HealthEngine: An Integrated Healthcare Analytics Model Using Multimodal Transformer, Deep Multitask Neural Networks, and SHAP (Supplementary Materials)

Moshedayan Sirapangi, Gopikrishnan Sundaram, Norbert Herencsar, *Senior Member, IEEE,* Meng Li, *Senior Member, IEEE,* and Gautam Srivastava, *Senior Member, IEEE*

***Abstract*—The development of more accurate and explainable health predictions is paramount in view of the increasing prevalence rates of chronic diseases and mental illnesses. Existing methods often under-utilize the rich, heterogeneous data streams coming from wearable devices, environmental sensors, and behavioural data, and hence fall short in making predictions that are both accurate and actionable. Most models lack transparency and cannot avoid privacy concerns due to the samples used from sensitive health data. This study specifically addresses the challenge of building a predictive healthcare framework capable of handling multimodal data, data streams that originate from different modalities (physiological, behavioural, and environmental), and exhibit intrinsic heterogeneity. Unlike general heterogeneous datasets, multimodal health data demand models that can effectively integrate structured, semi-structured, and temporal information while ensuring privacy and interpretability. In this work, we propose an integrated comprehensive multimodal health prediction framework with five advanced methods, namely Multimodal Transformer Networks (MTN), Deep Multitask Neural Networks (DMNN), SHapley Additive exPlanations (SHAP) based explainability, Federated Learning, and Bayesian Neural Networks (BNN). MTN uses attention mechanisms to fuse different data modalities and captures cross-modal dependencies effectively, achieving an improvement of 8 to 12% in prediction accuracy. DMNN leverages multitask learning to share knowledge between related health prediction tasks, reducing error rates by 10–15%. SHAP is used to provide localized, patient-specific explanations that improve clinical trust by up to 85%. This is done by privately training the models on decentralized data sets, which provides results without more than a 3% drop in precision compared to centralized models. Third, BNNs are used to quantify uncertainty in predictions, providing useful confidence intervals that improve clinical decision-making by 20%. The results obtained suggest significant improvements in predictive accuracy, transparency, and privacy preservation. This research not only improves health predictions through multimodal analysis but also tackles significant limitations in privacy, interpretability, and uncertainty quantification, thus promoting informed clinical decisions and customized patient care.**

***Index Terms*—Health Prediction, Multimodal Analysis, Privacy-Preserving Models, SHAP Explainability, Transformer Networks.**

Corresponding Author: Gautam Srivastava

M. Sirapangi and G. Sundaram are with School of Computer Science and Engineering, VIT-AP University, Amaravati, 522237, Andhra Pradesh, India (email: dayan.21phd7096@vitap.ac.in; gopikrishnan.s@vitap.ac.in).

N. Herencsar is with the Department of Telecommunications, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technicka 3082/12, 61600 Brno, Czech Republic (email: herencsn@vut.cz, herencsn@ieee.org).

M. Li is with School of Computer Science and Information Engineering, Hefei University of Technology, 230601 Hefei, Anhui, China (email: mengli@hfut.edu.cn).

G. Srivastava is with the Department of Math and Computer Science, Brandon University, Brandon, R7A 6A9, Manitoba, Canada, and the Research Centre for Interneural Computing, China Medical University, Taichung, 40402, Taiwan as well as the Centre for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, India (email: srivastavag@brandonu.ca).

Manuscript received MM DD, YYYY; revised MM DD, YYYY.

S1. Behaviour Analysis with Machine Learning Models

Another very promising area of machine learning is within behavioural analysis. Machine learning in studies [1], [2] shows the ability to predict and automate human behaviours in fields as varied as the analysis of mosquito behaviour to the behaviour of pedestrians crossing. Such works also reported an increase in prediction accuracy in a range of 17 to 23%, suggesting many safer and more efficient systems in areas related to public health or urban planning. However, these models tend to be predicated on sizable pre-labelled data and are inherently bound by those very data, or restricted to particular scenarios. For example, though the work in [2] focuses on pedestrian behaviour at unsignalized intersections, it is not clear whether the model will perform equally well in different urban settings or more diverse populations.

Behaviour monitoring systems show promising results. Real-time monitoring using multi-sensor data increased accuracy by 18%, although continuous data collection poses practical challenges [3]. The precision of behaviour segmentation improved by 14% with matrix factorization, but the high complexity of the data limits the scalability [4]. Simulated driving studies improved merging accuracy by 13%, though real-world applications are still needed [5]. In home robotics, ML improved behaviour recognition by 20%, but extensive labelled datasets are required for wider use [6]. Meanwhile, studies on the impact of COVID-19 revealed a 30% increase in sedentary lifestyles, highlighting the need for predictive and intervention strategies [7].

Explainable Artificial Intelligence has become a crucial area of focus in healthcare machine learning applications driven by the need for transparency and trust in clinical decision-making. Popular model-agnostic methods like LIME [8], and SHAP

[9] allow for instance-level explanation of model outputs by approximating feature contributions. Recently, attention-based models and transformer architectures have also been explored for their inherent interpretability in sequential healthcare predictions [10]. Integrating SHAP within multimodal learning pipelines, as performed in this work, ensures patient-specific feature-level transparency, thus promoting clinician adoption of AI-assisted healthcare tools.

