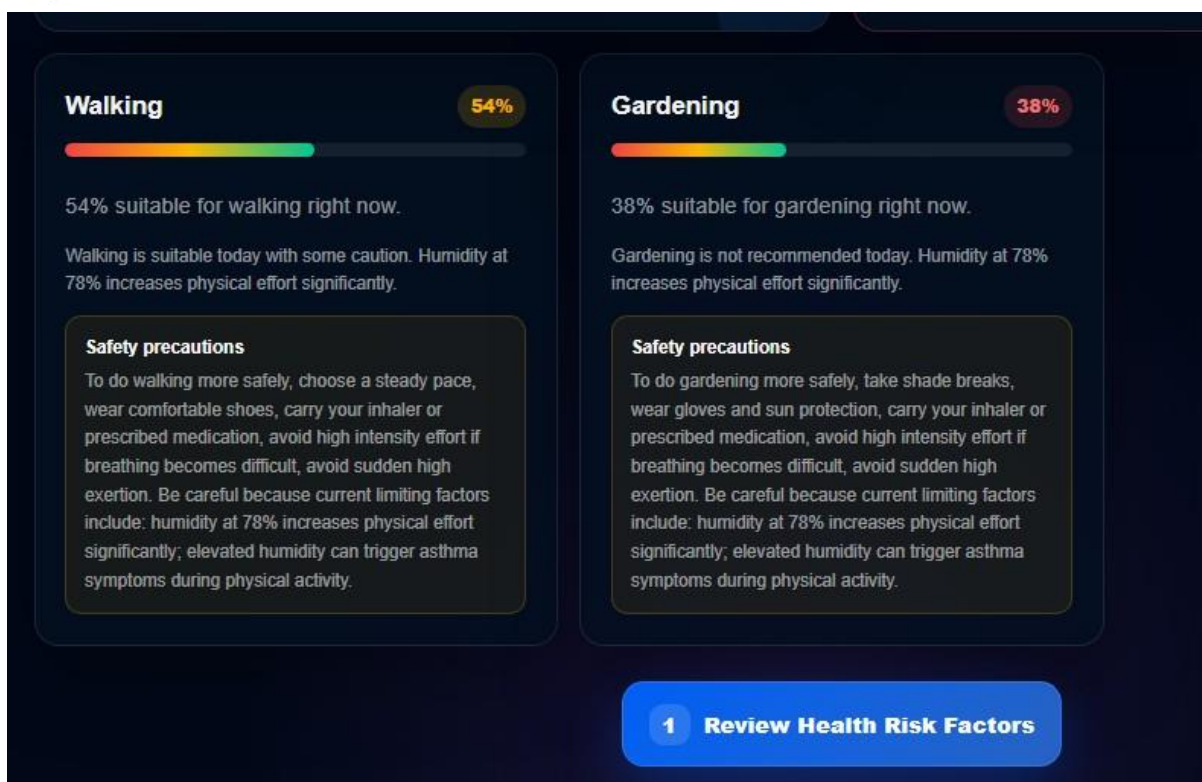


Figure 6.3: Activity suitability output cards showing suitability percentage, recommendation status and safety precautions for selected outdoor activities

Health-risk classification was the most safety-relevant task. Eight algorithms were compared, and the results appear in Table 6.3 and Figure 6.4. Random Forest achieved the highest raw accuracy at 0.995, but the Neural Network MLP was selected with accuracy = 0.992 because the task involves interacting environmental and personal variables that are conceptually suited to non-linear layered representation (Goodfellow et al., 2016). This selection was methodologically

defensible because the difference in overall accuracy was small. However, Figure 6.5 shows that per-class performance is more important than headline accuracy. The MLP's high-risk recall was 0.719, meaning that approximately 28 percent of true high-risk situations in the test set were not classified as high risk by the model alone. In a health-aware system, this is the most significant technical limitation.

Table 6.3: Algorithm comparison for health risk classification

Algorithm	Weighted F1	Accuracy
Logistic Regression	0.978	0.978
Decision Tree	0.994	0.994
Gradient Boosting	0.993	0.993
XGBoost	0.994	0.994
LightGBM	0.994	0.993
CatBoost	0.994	0.994

Random Forest	0.995	0.995
Neural Network MLP	0.992	0.992



Figure 6.4: Algorithm comparison for health risk classification accuracy by algorithm

