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

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.**

I. Introduction

Diversified data streams combined with contextual data, such as air quality and sleep patterns, develop the potential to significantly improve precision and timeliness in the prediction of the end of the health condition [1]–[3]. However, how to effectively integrate and extract meaningful insights from such heterogeneity in big data raises major concerns in the healthcare domain. For heterogeneous data characterized with multimodal information, while the basic predictive power can be achieved to a satisfactory level using traditional machine learning approaches, dealing with complexity and data volume often leads to less informative predictions and poor clinical application. The main limitation of health prediction models at present lies in the reliance on simple concatenation fusion, which does not capture relationships between different data modalities. Such approaches leave important information on the Table I and might handle the modalities as independent sources of information, while taking into account these modalities in a functionally related aspect, hopefully giving us a better pattern of a patient's health.

It is important to clarify that while all multimodal datasets are inherently heterogeneous, the reverse is not necessarily true; not all heterogeneous datasets involve distinct modality types. In this work, our focus is specifically on multimodal health data, which comprises structured clinical data, time-series physiological signals, behavioural logs, and environmental sensor streams. This distinction refines our problem formulation and highlights the need for specialized fusion and interpretation techniques that respect both modality-specific structures and cross-modal interactions.

In addition, most traditional models do not take into account the sequential nature of health data and, more importantly, time-series wearable data, which leads to a loss of important temporal dependencies. This usually results in a lack of

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.

TABLE I: Summary of Existing Methods with Machine Learning.

Ref.	Method Used	Findings	Results	Limitations
[2]	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.
[4]	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.
[5]	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.
[6]	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.
[7]	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.
[8]	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.
[9]	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.
[10]	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.
[11]	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.
[12]	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.
[13]	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.

high predictive accuracy for complex health conditions that develop over time, such as cardiovascular diseases, diabetes, and mental health disorders [4], [5], [14]. In addition to these limitations in data fusion and temporal modelling, current methods often face significant challenges regarding transparency and interpretability. Clinicians or healthcare providers need to know more than what the model predicts; they need to understand the reason behind the prediction with which to suggest an informed decision regarding patient care sets. Black-box models explain little about their outputs, hence breeding suspicion and less trust in practitioners clinically. This is a problem in high-stakes settings, such as health, where the results of wrong predictions will actually be catastrophic during clinical scenarios. This may also hinder the personalization of care because one will not be in a position to set the models to explainable reasons. This goes further to where it is quite hard to personalize treatments to individual patients without a clear understanding of what drives model inferences.

Another thing is that health data are sensitive in nature and therefore raise huge privacy concerns. Centralized models aggregate data from many sources to train a global predictive model; this requires the need to collect patient information in a central repository. This not only increases the vulnerability related to data breaches, but also creates serious regulatory challenges with respect to data protection and HIPAA and GDPR compliance. With healthcare becoming more personalized and data-driven, the need for privacy and security is gaining the front seat. Given these limitations, this research introduces new opportunities for predictive healthcare, as data from shared origins, including wearable devices, environmental sensors, and behavioural monitoring systems, are linked to health. This research proposes an integrated design that combines several advanced machine learning techniques, each of which is selected for a specific purpose when applied to the current state of predictive healthcare systems.

The core research problem addressed in this work is the design of an integrated, explainable, privacy-preserving, and uncertainty-aware predictive framework that can model temporal, cross-modal, and structural relationships in complex multimodal healthcare datasets. This requires not only powerful modelling techniques, but also the incorporation of interpretability, privacy, and uncertainty quantification mechanisms to meet clinical deployment standards.

The proposed system introduces a fully unified design that brings together multimodal transformer learning, multitask prediction, federated training with respect to privacy, patient-centred explanation, and uncertainty estimation in a continuous data-to-decision pipeline. Previous works treat

these ideas as isolated modules that do not interact with one another. Our system uses the multimodal transformer to create deep shared representations that capture temporal and cross-modal structure across physiological, behavioural, and environmental streams. These representations are then used by the multitask model to learn common risk factors in several health outcomes. The federated setting allows these models to learn from several sites without collecting sensitive data in a central store. The uncertainty-aware prediction head uses a distribution-based view of the shared representation to provide confidence ranges for every clinical output. The SHAP-based explanation stage is applied directly on the federated transformer and multitask outputs and produces clear feature-level contributions for every subject. This unified design offers a new direction in clinical artificial intelligence by addressing fusion, privacy, trust, and uncertainty together instead of treating them as separate concerns.