S2. Methodological Rationale and Selection Justification

The proposed model is not merely a stack of well-known methods, but a carefully designed synergy where each component addresses critical gaps identified in predictive healthcare. MTN specializes in temporal and cross-modal fusion, handling the inherent heterogeneity in health data streams. Deep Multitask Neural Networks (DMNN) capitalize on the correlation across health conditions, allowing risk factors to propagate across tasks for better generalization. SHAP-based explainability bridges the gap between complex deep models and clinical interpretability, while Federated Learning ensures that predictive power is obtained without compromising patient privacy. Finally, BNNs add an essential layer of uncertainty quantification, offering clinicians actionable confidence intervals. The combination thus forms a holistic, privacy-preserving, interpretable, and clinically reliable framework, specifically tailored to the nuanced challenges of predictive analytics in healthcare.

The methodology in this work was not selected arbitrarily but instead based on a systematic evaluation of the core challenges in multimodal healthcare analytics. Each component addresses a specific limitation observed in prior studies:

- **MTN** were chosen to handle the temporal and cross-modal dependencies inherent in heterogeneous health data. Unlike early fusion methods or simple recurrent models, MTNs can dynamically attend to important timestamps and modalities, enhancing predictive precision.
- **DMNN** were integrated to model multiple related health outcomes simultaneously. This choice allows the model to exploit the biological and behavioural correlations among diseases, a critical property ignored by many single-task baselines.
- **Explainability with SHAP** was selected after evaluating alternatives such as LIME. SHAP provides consistent feature attributions based on cooperative game theory, making it more theoretically grounded than LIME, which relies on local linear approximations that are less stable across complex decision boundaries [9].
- **Federated Learning** was adopted to ensure privacy-preserving training across decentralized patient datasets. We acknowledge that Federated Learning entails computational overhead and communication challenges. These trade-offs were deemed acceptable considering the regulatory requirements (e.g., HIPAA, GDPR) for real-world healthcare deployment.

Thus, the integration of MTN, DMNN, SHAP, and Federated Learning is a targeted design to achieve predictive accuracy, transparency, and privacy—essential pillars for trustworthy healthcare AI systems.

S3. Improving the Healthcare Prediction using SHAP

SHAP for multimodal data allows for the implementation and integration of an extended interpretability framework that is necessary due to complex health predictions originating from a myriad of data sources, such as heart rate, step count, and air quality. SHAP is based on cooperative game theory that attributes the contribution of every feature toward the model's prediction. This comes with the guarantee of transparency in its localized explanations for each patient. It plays a great role in health care, where the explainability of a model justifies predictions on interpretive and meaningful feature contributions that gain trust in machine learning models. SHAP works by calculating Shapley values-the marginal contribution of each feature toward the outcome in a prediction-while accounting for all combinations of feature inputs to the process. For an explanation variable $x = [x_1, x_2, \ldots, x_n]$ and a model *f*, the Shapley value for a feature x_i is the weighted expectation of its contribution to all subsets of the feature set $S \subseteq x$, excluding the x_i sets. Mathematically speaking, the Shapley value φ_i for feature x_i given using Eq. (1),

$$\phi_i = \sum_{S \subseteq x \setminus \{x_i\}} \frac{S!(n-S-1)!}{n!} \left[f(S \cup \{x_i\}) - f(S) \right]. \quad (1)$$

Here, *n* denotes the total number of features, $|S|$ is the size of the subset *S*, and $f_s\ U_{xi}$, where f_s gives the marginal contribution of x_i when added to the subset *S*. To that end, combinatorial weights $|S!|$ and $|(n-S-1)!|$ ensure that the attribution is fair towards all possible combinations of the input features. Furthermore, the method ensures that the SHAP value is inclined to capture interaction effects across modalities. In the context of health prediction, in order to apply SHAP, first, the model will compute the prediction y' for a patient using a combination of features across different modalities. Suppose that $f(x)$ gives some sort of probability or a risk score of certain health conditions, such as the likelihood of a cardiovascular event. SHAP sums the prediction into the sum of contributions from each single feature using Eq. (2),

$$f(x) = \varphi_0 + \sum_{i=1}^{n} \varphi_i, \quad (2)$$

where φ_0 is the baseline prediction, which means that φ_0 is the predicted value if no features are provided, and φ_i is the Shapley value for the set of characteristics x_i. The baseline φ_0 is calculated as an average forecast for all patients within the dataset and provides a reference from which to compare each of the characteristic contributions individually with respect to the predictions sets $f(x)$. Then, interpret the importance of the feature with that magnitude and the sign of the Shapley values φ_i. For example, a positive Shapley Value about the heart rate would mean that having a high heart rate greatly increases

TABLE S1: Summary of Existing Methods with Machine Learning.