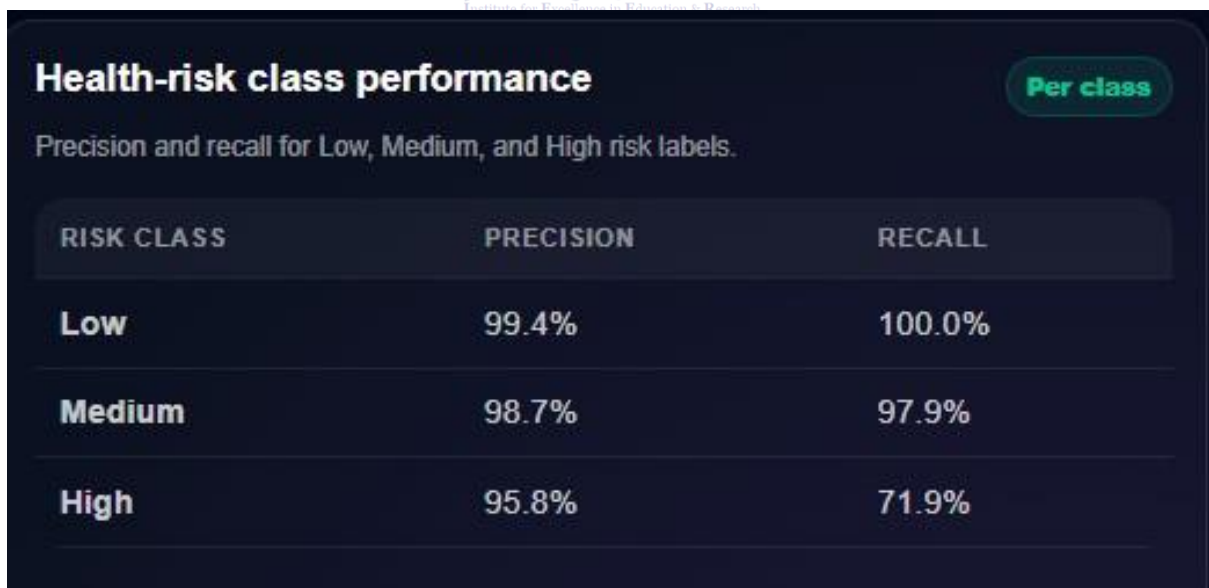


Figure 6.5: Per-class precision and recall for the Neural Network MLP health risk classifier

Feature importance analysis for the perceived temperature Random Forest model showed that temperature dominated prediction, followed by

humidity and age, as summarised in Table 6.4 and Figure 6.6. This pattern is physically plausible and aligns with thermal comfort

literature (Luo et al., 2020; Zhang et al., 2024). The relatively lower importance of sensitivity variables reinforces the need for the rule-based personalisation offset. The model cannot learn

detailed sensitivity effects from training labels that do not explicitly encode those health characteristics.

Table 6.4: Feature importance ranking for the Random Forest perceived temperature model

Rank	Feature	Relative Importance
1	feature_temp	High
2	feature_humidity	Medium-High
3	feature_age	Medium
4	feature_wind	Medium
5	feature_health_score	Low-Medium
6	feature_aqi	Low
7	feature_heat_sens	Low
8	feature_cold_sens	Low



Figure 6.6: Feature importance for the Random Forest perceived temperature model

Evaluation against objectives is summarised in Table 6.5. Six objectives were fully achieved. The system successfully combines weather and health

profile data, compares algorithms systematically, predicts perceived temperature, assesses activity suitability, recommends forecast time windows

and implements a working FastAPI/React prototype. The critical evaluation objective was also achieved because the limitations are explicitly identified. The health-risk objective was only partially achieved: despite high overall

accuracy, high-risk recall is not adequate for independent safety-critical deployment. Figures 6.7, 6.8 and 6.9 retain the prediction summary, health-risk analysis and forecast activity timing screens.

Table 6.5: Research objectives achievement summary

Objective	Status	Evidence
1. Combine weather data and health profile	Fully achieved	Four personalised outputs produced per request
2. Systematic algorithm comparison	Fully achieved	6 to 8 algorithms compared per task with consistent protocol
3. Perceived temperature regression	Fully achieved	MAE 0.216°C, CV MAE 0.314°C
4. Health risk classification	Partially achieved	Accuracy 0.992 but high-risk recall 0.719
5. Activity suitability classification	Fully achieved	Weighted F1 0.968, CV F1 0.950
6. Forecast recommendation component	Fully achieved	Ranked windows returned with health context
7. FastAPI and React implementation	Fully achieved	Running system accessible at /ui
8. Critical evaluation and appraisal	Fully achieved	Addressed in this chapter