A. Motivation & Contribution

The growing demand for precise and personalized health prediction systems capable of managing increasingly complex and varied multimodal data samples is the primary motivation for this research. As wearable technologies, behavioural monitoring systems, and environmental sensors become more widespread, there is an exponential increase in the volume and variety of health data available for predictive modelling. While there exists the intention of development, most of the predictive models are not very good at integrating these data streams because of either naive fusion techniques or a model's incapability to capture temporal and cross-modal dependencies. In addition, the lack of interpretability and privacy concerns within these models raises significant challenges for translation into clinical settings. The current paper, therefore, attempts to address these lacunae by designing an integrated model that leverages state-of-the-art machine learning techniques to improve health prediction accuracy with interpretability and privacy.

The contributions of the current work are to be understood from multiple points of view. The first lies in the application of the Multimodal Transformer Networks, introducing an intricate mechanism for the fusion of heterogeneous data sources, capturing temporal and cross-modal relationships in a manner not realizable with other methods. The second lies in applying Deep Multitask Neural Networks with the ability to enable higher predictive accuracy for multiple health conditions due to the sharing of risk factors among them for different scenarios. Third, SHapley Additive exPlanations for explainability result in increased clinical trust through transparent model predictions at a patient-specific level. Fourth, incorporating Federated Learning allows for privacy-preserving predictive modelling, addressing a crucial concern related to the security of healthcare data. Finally, the application of Bayesian Neural Networks provides uncertainty quantification, thus increasing the reliability of the model outputs and providing clinicians with valuable insights about the confidence of the predictions. Together, these contributions advance the field of predictive healthcare and provide a complete and innovative solution to challenges related to the analysis of multimodal data samples.

II. Review of Existing Models used for Healthcare Analysis

The rapidly advancing field of machine learning and artificial intelligence has infiltrated numerous aspects of human endeavours, including healthcare, behavioural analysis, industrial automation, and human-robot collaboration. In this perspective, with the growing embedding of AI in diverse fields, researchers are trying to investigate how machine learning could help improve predictive accuracy, interpretability, and decision-making while addressing essential challenges in privacy concerns, model generalization, and real-time adaptability. The review in this section is written with this broad scope and focuses on the current developments, opportunities, and limitations of the field. Machine learning models have increasingly found applications in health care for the prediction, diagnosis, and management of a variety of health-related conditions. Articles like [1], [5] focus on improving clinical outcomes by predicting stroke or cardiovascular risks in diabetic patients. These works report impressive accuracy, ranging from the work in [1] with a stroke prediction rate of 92% to the work in [5], where the process reduces cardiovascular incidents to up to 10%. The clear potential of machine learning to improve patient care is undeniable, although these approaches also reveal the limitations of applying models to specific diseases. The former are highly specialized, and their generalization to other health conditions remains really limited. Consequently, while models [1], [5] often excel in specific domains, frameworks that expand their applications across diverse health conditions are still needed.

To more clearly emphasize the originality of the proposed architecture, we present a systematic, itemized comparison that delineates the key differences between prior approaches and the current system. The comparison focuses on the type of health data, the main prediction model, the presence of multitask learning, the use of privacy-preserving training, the availability of explanation, the presence of uncertainty estimates, and the key missing gaps. This summary shows that earlier clinical prediction studies usually address only one or two of these aspects at a time. In contrast, our system jointly supports multimodal fusion, multitask prediction, privacy-sensitive learning, explanation at the subject level, and uncertainty estimation within a single clinical decision pipeline.

A. Behaviour Analysis with Machine Learning Models

Another very promising area of machine learning is within behavioural analysis. Machine learning in studies [12], [15] 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 [12] 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 [13]. The precision of behaviour segmentation improved by 14% with matrix factorization, but the high complexity of the data limits the scalability [16]. Simulated driving studies improved merging accuracy by 13%, though real-world applications are still needed [17]. In home robotics, ML improved behaviour recognition by 20%, but extensive labelled datasets are required for wider use [18]. Meanwhile, studies on the impact of COVID-19 revealed a 30% increase in sedentary lifestyles, highlighting the need for predictive and intervention strategies [19].

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 [20], and SHAP [21] 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 [22]. 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.