Ref.	Method Used	Findings	Results	Limitations
[11]	Personalized Machine Learning	Predicted well-being and empathy in healthcare professionals.	Improved empathy prediction by 18% using wearable sensor data samples.	Lacks scalability to larger datasets.
[12]	Machine Learning in Personalized Health Management	Reduced barriers for personalized health risk assessments.	Improved accessibility to health risk predictions by 25%.	Lack of long-term evaluation on patient outcomes.
[13]	Machine Learning for Cardiovascular Health in Diabetic Patients	Applied ML models for cardiovascular management in diabetic patients.	Achieved a 10% reduction in cardiovascular incidents in diabetic patients.	Model is specific to diabetic populations and cannot be generalized.
[14]	Machine Learning to Assess COVID-19 Dynamics Due to War Effects	Modeled the impact of the Russian invasion on Italy's COVID-19 trajectory using machine learning.	Identified a 15% increase in COVID-19 cases due to the war's effects.	The model lacks real-time adaptability due to the static dataset.
[15]	IoT-Based Framework with Machine Learning	Personalized health assessment using an IoT-based ML framework.	Improved personal health recommendations by 20%.	High reliance on continuous data streams may affect reliability.
[16]	Machine Learning Algorithms for Diabetes Diagnosis	Assisted in diabetes diagnosis using ML algorithms.	Increased diagnosis accuracy by 15%.	Only applies to diabetes; lacks cross-disease validation.
[17]	Machine Learning for Type 2 Diabetes Prediction	Modeled diabetes risk based on lifestyle behaviors using machine learning.	Improved diabetes risk prediction by 12%.	Requires further personalization for individual patients.
[18]	Strategic ML for Cardiovascular Disease Prediction	Used strategic machine learning optimization to identify high-risk cardiovascular patients.	Improved high-risk patient identification by 17%.	Model complexity leads to high computational overhead.
[19]	Supervised ML Models for Mental Health Diagnosis	Evaluated AI models trained on mental health diagnostic data samples.	Improved diagnostic accuracy by 19%.	Limited explainability reduces clinical trust.
[2]	ML for Automation and Prediction of Human Behaviors	Automated human behavior prediction using machine learning.	Improved prediction accuracy by 18%.	Limited scalability for diverse human behavior scenarios.
[3]	Real-Time Human Behavior Monitoring with Multi-Sensor Modalities	Used multi-sensor modalities for real-time monitoring of human behavior.	Enhanced real-time monitoring accuracy by 18%.	Requires continuous data collection from multiple sensors.

the risks of a cardiovascular event for the patient, whereas a negative Shapley Value about the air quality might seem to suggest that it reduces the risks. These individual explanations will help the clinicians understand why the model came up with a certain prediction and further check if the model's reasoning based on clinical knowledge sets was appropriate.

A critical saliency feature of SHAP pertains to decomposing the impact of multimodal features across temporal instances in sequential data; examples include time series data from wearable devices and deployments. Let X_t be the vector of features in the temporal instance t of a modality, and consider that the model integrates temporal information using a sequence model, such as a Transformer, in that process. Then, the contribution of the time series data at each timestamp can be captured by extending the Shapley values over time, computing a Shapley value $\varphi_{(i,t)}$ for each feature x_i in the temporal instance t sets. The total contribution of the feature x_i can be represented using Eq. (3).

$$\varphi_i = \sum_{t=1}^{T} \varphi(i,t), \tag{3}$$

where T is the total number of timestamp instances. This temporal decomposition is very useful for clinicians because they can refer to specific instances in the temporal instance that were most important for the model's prediction, a spike of heart rate at a specific period of elevated physical activity or stress.

By integrating SHAP with other predictive techniques, the model will generate health predictions accurately and in a way that clinicians can explain with appropriate reasoning. Therefore, SHAP forms an integral part of the development of good recipes in trustworthy AI systems, especially healthcare applications, to enable further improvements in patient outcomes through the facilitation of informed data-driven decisions. Then, the integration of the Federated Learning framework for privacy-preserving multimodal predictive models has been performed to address the urgent need to ensure sensitive patient data privacy while still conducting effective predictive model training from decentralized datasets. Federated learning basically provides an enabling framework for developing a global predictive model by aggregating the knowledge learned across different devices or even institutions, such as hospitals, personal devices, and other IoTs, without physically centralizing raw data samples.

Preserving privacy by maintaining data within its native setting is essential for numerous applications, particularly when dealing with sensitive information in healthcare. Federated learning thus ensures data privacy, yet takes full advantage of collective intelligence gained from diverse

TABLE S2: Summary of Existing Methods using SHAP based Learning Models.