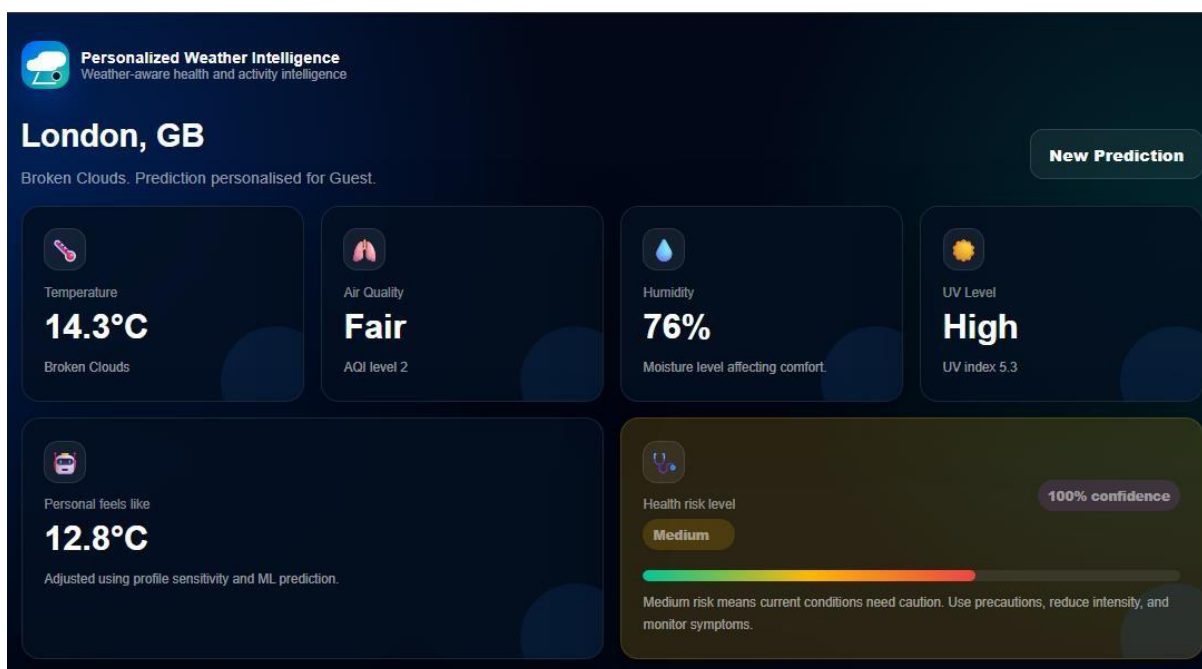


Figure 6.7: Prediction summary page showing current weather conditions, personalised feels-like temperature, health risk level and activity suitability cards

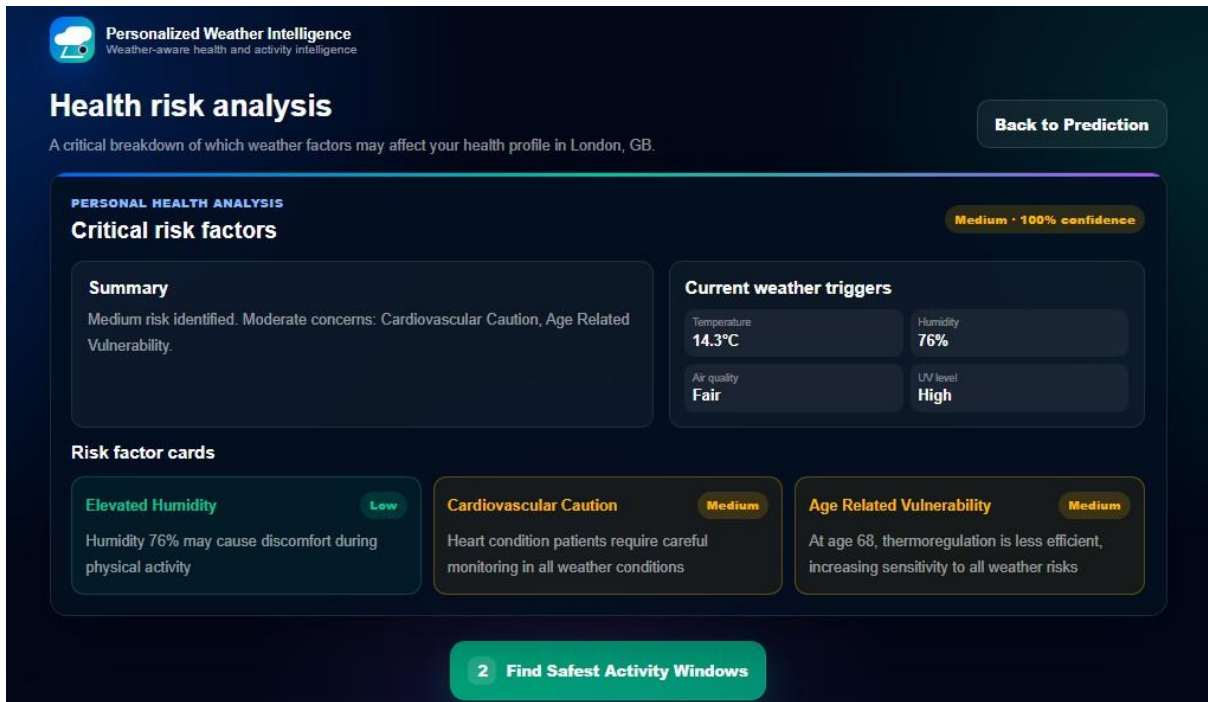


Figure 6.8: Health risk analysis page showing risk level, confidence score, weather triggers and contributing risk factor breakdown



Figure 6.9: Forecast activity timing page showing ranked future time windows with suitability percentages and health guidance

Functional testing covered valid and invalid city inputs, standard and vulnerable health profiles, multiple activity selections, forecast access, metrics dashboard rendering, UV endpoint failure and forecast retrieval failure. All nine tests passed, as shown in Table 6.7. The model

evaluation dashboard and error-handling screen are retained in Figures 6.10 and 6.11. These results show that the prototype is functionally stable for demonstration and evaluation, even though it is not ready for real-world deployment as a regulated health product.