B. *Existing ML Models for Healthcare Analytics*

Recent advances in ML have significantly transformed various domains of health and behaviour monitoring, offering improved accuracy and efficiency while revealing critical limitations. For example, supervised ML models have demonstrated a 19% improvement in the accuracy of mental health diagnosis [11], while ensemble frameworks have shown promise in early detection of depression, improving early identification by 14% [23]. However, these advances are often hindered by issues such as limited explainability, which reduces clinical trust, and vulnerability to incomplete data streams. For chronic diseases, ML has improved monitoring and prediction. The monitoring of rheumatoid arthritis improved by 22%, and the unobtrusive monitoring of mental health achieved a gain of 15% using quantum ML. However, scalability and real-time applications remain hurdles [24], [25]. Behaviour prediction systems have also seen an 18% improvement, but struggle to adapt to various scenarios [12].

Table I indicates the rapid development of machine learning as a powerful tool in many different applications. In healthcare, several promising applications of machine learning have emerged with the promise of improving diagnostic accuracy, personalizing treatment planning, and enabling continuous health monitoring through multimodal data fusion. For example, the accuracy for the prediction of stroke and diabetes has been enhanced in works [1], [7] through sophisticated real-time data stream machine learning models. Such developments are crucial as healthcare systems shift towards proactive and preventive care models. However, the challenges of explainability, privacy, and generalizability need to be resolved to secure the confidence of both clinicians and patients in these models. Not surprisingly, explainability is one of the key challenges that emerges from this review. In fact, while the models proposed in [11], [26] both attempted to make machine learning predictions interpretable, a lot remains to be done regarding the ability of non-expert users to understand AI-driven predictions. This is a problem not confined to healthcare, and in industrial applications [27], [28], there is an ever-increasing demand for AI systems to explain their decisions, particularly as automation is becoming increasingly prevalent in safety-critical settings. In fact, developing highly accurate models capable of providing transparency will be the key to establishing trust in AI by the general public.

Recent transformer models designed for structured health records, BEHRT [29] and Med-BERT [30], show that sequential attention can model long-range clinical patterns in large-scale electronic records. These works focus on structured event sequences, while the present study extends the transformer idea to multimodal data streams that include physiological signals, behaviour logs, and environmental readings. The present framework also adds multitask prediction, SHAP-based interpretation, federated training, and uncertainty estimation, which are not covered in these earlier models. The inclusion of these references places the proposed approach within the broader space of transformer-driven health analytics.

C. *Reviews on SHAP-based Federated Learning Models*

Federated Learning has emerged as a privacy-preserving paradigm that allows decentralized model training without transferring raw patient data to centralized servers. Pioneering works FedAvg [31] have demonstrated the viability of FL in large-scale, non-IID settings. In healthcare, applications of FL have focused on cross-institutional collaborations [32], [33], enabling models to learn from distributed clinical records while adhering to regulatory requirements like HIPAA and GDPR. Our work adopts and extends the federated setting by incorporating multimodal health data fusion, personalized local training, and privacy-preserving aggregation strategies, aligning with the growing emphasis on secure, distributed healthcare analytics.

Another important aspect is privacy in relation to the avalanche of data from wearable devices, IoT systems, and other sensors. Federated learning, as argued in [7], [39], holds promise for a good approach to keep user privacy intact, yet enables machine learning. In this way, with the sensitive data on the local device level in this decentralized model training approach, global improvements continue to occur in model performance due to shared knowledge. Federated learning can be a critical means of addressing these priorities, which are currently at odds, in some of the most important AI

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

Ref.	Method Used	Findings	Results	Limitations
[1]	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.
[14]	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.
[16]	Matrix Factorization for Human Behavior Segmentation	Applied matrix factorization for human behavior segmentation.	Improved segmentation accuracy by 14%.	High dimensionality limits model scalability.
[18]	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.
[23]	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.
[26]	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.
[34]	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.
[35]	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.
[36]	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.
[37]	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.
[38]	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.

application areas, such as healthcare and finance. However, this approach also has multiple associated technical challenges, before federated learning can fully deliver on its promise, which include how to coordinate across decentralized systems and a rise in computational costs. Generalization to other settings or populations remains a persistent challenge for most machine learning models in any domain. The models outlined in [12], [15] have been successful in enhancing the prediction of behaviour and automation within certain settings. However, these models have a small generalization to other settings or populations. This is dramatically the case in models built to perform rigidly for some task of interest in animals; see, for example, the poultry task management model in [27]. Consequently, future machine learning studies should be directed toward making models more flexible and adaptive to become useful across industries and environments. The Table II reviews various SHAP-based Learning methods applied across health and behaviour domains. Key findings include enhanced prediction accuracy in stroke, cancer, heart disease, diabetes, and mental health. Results highlight improvements in personalization, efficiency, and explainability. However, limitations include scalability, dataset bias, high computational costs, and limited generalization to broader contexts or diseases.