Ref.	Method Used	Findings	Results	Limitations
[20]	Ensemble Machine Learning	Enhanced stroke prediction by integrating various models.	Achieved a 92% accuracy in stroke prediction.	Limited to stroke prediction; lacks generalization to other diseases.
[21]	Integrative Cancer Risk Prediction	Created a predictive model for pancreatic cancer using UK Biobank data samples.	Achieved a 78% sensitivity in pancreatic cancer prediction.	Limited by the underrepresentation of minorities in the dataset.
[4]	Matrix Factorization for Human Behavior Segmentation	Applied matrix factorization for human behavior segmentation.	Improved segmentation accuracy by 14%.	High dimensionality limits model scalability.
[6]	Deep Neural Network for Human Behaviour Recognition in Home Robotics	Used ensemble three-stream deep neural networks for behaviour recognition in home service robots.	Achieved a 20% increase in behaviour recognition accuracy.	Requires extensive labelled data for diverse behaviours.
[22]	Ensemble ML Framework for Early Depression Detection	Applied an ensemble machine learning framework for early detection of depression.	Improved early detection accuracy by 14%.	Model performance degrades with incomplete data streams.
[23]	Explainability in Mental Health Prediction	Used human-machine interaction to enhance explainability in mental health disorder predictions.	Improved model explainability by 20% with interactive features.	Lack of generalization to non-mental health disorders.
[24]	Hybrid WT-CNN Model for Heart Disease Prediction	Proposed a wavelet transform-CNN hybrid model for predicting heart disease.	Achieved 89% accuracy in heart disease detection.	High computational costs for real-time applications.
[25]	Big Data Intelligence for Diabetes Monitoring	Developed a big data framework for continuous diabetes monitoring.	Improved diabetes monitoring by 12%.	Lack of explainability in the model's decision-making.
[26]	Reinforcement Learning for Precision Medicine in Hypertension	Applied reinforcement learning to personalize hypertension treatments for diabetes patients.	Improved treatment efficacy by 14%.	Requires more clinical trial data for validation.
[27]	Nature-Inspired Computing for Disease Prediction	Applied nature-inspired algorithms to predict human diseases using ML.	Improved disease prediction accuracy by 12%.	Lacks personalization for specific patient cohorts.
[28]	Human-in-the-Loop Systems for Behavior Learning	Developed adaptive models for behavior learning in human-in-the-loop systems.	Improved task performance prediction accuracy by 15%.	Requires extensive calibration for different tasks.

and geographically dispersed multimodal data sources. In federated learning, the overall workflow begins when every client-participated training trains their local models on respective datasets: hospitals or personal devices. Consider the local set of training data at client k, which can be defined by D_k and whose elements are multimodal features X_k illustrated by the following types of data: wearable sensor, behavioral data, and medical history. Each of the local models is separately trained with the local data to obtain W_k-parameterized models. This process is mathematically formulated as the minimization of a local loss function L_k, W_k, which in turn is specific to the data distribution in client k using Eq. (4).

$$W_k^* = \arg\min_{W_k} L_k(W_k; D_k). \tag{4}$$

The central server views only the locally optimized model parameters W_k* from each of the client sets and not the local data, in the process of updating the global model, W_{global}, by aggregating these model parameters from all clients. A common aggregation scheme is weighted averaging, where the contribution of each client's model update is directly proportional to the number of samples $|D_k|$ in its local dataset samples. The global model update is given as Eq. (5).

$$W_{\text{global}} = \sum_{k=1}^{K} \frac{|D_k|}{\sum_{j=1}^{K} |D_j|} W_k^*, \tag{5}$$

where K is the total number of participating clients and $|D_k|$ is the size of the data points sets in client k. The process repeats, whereby the global model is sent back to each client for further training at the local level to arrive at a continuous refinement of the global predictive model over successive rounds of communication. Other major benefits of using federated learning in such a multimodal health prediction setting are cross-population generalization without breach of privacy regulations. For example, data coming from different hospitals could differ in terms of patient demographics or local environmental conditions and would introduce significant bias if the corresponding banks of data used during computations were centralized. This contrasts with federated learning, where each institution practices local control of its data while participating in the unison training of the global model, making its outcome more resilient regarding differences in the distribution of data between processes.

The federated design in this study separates the shared global structure from optional client-level adaptation. During the main training stage, all clients optimise the same set of

TABLE S3: Comparison of Existing Methods and Identified Gaps in Health Analytics.

Study	**Data Type and Modality**	**Prediction Model**	**Multitask Learning**	**Privacy Aware Training**	**Explanation and Uncertainty**	**Main Gap With Respect to Proposed System**
Classical clinical risk models using structured records [13], [20]	Single or few structured clinical variables and scores	Tree ensemble or shallow neural model	Not used	Not used	No explanation and no uncertainty	Limited to narrow feature sets and cannot exploit rich multimodal streams
Multimodal health studies using sensor and behaviour data, [3], [15]	Wearable and sensor data and behaviour logs	Deep learning model with feature concatenation	Rarely used	Not used	Limited explanation and no uncertainty	Do not employ transformer-based cross-modal attention or multitask structure
Federated learning studies in healthcare [29], [30]	Structured records or imaging data across sites	Federated neural or logistic model	Usually single task	Used	No local explanation and no uncertainty	Provide privacy but do not support multimodal fusion or patient-level explanation of predictions
Explainable clinical models based on SHAP or similar methods [9], [23]	Structured clinical records or selected sensor variables	Gradient boosted trees or feedforward networks	Not used	Not used	Global or local explanation only	Do not combine explanation with federated learning or multimodal transformer-based fusion
Uncertainty-aware medical prediction models using Bayesian approaches [2], [31]	Structured or imaging data	Bayesian neural model	Mostly single task	Not used	Provide uncertainty but no detailed feature attribution	Do not link uncertainty with multitask outputs and do not operate in a federated multimodal setting
Proposed HealthEngine system	Physiological streams, behaviour data and environmental signals combined across sites	Multimodal transformer with deep multitask prediction and Bayesian heads	Used for several related health outcomes	Used through federated optimisation of shared and task specific parts	Provides subject specific SHAP based explanation and prediction intervals for every task	Addresses multimodal fusion, knowledge transfer across tasks, privacy, explanation and uncertainty together in one integrated system

shared parameters W^{shared}, which includes the layers of the multimodal transformer and the common part of the multi-task model. The task-specific output branches are also trained jointly and aggregated into global parameters W^{task} through the server-side averaging procedure described earlier. The quantitative results reported in the tables are obtained from this global model, which is then deployed at every client site without further modification in order to maintain a consistent basis for comparison with baseline systems.