Table 6.7: Functional testing

Test Case	Input	Expected Output	Result
Valid city, standard profile	London, age 25, no conditions	Full prediction response returned	Pass
Valid city, multiple conditions	Manchester, age 68, heart condition and asthma	Elevated risk, adjusted feels-like	Pass
Invalid city name	Nonexistentcity123	Error message returned to frontend	Pass
High heat sensitivity	Dubai, age 35, high heat sensitivity	Higher feels-like than standard profile	Pass
Multiple activities selected	Walking, Running, Cycling	Three separate structured activity results	Pass
Forecast endpoint accessible	London	Ranked forecast windows returned	Pass
Model evaluation dashboard	GET /api/metrics	Metrics JSON returned and rendered	Pass
UV endpoint unavailable	Simulated API failure	System defaults to zero, no crash	Pass
Forecast retrieval failure	Simulated forecast failure	Current predictions still returned with error field	Pass

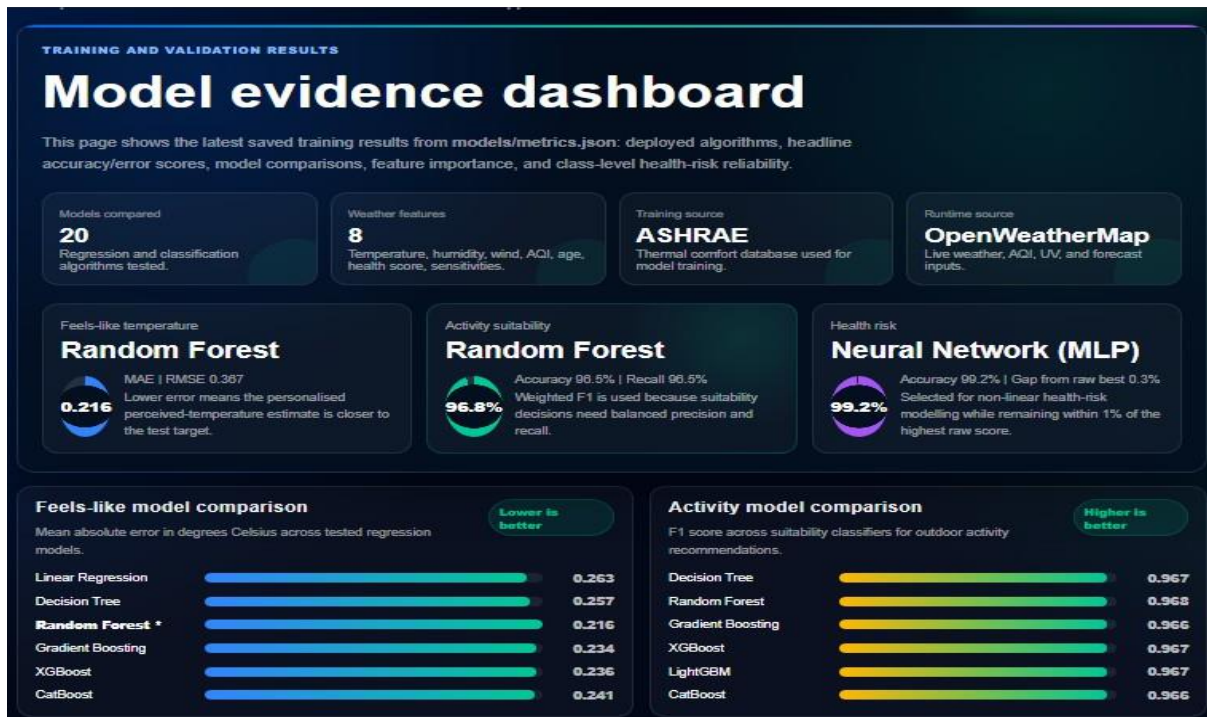


Figure 6.10: Model evaluation dashboard showing algorithm comparison graphs and training metrics for all three models

Tell us about your weather needs

Your profile is used to personalise feels-like temperature, risk level, and activity suitability.

City 'Nonexistentcity123' not found. Please check the spelling.

City: Nonexistentcity123

Name: Guest | Age: 68

Heat sensitivity: High | Cold sensitivity: Medium

Health conditions: None, Asthma, Heart condition, Diabetes, Arthritis, Respiratory illness

Activity preferences: Walking, Running, Cycling, Swimming, Gardening

Figure 6.11: System error handling showing a meaningful error message returned when an invalid city name is submitted

The commercial risk assessment is significant because the system occupies a boundary between consumer software and health-related decision support. If framed carelessly, a personalised weather recommendation system could be

misinterpreted as clinical advice. If it stored longitudinal health profiles, it would introduce stronger privacy, security and governance obligations. If it were marketed as preventing harm or managing disease, regulators could treat

it more like a medical device. The study addresses these risks by presenting the prototype as advisory, non-diagnostic and non-persistent. Nonetheless, the risk table shows that technical performance is only one condition for viability. A responsible future version would need legal review, professional input, user testing, robust security engineering and alternative weather provider strategies.