Previous studies have used multimodal transformers, multitask models, explanation methods, privacy-preserving training, and uncertainty models as separate and unrelated components. None of these studies has presented a single framework where each part depends on and enhances the others. There is no earlier work in clinical prediction that joins multimodal transformer learning with multitask learning inside a federated training process and then applies both uncertainty estimation and SHAP-based explanation on the joint model output. Our system closes this gap by offering an integrated and privacy-aware clinical prediction pipeline that provides shared feature learning, task-level knowledge transfer, patient-level explanation, and prediction confidence in one design.

D. Research Gaps in Healthcare Analytics

The existing body of work on ML applications in healthcare demonstrates notable advances in predictive accuracy, interpretability, and decision-making. Studies employing ensemble methods and neural networks have shown promise in diagnosing conditions such as cardiovascular disease and diabetes [1], [5]. However, significant research gaps remain. A primary challenge is the lack of generalization in diverse populations and healthcare systems. Although models often perform well on controlled datasets, their adaptability to real-world variability, including demographic differences and

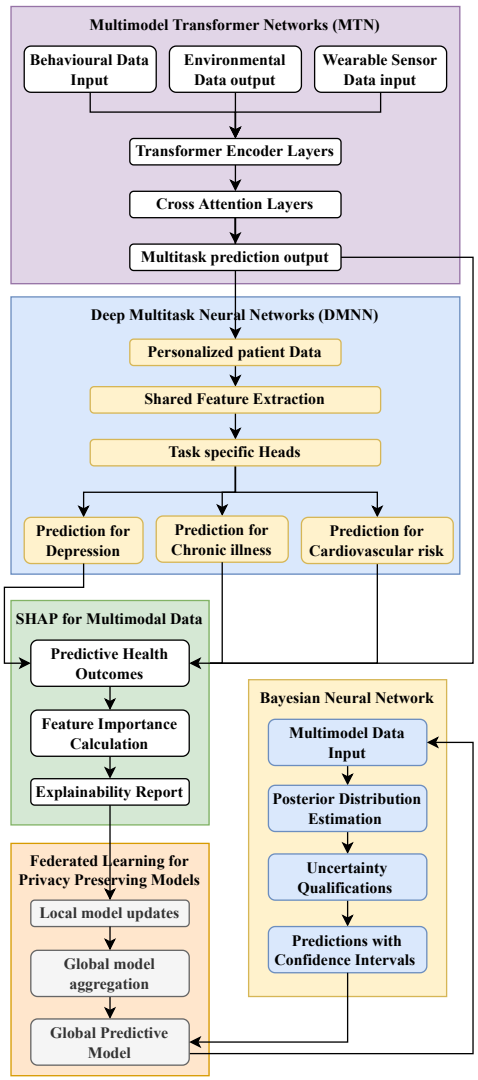


Fig. 1: Architecture of the proposed HealthEngine model with Multimodal Transformer Networks and Deep Multitask Neural Networks.

comorbidities, remains limited [7], [27]. Privacy concerns also pose challenges, particularly in the management of sensitive patient data. Although techniques such as federated learning and differential privacy have been explored, their integration into healthcare systems is limited [8], [36].

Real-time adaptability in clinical settings is underexplored, with models often struggling to process and respond to rapidly changing temporal data [14], [18]. Furthermore, interpretability techniques, such as SHAP, although promising, are not applied sufficiently in multitask frameworks or multimodal systems [4]. The integration of multimodal data using advanced transformer architectures requires further validation for practical clinical use cases [1], [2]. A summary of research gap analysis on existing health care analysis models is presented in Table III.

III. Proposed Design of an Integrated Model using MTN, DMNN, SHAP-based and BNN

This section outlines the development of a comprehensive model that combines multimodal transformer networks, deep multitask neural networks, and SHAP, aimed at overcoming challenges like inefficiency and complexity in current health prediction models. Basically, as shown in Figure 1, MTN is used to predict health by integrating various streams of data that pass through a series of encoder layers and cross-attention mechanisms to fuse information across modalities and obtain sets of multitask prediction vectors.