The framework also supports a light form of personalization once global training has converged. After the final global parameters W^{shared} and W^{task} are broadcast, each client is allowed to refine a small set of local bias terms in the task outputs using its own validation data while keeping W^{shared} fixed. Let the local bias vector for client k and task j be written as b_kj. During a short local adaptation phase, client k performs a few gradient steps on b_kj for each task while the shared representation and all other weights remain unchanged. This produces a personalised adjustment of decision thresholds that reflects the case mix and data distribution of that client while preserving the common structure learned across all sites. In our experiments, this local tuning provided modest gains for clients with strong domain shift. However, to ensure clarity, we report the performance of the global model in the main tables and mention the personalised variant as an additional capability of the framework.

The convergence of global models is one of the key components in protected federated learning, particularly when the distribution of data among the various clients is not identical. This challenge is addressed through a combination of careful initialization of the global model and adaptive learning rates for local updates. In particular, the local models are initialized by using the global model parameters before each training round, represented as $W_{\text{global}}(t)$ for the t^{th} communication rounds. Local optimization involves updating the model parameters using gradient-based methods such as stochastic gradient descent using Eq. (6).

$$W_k(t+1) = W_k(t) - \eta \nabla L_k(W_k(t)). \tag{6}$$

Here η is the learning rate, and $\nabla L_k(W_k(t))$ indicates a gradient of a local loss function related to model parameters in the process in the round *T*. The global model at each time step is then updated by aggregating the local updates during each round, via the following process: Federated learning is most suitable for the healthcare domain because it handles

sensitive data samples on health issues while maintaining important privacy. Traditional machine learning approaches, centrally modeled, require aggregating data from all sources into a single repository, raising the risk of data breaches, and this seems very doubtful in terms of GDPR and HIPAA privacy regulations. As the data for each patient resides within the respective client device and only the model parameters are exchanged between the clients and the central server, federated learning does not face these challenges.

S4. Dataset Description and Training

In a federated learning paradigm, local models are separately trained in each contributing institution or other personal device within a privacy-preserving protocol where raw data are never shared with the central server. The sample size for the local datasets ranged from 10,000 to 50,000 depending on the size of the institution and the population of patients. Local models were initialized with random models and optimized by stochastic gradient descent using a learning rate of 0.001 and a batch size of 128. Training was performed for 20 rounds of communication. The model updates were then averaged using weights proportional to the size of the contributed data set for each institution. The global model was evaluated every 5 rounds using a validation set, which consists of 15% of the total data set. Stratification was performed to ensure equal proportions of all health conditions in the test set. Predictions were made using an Aggregated Global Model in a Bayesian Neural Network with quantified uncertainty. Point estimates computed from uncertainty estimates by variational inference with a Gaussian prior are reported.

The datasets used in this study contain gaps in several streams and exhibit an imbalance between clinical labels. For missing values in time series signals, we applied forward fill for short gaps and interpolation with cubic splines for longer gaps. For static attributes, we used mean imputation for numerical fields and mode imputation for categorical fields. To address label imbalance, we applied class weight adjustment in the loss function where the weight for class c is set to $w_c = \frac{1}{f_c}$ where f_c is the observed frequency of the class in the training split. We also used random minority oversampling to ensure that each class appears in every mini-batch. These steps reduce training bias and provide stable gradients in all health prediction tasks.

In all estimates, the posterior distribution over the weights was approximated by Monte Carlo sampling with 50 samples. For each prediction, SHAP values are calculated, resulting in highly detailed feature attributions. These attributions have been manually confirmed to ensure that high levels of interpretability and clinical relevance were achieved. In these experiments, a data set was created by combining the information in databases available to public health and private clinical data sets. For example, detailed patient health records were obtained through the MIMIC-IV [32] clinical database, while other sources of wearable sensor data included the PhysioNet signal data set, such as ECG, PPG, and accelerometer data from smartwatches. Emulations of phone usage patterns and social interaction were found in the StudentLife data set. The environmental data was obtained from the UCI Machine Learning Repository for its comprehensive environmental sensor data spread in various geographies. Data were split 70%: 15%: 15% for training, validation, and testing. Stratified sampling was used to ensure that all subsets were well represented for chronic diseases and mental health conditions.

The study uses both real and synthetic sources. Real data come from MIMIC IV, StudentLife and the UCI environmental set. Synthetic behaviour and environmental streams were created only for variables that were not available in the real sets. These generated streams follow the same temporal grid as the real data and use simple rules to model activity level and ambient readings. They serve to complete missing channels for multi-stream fusion and do not replace any real clinical values.

TABLE S4: Hyperparameter Settings for Model Training

Hyperparameter	**Value**
Learning Rate	0.001 minutes
Batch Size	128 (Local: 64 for FL)
Optimizer	Adam (β_1 = 0.9, β_2 = 0.999)
Dropout Rate	0.3
Number of Training Rounds (Local)	20
Global Aggregation Frequency	Every 5 rounds
Task-Specific Loss Function Weights	1.0

TABLE S5: Hardware Configuration Used for Training

Component	**Specification**
GPU	Nvidia Tesla V100
CPU	Intel Xeon with 128 GB of RAM
Storage	Sufficient for large datasets and model checkpoints

TABLE S6: Training Time per Round for Federated Learning

Step	**Time per Round**
Local Model Updates	10 minutes
Global Aggregation	5 minutes
Total Federated Learning Round Time	15 minutes

A. Hyperparameters and System Configuration

In this work, we used the following hyperparameters for training the HealthEngine model, as shown in Table S4.