The limitations define the boundary of the findings. The largest data limitation is the use of indoor-focused ASHRAE records for an outdoor weather recommendation problem. Outdoor environments contain solar radiation, rain, radiant surface heat, variable gusts and physical activity effects that indoor datasets cannot fully capture. Future outdoor datasets such as field-based park or public-space studies would improve external validity (Fan et al., 2023; Tian et al., 2025). The largest model limitation is the high-risk recall of 0.719. Future work should use dedicated high-risk data collection, Synthetic Minority Oversampling Technique, cost-sensitive learning and expert review of risk thresholds (Togunwa et al., 2023). The largest evaluation limitation is the absence of a formal user study with older adults or people with chronic health conditions. Such a study is essential to test whether recommendations are understandable, accessible and correctly interpreted as advisory rather than clinical.

Legal, ethical and commercial risks were also assessed. Health profile inputs are sensitive, and GDPR requires special protection for health-

related data (European Parliament, 2016). The prototype reduces this risk by not persistently storing profiles, but future versions with accounts or history would require privacy impact assessment, consent flows and stronger technical safeguards. Because the system makes health-related recommendations, disclaimers and user education are essential (Hatherley, 2024). Commercial risks include API pricing changes, ASHRAE licensing constraints, user misinterpretation, possible medical-device classification, data breach risk and competition from large platforms. These risks and mitigations are retained in Table 6.9.

Taken together, the results show a prototype that is stronger as a research contribution than as a deployable product. Its contribution is strongest where it demonstrates integration: live API retrieval, feature engineering, multiple trained models, rule-based adjustment, frontend presentation and evaluation dashboards operate as a coherent system. Its limitations are strongest where the target context requires evidence that was not available within the project timeframe: outdoor-specific data, clinical threshold validation and real user interpretation studies.

This distinction is important because it avoids overstating the achievement. The system proves that personalised weather intelligence can be built and evaluated, but it also identifies the exact areas that must be solved before claims of safety, reliability or commercial readiness would be justified.

Table 6.9: Commercial and business risk assessment

Risk	Likelihood	Impact	Mitigation
OpenWeatherMap API pricing change or discontinuation	Medium	High	Integrate multiple weather data providers
ASHRAE dataset licence restrictions on commercial use	Medium	High	Commission dedicated outdoor comfort dataset
User misinterpretation of health recommendations	Medium	High	Strengthen disclaimers, add professional review
Regulatory classification	Low	Very High	Maintain strict decision-

as a medical device			support framing
Data breach exposing user health information	Low	Very High	End-to-end encryption, no persistent storage
Competition from large platforms adding health features	High	Medium	Focus on depth of personalization as differentiator

7. Conclusion and Future Work

Three principal findings emerge. First, personalised weather intelligence is technically feasible. The system can combine live weather data, user health profiles and trained models to produce differentiated recommendations for users in the same city under the same weather conditions. Second, the ASHRAE Global Thermal Comfort Database II is a useful but imperfect foundation. It supports modelling because it is large and contains both personal and environmental variables, but its indoor orientation limits outdoor validity. Third, hybrid decision logic is necessary in health-aware machine learning systems. Pure model prediction produced high overall accuracy, but high-risk recall revealed a safety gap. The rule-based validation layer does not eliminate that limitation, but it provides a transparent safeguard consistent with Shortliffe and Sepúlveda's (2018) recommendation that safety-relevant decision support should combine data-driven and expert-informed logic.

A second practical finding concerns communication. The system does not merely predict values; it has to present them in a way that supports action. A personalised perceived temperature is useful only if the user understands that it reflects health and sensitivity adjustments. A health-risk label is useful only if the factors that triggered it are visible. An activity suitability score is useful only if accompanied by precautions and alternative timing. This is why the prototype emphasises explanation, limiting factors and forecast windows rather than presenting model outputs alone. Future evaluation should test these explanations with target users, but the design already recognises that interpretability is part of usability, not an optional add-on.

The study therefore demonstrates a meaningful extension of conventional weather applications. Instead of giving all users the same forecast interpretation, the prototype shows how meteorological data can be contextualised through personal health information and activity goals. The Random Forest perceived-temperature model achieved a very low test error; the activity suitability model provided stable classifications; and the MLP health-risk model achieved high overall accuracy while exposing an important class-specific weakness. The practical value of the work lies not only in the working system but also in the evidence-based boundary it establishes: personalised environmental decision support is possible, but it must be evaluated at the level of vulnerable classes, not only average metrics.

Future work should prioritise four areas. The first is improving high-risk recall through better representation of extreme weather and vulnerable users in the training data. The second is retraining and validating models on outdoor thermal comfort datasets. The third is conducting formal usability testing with health-vulnerable participants, including think-aloud tasks and post-task questionnaires. The fourth is strengthening governance for any persistent data storage, including privacy-by-design, consent management and independent review. In its current form, the system is best understood as a robust academic prototype and proof of concept. With better outdoor data, clinical review, user-centred evaluation and stronger operational safeguards, it could become a foundation for safer, more personalised weather intelligence.

The article also highlights a broader lesson for applied machine learning: model choice must be interpreted in context. The Random Forest was selected for two tasks because it delivered the strongest or most stable empirical results and is

well suited to structured tabular data. The MLP was selected for health risk because it was conceptually appropriate for multi-factor interactions, but its high-risk recall exposed the danger of relying on model architecture arguments without class-level evaluation. The most defensible conclusion is therefore not that one algorithm is universally superior, but that health-aware systems require a combination of empirical comparison, conceptual reasoning, error analysis and transparent safeguards.

References

- Abd El-khalik, W. (2022) 'A machine learning approach for improved thermal comfort prediction in sustainable built environments', *Sustainable Machine Intelligence Journal*, 1. Available at: <https://doi.org/10.55708/js0101004> (Accessed: 14 March 2025).
- Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M. and Steggle, P. (1999) 'Towards a better understanding of context and context-awareness', in *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing*, Karlsruhe, Germany, September 1999. Berlin: Springer, pp. 304–307.
- Achebak, H., Rey, G., Lloyd, S.J., Quijal-Zamorano, M., Méndez-Turrubiates, R.F. and Ballester, J. (2024) 'Ambient temperature and risk of cardiovascular and respiratory adverse health outcomes: a nationwide cross-sectional study from Spain', *European Journal of Preventive Cardiology*, 31(9), pp. 1080–1089. Available at: <https://doi.org/10.1093/eurjpc/zwae021> (Accessed: 15 March 2025).
- ASHRAE (2020) *ASHRAE Standard 55: Thermal Environmental Conditions for Human Occupancy*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers.
- Ayman, N., Alaa, S., Hussein, M. and Hamdi, A. (2025) 'Personalised and safe route planning for asthma patients using real-time environmental data', arXiv preprint, arXiv:2501.10372. Available at: <https://arxiv.org/abs/2501.10372> (Accessed: 17 March 2025).
- Bass, L., Clements, P. and Kazman, R. (2021) *Software Architecture in Practice*. 4th edn. Boston: Addison-Wesley.
- BCS (2022) *BCS Code of Conduct*. Swindon: British Computer Society. Available at: <https://www.bcs.org/membership-and-registrations/become-a-member/bcs-code-of-conduct/> (Accessed: 15 March 2025).
- Breiman, L. (2001) 'Random forests', *Machine Learning*, 45(1), pp. 5–32. Available at: <https://doi.org/10.1023/A:1010933404324> (Accessed: 13 March 2025).
- Cappelli, F., Tauro, F., Apollonio, C., Petroselli, A., Borgonovo, E. and Grimaldi, S. (2024) 'Random forest and feature importance measures for discriminating the most influential environmental factors in predicting cardiovascular and respiratory diseases', *International Journal of Environmental Research and Public Health*, 21(7), p. 867. Available at: <https://doi.org/10.3390/ijerph21070867> (Accessed: 16 March 2025).
- Chen, T. and Guestrin, C. (2016) 'XGBoost: a scalable tree boosting system', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, California, August 2016. New York: ACM, pp. 785–794. Available at: <https://doi.org/10.1145/2939672.2939785> (Accessed: 14 March 2025).
- Conduah, A.K., Ofoe, S. and Siaw-Marfo, D. (2025) 'Data privacy in healthcare: global challenges and solutions', *Digital Health*, 11. Available at: <https://doi.org/10.1177/20552076251343959> (Accessed: 17 March 2025).