1) Hardware Configuration: The experiments were conducted on the following hardware setup, as shown in Table S5.

2) Training Time per Round: The training time per round of federated learning was measured and summarized in Table S6.

The federated learning setup ensured data privacy while enabling the model to aggregate knowledge from decentralized datasets across different client devices.

S5. Ablation Study and Component-wise Analysis

To rigorously assess the contribution of each major component—MTN, DMNN, SHAP, Federated Learning, and BNN—an ablation study was conducted. Table S7 presents the performance metrics as each module is progressively

TABLE S7: Ablation Study: Contribution of Each Component

Configuration	Accuracy	F1 Score	RMSE	UQ (±%)
Baseline (Concatenation +Single-task)	78.5	0.820	0.138	–
Baseline + MTN	82.7	0.851	0.121	–
Baseline + MTN + DMNN	85.3	0.872	0.108	–
Baseline + MTN + DMNN + SHAP	85.7	0.877	0.105	–
Baseline + MTN + DMNN + SHAP + Federated Learning	87.0	0.887	0.097	–
Full Model (All + BNN)	87.2	0.891	0.094	±6.3

integrated into the baseline model. The baseline model uses a simple concatenation of features and a single-task neural network.

Starting from the baseline, we observe that the simple concatenation approach struggles to capture complex inter-modality and temporal relationships, reflected in modest accuracy (78.5%) and higher RMSE (0.138). When MTN is introduced, a significant jump in the accuracy of approximately 4.2%. This gain is attributable to the MTN's ability to model cross-modal temporal dependencies through self-attention and cross-attention mechanisms, which simple concatenation fails to capture. Adding DMNN further increases performance by exploiting shared risk factors in all health conditions. Multitask learning encourages the model to take advantage of commonalities in the data, leading to an additional gain of 2.6% in accuracy and a notable reduction in RMSE.

Integration of SHAP-based explainability does not directly impact predictive performance metrics but plays a critical role in clinical interpretability. Its inclusion ensures that each prediction is accompanied by transparent patient-specific explanations, significantly increasing the trustworthiness of the system in clinical deployment. Introducing Federated Learning preserves patient data privacy while simultaneously aggregating knowledge across decentralized data sources. This leads to a further improvement in performance (+1.3% accuracy), as the model benefits from a wider variety of patient data without violating privacy norms. Shows that collaborative learning across institutions enhances generalization capabilities.

Finally, BNNs are incorporated to quantify uncertainty. Although the absolute gain in accuracy is marginal (+0.2%), the added capability to output uncertainty limits (±6.3%) is clinically invaluable. In high-risk domains like healthcare, knowing the confidence level associated with predictions is crucial to decision-making, particularly in borderline cases. The ablation study clearly demonstrates that each component systematically contributes to the robustness, interpretability, privacy compliance, and clinical utility of the model. Rather than being a superficial aggregation, the integration strategy is carefully designed to address the multifaceted challenges inherent to predictive healthcare modelling.

S6. Practical Deployment and Ethical Considerations

Although the proposed framework demonstrates strong potential in predictive healthcare analytics, several practical deployment challenges must be recognized. Training Multimodal Transformer Networks within a federated setting introduces substantial communication overhead and computational complexity, particularly due to the size and dynamic nature of healthcare data. Deployment strategies involving hybrid edge-cloud architectures are recommended to mitigate latency and bandwidth constraints. Furthermore, although Federated Learning inherently preserves data locality, it does not eliminate all privacy vulnerabilities; risks such as model inversion and gradient leakage attacks remain plausible. Incorporating secure aggregation protocols and differential privacy mechanisms will be essential to enhance the robustness of privacy guarantees. Another practical challenge concerns interpretability: while SHAP values provide localized explanations, interpreting these explanations consistently across highly dynamic, temporal data streams remains an open problem, requiring future advances in sequential explainability.

From an ethical standpoint, the development and deployment of AI in healthcare systems must be governed by the principles of fairness, transparency, and security. Predictive models can inadvertently perpetuate biases present in training datasets, particularly across demographic variables such as age, gender, and ethnicity. Continuous auditing and bias mitigation techniques are critical to ensure equitable clinical outcomes. In federated environments, even though raw data remain decentralized, model updates can leak sensitive information if not carefully protected; thus, techniques like secure multiparty computation and noise injection must be integrated. Moreover, while SHAP-based interpretability mechanisms improve transparency, they must be appropriately contextualized to avoid misinterpretations that could misguide clinical decisions. In general, ethical risk assessments and robust governance frameworks are indispensable for the trustworthy adoption of AI-driven healthcare solutions.

The present study does not report numerical measures of clinical trust because no structured user evaluation was carried out with clinicians. All statements about improved trust and transparency are therefore qualitative and arise from the ability of the SHAP-based analysis to show subject-level feature contributions for every prediction. A dedicated study collecting ratings and feedback from medical personnel is left as future work so that perceived trust can be rigorously quantified.