- Cosma, A.C. and Simha, R. (2019) 'Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions', *Building and Environment*, 148, pp. 372–383. Available at: <https://doi.org/10.1016/j.buildenv.2018.11.017> (Accessed: 14 March 2025).
- Creswell, J.W. and Creswell, J.D. (2018) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 5th edn. London: Sage Publications.
- Environment and Climate Change Canada (2001) *Wind Chill Index*. Ottawa: Government of Canada. Available at: <https://www.canada.ca/en/environment-climate-change/services/weather-general-tools-resources/wind-chill-cold-weather.html> (Accessed: 13 March 2025).
- European Parliament (2016) Regulation (EU) 2016/679 of the European Parliament and of the Council – General Data Protection Regulation. Brussels: Official Journal of the European Union. Available at: <https://gdpr-info.eu/> (Accessed: 15 March 2025).
- Fan, Y., Wang, J., Obradovich, N. and Zheng, S. (2023) 'Intraday adaptation to extreme temperatures in outdoor activity', *Scientific Reports*, 13, p. 445. Available at: <https://doi.org/10.1038/s41598-022-26928-y> (Accessed: 16 March 2025).
- Fang, W., Li, Z., Gao, J., Meng, R., He, G., Hou, Z., Zhu, S., Zhou, M., Zhou, C. and Xiao, Y. (2023) 'The joint and interaction effect of high temperature and humidity on mortality in China', *Environment International*, 171, p. 107669. Available at: <https://doi.org/10.1016/j.envint.2022.107669> (Accessed: 16 March 2025).
- Fanger, P.O. (1970) *Thermal Comfort: Analysis and Applications in Environmental Engineering*. Copenhagen: Danish Technical Press.
- Földváry Ličina, V., Cheung, T., Zhang, H., de Dear, R., Parkinson, T., Arens, E., Chun, C., Schiavon, S., Luo, M., Brager, G., Li, P., Kaam, S., Adebamowo, S.N., Andamon, M.M., Babiak, J., Bako-Biro, Z., Bauman, F., Bednar, T., Brendon, T. and Zhai, Y. (2018) 'Development of the ASHRAE Global Thermal Comfort Database II', *Building and Environment*, 142, pp. 502–512. Available at: <https://doi.org/10.1016/j.buildenv.2018.06.022> (Accessed: 13 March 2025).
- Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. Cambridge, MA: MIT Press. Available at: <https://www.deeplearningbook.org/> (Accessed: 14 March 2025).
- Haque, R., Ho, S.B., Chai, I., Abdullah, A., Sim, K.Y., Hoo, M.H. and Elshaw, M. (2021) 'Optimised deep neural network model to predict asthma exacerbation based on personalised weather triggers', *F1000Research*, 10, p. 909. Available at: <https://doi.org/10.12688/f1000research.73026.1> (Accessed: 16 March 2025).
- Hatherley, J. (2024) 'Are clinicians ethically obligated to disclose their use of medical machine learning systems to patients?', *Journal of Medical Ethics*. Available at: <https://doi.org/10.1136/jme-2024-109905> (Accessed: 17 March 2025).
- Humphreys, M.A. and Nicol, J.F. (2002) 'The validity of ISO-PMV for predicting comfort votes in every-day thermal environments', *Energy and Buildings*, 34(6), pp. 667–684. Available at: [https://doi.org/10.1016/S0378-7788\(02\)00018-X](https://doi.org/10.1016/S0378-7788(02)00018-X) (Accessed: 13 March 2025).
- JetBrains (2024) *Python Developers Survey 2023: Results*. Prague: JetBrains. Available at: <https://www.jetbrains.com/lp/python-developers-survey-2023/> (Accessed: 15 March 2025).

- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y. (2017) 'LightGBM: a highly efficient gradient boosting decision tree', in *Advances in Neural Information Processing Systems* 30, Long Beach, California, December 2017. New York: Curran Associates, pp. 3146–3154.
- Kenney, W.L., Craighead, D.H. and Alexander, L.M. (2014) 'Heat waves, aging, and human cardiovascular health', *Medicine and Science in Sports and Exercise*, 46(10), pp. 1891–1899. Available at: <https://doi.org/10.1249/MSS.0000000000000325> (Accessed: 15 March 2025).
- Kolyshkina, I. and Simoff, S. (2021) 'Interpretability of machine learning solutions in public healthcare: the CRISP-ML approach', *Frontiers in Big Data*, 4, p. 660206. Available at: <https://doi.org/10.3389/fdata.2021.660206> (Accessed: 14 March 2025).
- Kovats, R.S. and Hajat, S. (2008) 'Heat stress and public health: a critical review', *Annual Review of Public Health*, 29, pp. 41–55. Available at: <https://doi.org/10.1146/annurev.publhealth.29.020907.090843> (Accessed: 15 March 2025).
- Kyung, N., Chan, J., Lim, S. and Lee, B. (2023) 'Contextual targeting in mHealth apps: harnessing weather information and message framing to increase physical activity', *Information Systems Research*, 35(3), pp. 1034–1051. Available at: <https://doi.org/10.1287/isre.2020.0119> (Accessed: 16 March 2025).
- Luo, M., Xie, J., Yan, Y., Ke, Z., Yu, P., Wang, Z. and Zhang, J. (2020) 'Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II', *Energy and Buildings*, 210, p. 109776. Available at: <https://doi.org/10.1016/j.enbuild.2020.109776> (Accessed: 13 March 2025).
- Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M.J. and Flach, P. (2021) 'CRISP-DM twenty years later: from data mining processes to data science trajectories', *IEEE Transactions on Knowledge and Data Engineering*, 33(8), pp. 3048–3061. Available at: <https://doi.org/10.1109/TKDE.2019.2962680> (Accessed: 14 March 2025).
- Obermeyer, Z. and Emanuel, E.J. (2016) 'Predicting the future: big data, machine learning, and clinical medicine', *New England Journal of Medicine*, 375(13), pp. 1216–1219. Available at: <https://doi.org/10.1056/NEJMp1606181> (Accessed: 15 March 2025).
- OWASP (2021) OWASP Top Ten: A02 Cryptographic Failures – Sensitive Data Exposure. Bel Air: Open Web Application Security Project. Available at: <https://owasp.org/Top10/> (Accessed: 16 March 2025).
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V. and Gulin, A. (2018) 'CatBoost: unbiased boosting with categorical features', in *Advances in Neural Information Processing Systems* 31, Montreal, Canada, December 2018. New York: Curran Associates, pp. 6638–6648.
- Public Health England (2021) Heatwave Plan for England: Protecting Health and Reducing Harm from Severe Heat and Heatwaves. London: PHE. Available at: <https://www.gov.uk/government/publications/heatwave-plan-for-england> (Accessed: 16 March 2025).
- Ramírez, S. (2021) FastAPI Documentation. Available at: <https://fastapi.tiangolo.com/> (Accessed: 15 March 2025).
- Rogers, Y., Sharp, H. and Preece, J. (2023) *Interaction Design: Beyond Human-Computer Interaction*. 6th edn. Chichester: Wiley.