A. Fairness and Bias in Predictive Healthcare

In addition to privacy concerns, fairness and bias are critical ethical issues when deploying AI in healthcare. It is crucial that predictive models do not disproportionately impact certain demographic groups or underrepresented populations, leading to biased or inequitable healthcare decisions.

1) Fairness and Bias in Healthcare AI: AI systems, including predictive healthcare models, are susceptible to the inherent biases in the training data. These biases may

arise from the overrepresentation or underrepresentation of certain demographic groups, such as race, gender, age, or socioeconomic status. To mitigate the risks of bias, we ensure that the datasets used for training include diverse representations across all relevant demographic variables. The datasets include a balanced mix of samples from different ethnicities, age groups, and genders, ensuring that the model does not favour one group over another.

2) Methods for Mitigating Bias: Several methods were applied during training to reduce the potential bias in model predictions:

- **Bias-aware loss functions:** We employed loss functions that penalize predictions that disproportionately affect certain demographic groups. This helps ensure that the model learns to make fair predictions across all groups.
- **Data balancing techniques:** Oversampling and undersampling techniques were used to balance the dataset, ensuring that underrepresented groups are adequately represented in the model's training process.
- **Bias detection and auditing:** A set of bias detection techniques was employed post-training to assess the fairness of the model's predictions. This included testing the model on specific subgroups to check if certain groups are unfairly disadvantaged in terms of prediction accuracy.

3) Impact on Underrepresented Groups: Despite precautions, it is important to acknowledge that complete fairness is challenging, especially when dealing with medical data. For example, certain diseases may be more prevalent in certain populations, which may inadvertently affect the performance of the model in different groups. We strive to minimize these effects through careful monitoring and continuous feedback loops. Furthermore, the inclusion of diverse data sources, such as wearable sensors and environmental data, allows the model to capture broader health patterns, reducing the risk of bias based solely on clinical records.

4) Ongoing Fairness Audits: To ensure that the model remains fair over time, we plan to implement regular fairness audits as part of the model's lifecycle. These audits will include:

- Regular evaluations of the model's performance on new, diverse datasets.
- Continuous monitoring of outcomes disparities based on demographics of the patient.
- Adjustments to the structure and training process of the model to address any identified biases.

By integrating fairness and bias mitigation strategies into the training and deployment of predictive healthcare models, we aim to ensure that the HealthEngine framework delivers equitable, reliable, and transparent predictions for all patient groups, ultimately promoting fairness in clinical decision-making.

REFERENCES

[1] Y. M. Qureshi, V. Voloshin, C. E. Towers, J. A. Covington, and D. P. Towers, "Double vision: 2D and 3D mosquito trajectories can be as valuable for behaviour analysis via machine learning," *Parasites & Vectors*, vol. 17, no. 1, p. 282, 2024.

[2] H. Jupalle, S. Kouser, A. B. Bhatia, N. Alam, R. R. Nadikattu, and P. Whig, "Automation of human behaviors and its prediction using machine learning," *Microsystem Technologies*, vol. 28, no. 8, pp. 1879–1887, 2022.

[3] S. Dávila-Montero, J. A. Dana-Lê, G. Bente, A. T. Hall, and A. J. Mason, "Review and challenges of technologies for real-time human behavior monitoring," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 1, pp. 2–28, 2021.

[4] H. Gao, C. Lv, T. Zhang, H. Zhao, L. Jiang, J. Zhou, Y. Liu, Y. Huang, and C. Han, "A structure constraint matrix factorization framework for human behavior segmentation," *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 12978–12988, 2021.

[5] O. Siebinga, A. Zgonnikov, and D. A. Abbink, "Human Merging Behaviour in a Coupled Driving Simulator: How Do We Resolve Conflicts?," *IEEE Open Journal of Intelligent Transportation Systems*, 2024.

[6] Y.-H. Byeon, D. Kim, J. Lee, and K.-C. Kwak, "Ensemble three-stream RGB-S deep neural network for human behavior recognition under intelligent home service robot environments," *IEEE Access*, vol. 9, pp. 73240–73250, 2021.

[7] D. Rawat, V. Dixit, S. Gulati, S. Gulati, and A. Gulati, "Impact of COVID-19 outbreak on lifestyle behaviour: A review of studies published in India," *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, vol. 15, no. 1, pp. 331–336, 2021.

[8] M. T. Ribeiro, S. Singh, and C. Guestrin, ""Why Should I Trust You?": Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA*, pp. 1135–1144, 2016.

[9] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), Long Beach, California, USA*, pp. 4768–4777, 2017.

[10] E. Choi, M. T. Bahadori, A. Schuetz, W. Stewart, and J. Sun, "RETAIN: An Interpretable Predictive Model for Healthcare Using Reverse Time Attention Mechanism," in *Advances in Neural Information Processing Systems*, vol. 29, pp. 3504–3512, 2016.

[11] J. Nan, M. S. Herbert, S. Purpura, A. N. Henneken, D. Ramanathan, and J. Mishra, "Personalized Machine Learning-Based Prediction of Wellbeing and Empathy in Healthcare Professionals," *Sensors*, vol. 24, no. 8, p. 2640, 2024.

[12] H. Park, S. Y. Jung, M. K. Han, Y. Jang, Y. R. Moon, T. Kim, S.-Y. Shin, and H. Hwang, "Lowering Barriers to Health Risk Assessments in Promoting Personalized Health Management," *Journal of Personalized Medicine*, vol. 14, no. 3, p. 316, 2024.

[13] R. Jose, F. Syed, A. Thomas, and M. Toma, "Cardiovascular health management in diabetic patients with machine-learning-driven predictions and interventions," *Applied Sciences*, vol. 14, no. 5, p. 2132, 2024.

[14] D. Chumachenko, T. Dudkina, T. Chumachenko, and P. P. Morita, "Epidemiological Implications of War: Machine Learning Estimations of the Russian Invasion's Effect on Italy's COVID-19 Dynamics," *Computation*, vol. 11, no. 11, p. 221, 2023.

[15] S. K. Jagatheesaperumal, S. Rajkumar, J. V. Suresh, A. H. Gumaei, N. Alhakbani, M. Z. Uddin, and M. M. Hassan, "An iot-based framework for personalized health assessment and recommendations using machine learning," *Mathematics*, vol. 11, no. 12, p. 2758, 2023.

[16] L. P. Nguyen, D. D. Tung, D. T. Nguyen, H. N. Le, T. Q. Tran, T. V. Binh, and D. T. N. Pham, "The utilization of machine learning algorithms for assisting physicians in the diagnosis of diabetes," *Diagnostics*, vol. 13, no. 12, p. 2087, 2023.

[17] S. R. Velu, V. Ravi, and K. Tabianan, "Machine learning implementation to predict type-2 diabetes mellitus based on lifestyle behaviour pattern using HBA1C status," *Health and Technology*, vol. 13, no. 3, pp. 437–447, 2023.

[18] K.-V. Tompra, G. Papageorgiou, and C. Tjortjis, "Strategic Machine Learning Optimization for Cardiovascular Disease Prediction and High-Risk Patient Identification," *Algorithms*, vol. 17, no. 5, p. 178, 2024.

[19] A. van Oosterzee, "AI and mental health: evaluating supervised machine learning models trained on diagnostic classifications," *AI & SOCIETY*, vol. 40, pp. 5077–5086, Aug. 2025.

[20] R. Wijaya, F. Saeed, P. Samimi, A. M. Albarrak, and S. N. Qasem, "An Ensemble Machine Learning and Data Mining Approach to Enhance Stroke Prediction," *Bioengineering*, vol. 11, no. 7, p. 672, 2024.

[21] T.-M. Ke, A. Lophatananon, and K. R. Muir, "An integrative pancreatic cancer risk prediction model in the UK biobank," *Biomedicines*, vol. 11, no. 12, p. 3206, 2023.

[22] I. Khan and R. Gupta, "Early depression detection using ensemble machine learning framework," *International Journal of Information Technology*, vol. 16, pp. 3791–3798, Aug. 2024.
[23] I. Kaur, Kamini, J. Kaur, Gagandeep, S. P. Singh, and U. Gupta, "Enhancing explainability in predicting mental health disorders using human–machine interaction," *Multimedia Tools and Applications*, vol. 84, pp. 34945–34966, Sep. 2025.
[24] F. Mohammad and S. Al-Ahmadi, "WT-CNN: a hybrid machine learning model for heart disease prediction," *Mathematics*, vol. 11, no. 22, p. 4681, 2023.
[25] S. AlZu'bi, M. Elbes, A. Mughaid, N. Bdair, L. Abualigah, A. Forestiero, and R. A. Zitar, "Diabetes monitoring system in smart health cities based on big data intelligence," *Future Internet*, vol. 15, no. 2, p. 85, 2023.
[26] S. H. Oh, S. J. Lee, and J. Park, "Precision medicine for hypertension patients with type 2 diabetes via reinforcement learning," *Journal of Personalized Medicine*, vol. 12, no. 1, p. 87, 2022.
[27] MunishKhanna, L. K. Singh, and H. Garg, "A novel approach for human diseases prediction using nature inspired computing & machine learning approach," *Multimedia Tools and Applications*, vol. 83, no. 6, pp. 17773–17809, 2024.
[28] H.-N. Wu, "Online learning human behavior for a class of human-in-the-loop systems via adaptive inverse optimal control," *IEEE Transactions on Human-Machine Systems*, vol. 52, no. 5, pp. 1004–1014, 2022.
[29] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, and W. Shi, "Federated Learning of Predictive Models from Federated Electronic Health Records," *International Journal of Medical Informatics*, vol. 112, pp. 59–67, 2018.
[30] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, and S. Bakas, "Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data," *Scientific Reports*, vol. 10, p. 12598, Jul. 2020.
[31] D. Singh, P. Das, and I. Ghosh, "Prediction of pedestrian crossing behaviour at unsignalized intersections using machine learning algorithms: analysis and comparison," *Journal on Multimodal User Interfaces*, vol. 18, pp. 239–256, Sep. 2024.
[32] A. E. W. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, S. Hao, B. Moody, B. Gow, L.-w. H. Lehman, L. A. Celi, and R. G. Mark, "MIMIC-IV, a freely accessible electronic health record dataset," *Scientific Data*, vol. 10, p. 1, Jan. 2023.