- Rosser, F., Han, Y.Y., Rothenberger, S.D., Forno, E., Mair, C. and Celedón, J.C. (2022) 'Air quality index and emergency department visits and hospitalizations for childhood asthma', *Annals of the American Thoracic Society*, 19(7), pp. 1139–1148. Available at: <https://doi.org/10.1513/AnnalsATS.202105-539OC> (Accessed: 16 March 2025).
- Sarkar, S., Ahmed, T. and Islam, R. (2024) 'Machine learning thermal comfort prediction models based on occupant demographic characteristics', *Energy and Buildings*, 316, p. 114379. Available at: <https://doi.org/10.1016/j.enbuild.2024.114379> (Accessed: 14 March 2025).
- Schröer, C., Kruse, F. and Gómez, J.M. (2021) 'A systematic literature review on applying CRISP-DM process model', *Procedia Computer Science*, 181, pp. 526–534. Available at: <https://doi.org/10.1016/j.procs.2021.01.199> (Accessed: 14 March 2025).
- scikit-learn developers (2024) *Scikit-learn: Machine Learning in Python – User Guide*. Available at: https://scikit-learn.org/stable/user_guide.html (Accessed: 15 March 2025).
- Shimaoka, P., Dias, T. and Carvalho, A. (2023) 'The evolution of CRISP-DM for data science: methods, processes and frameworks', *Journal of Information Systems Engineering and Management*, 8(2). Available at: <https://doi.org/10.55267/iadt.07.13215> (Accessed: 14 March 2025).
- Shortliffe, E.H. and Sepúlveda, M.J. (2018) 'Clinical decision support in the era of artificial intelligence', *JAMA*, 320(21), pp. 2199–2200. Available at: <https://doi.org/10.1001/jama.2018.17163> (Accessed: 15 March 2025).
- Sokolova, M. and Lapalme, G. (2009) 'A systematic analysis of performance measures for classification tasks', *Information Processing and Management*, 45(4), pp. 427–437. Available at: <https://doi.org/10.1016/j.ipm.2009.03.002> (Accessed: 16 March 2025).
- Ssebyala, S.N., Kintu, T.M., Muganzi, D.J., Dresser, C., Demetres, M.R., Lai, Y. and Doshi, R.H. (2024) 'Use of machine learning tools to predict health risks from climate-sensitive extreme weather events: a scoping review', *PLOS Climate*, 3(1), p. e0000338. Available at: <https://doi.org/10.1371/journal.pclm.0000338> (Accessed: 15 March 2025).
- Steadman, R.G. (1984) 'A universal scale of apparent temperature', *Journal of Applied Meteorology and Climatology*, 23(12), pp. 1674–1687. Available at: [https://doi.org/10.1175/1520-0450\(1984\)023](https://doi.org/10.1175/1520-0450(1984)023) (Accessed: 13 March 2025).
- Tian, Z., Chen, Y., Wu, R. and Wang, L. (2025) 'Machine learning-based prediction of thermal comfort: exploring building types, climate, ventilation strategies, and seasonal variations', *Building Research and Information*, 54(1), pp. 100–117. Available at: <https://doi.org/10.1080/09613218.2025.2462932> (Accessed: 14 March 2025).
- Togunwa, T.O., Babatunde, A.O. and Abdullah, K. (2023) 'Deep hybrid model for maternal health risk classification in pregnancy: synergy of ANN and random forest', *Frontiers in Artificial Intelligence*, 6, p. 1213436. Available at: <https://doi.org/10.3389/frai.2023.1213436> (Accessed: 16 March 2025).
- Wang, H., Song, Y. and Bi, H. (2025) 'Optimizing public health management with predictive analytics: leveraging the power of random forest', *Frontiers in Big Data*, 8, p. 1574683. Available at: <https://doi.org/10.3389/fdata.2025.1574683> (Accessed: 16 March 2025).

- World Health Organisation (2021) Climate Change and Health. Geneva: WHO. Available at: <https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health> (Accessed: 14 March 2025).
- Zhang, Y., Li, X., Wang, Z. and Chen, H. (2024) 'Advancing personal thermal comfort prediction: a data-driven framework integrating environmental and occupant dynamics using machine learning', *Building and Environment*, 261, p. 111736. Available at: <https://doi.org/10.1016/j.buildenv.2024.111736> (Accessed: 14 March 2025).