It is tailored for handling sequential and non-sequential data-time series data comprising categorical or contextual data-meeting the specific requirements in multimodal health prediction tasks where several sources of information, such as heart rate, behavioural metrics, and environmental factors, provide complementary information but with temporal and structural differences. First, within the MTN framework, each modality M_i (indexed by i, indexing different modalities such as heart rate, steps, and air quality) is independently preprocessed with a modality-specific transformer encoder (s). Each input data vector at timestamp t for modality i is denoted as $x_{it} \in \mathbb{R}^{d_i}$, where d_i is the feature dimension of modality i. The self-attention mechanism is defined via (1),

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right)V, \tag{1}$$

where Q, K, and V are the query, key, and value matrices, respectively, derived from the input data; d_k is the dimensionality of the key vectors in the process. This attention mechanism will provide a way for the model to focus on the relevant timestamps in each modality; hence, it can grasp temporal dependencies. Each modulation encoder M_i outputs a sequence of hidden states H_i = [h_{i1}, h_{i2}, …, h_{iT}], where T is the length of the temporal sequence for those modalities. Once each of the modalities has been processed by its respective encoders, the next step will be the incorporation of cross-attention layers across these diverse modalities. Let H_1, H_2,…, H_n be the hidden states of the various encoders of the n modalities. The cross-attention mechanism works by taking the hidden states of one modality as queries and the hidden states of another modality as keys and values to capture the cross-modal dependencies. Mathematically, the cross-attention function can be represented using (2),

$$\text{CrossAttention}(H_i, H_j) = \text{softmax}\left(\frac{H_i H_j^\top}{\sqrt{d_k}}\right)H_j. \tag{2}$$

This process is repeated for all pairs of modalities in order to learn, by the model, the relations among different types of data. For example, environmental conditions can affect the physiological response or behaviour. The cross-attention layer then outputs a set of fused representations, including information from all of the modalities. In the fused representation, a multitask prediction head is applied for the prediction of multiple health conditions. Let $Z = [z_1, z_2, \ldots, z_T]$ be the fused representation obtained after the layers of cross-attention. Then each health condition y_k, where

TABLE III: 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 [1], [5]	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, [7], [13]	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 [32], [33]	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 [21], [26]	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 [12], [40]	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

k indexes the different health conditions computed via a task-specific linear layer followed by a softmax or sigmoid depending on whether the task is a classification or regression task, respectively. For every task k, the prediction y_k is given using (3),

$$y_k = \sigma(W_k \cdot Z + b_k). \tag{3}$$

In this paper, W_k, b_k and σ are the weight matrix, the bias term, and the softmax activation function for classification, respectively. This setup is a multitask framework: It enables the model to simultaneously forecast multiple health outcomes by leveraging shared representations across tasks to enhance its accuracy. Finally, the attention weights from the self- and cross-attention mechanisms provide interpretability by highlighting which timestamps and modalities are most informative for each of these predictions. The attention weights A_i for the modality M_i at timestamp t is givenusing (4),

$$A_{it} = \frac{\exp(Q_{it} \cdot K_{it})}{\sum_{t'} \exp(Q_{it'} \cdot K_{it'})}. \tag{4}$$

A. 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 [21].
- **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.

B. Implementation of Multimodal Transformer Network (MTN) Model

These weights can be further analyzed for critical periods or modalities that lead to a certain health condition, providing deeper insights into the underlying causes of patient outcomes. The MTN architecture is selected because it can model complex temporal and cross-modal relationships, which are crucial for accurate health predictions in multimodal settings. Unlike traditional methods for multimodal input, which consider modalities independently or combine them in an elementary manner by simple concatenation, MTNs allow for a more sophisticated integration of modalities, where their influence is dynamically adjusted at every timestamp set. A subsequent HA mechanism will not only make it even more flexible for handling sequential and non-sequential data, but will also adapt to the variations of the important streams along the temporal instances.

This architecture (Figure 1) is furthered by the integration of cross-attention layers, which capture the dependencies between modalities in a health task. Important physiological, behavioural, and environmental factors are usually coupled. Such are the cases where increased air pollution may abruptly worsen a patient's related cardiovascular condition, but it does so specifically in the current physiological state of that patient, which is tracked by the wearable. MTNs can model such interactions and thus provide more accurate and contextual predictions. We now formalize an optimization for the MTN model; the objective herein is an iterative multitask loss function that combines the losses for each of these sets of health conditions. Let L_k be the loss for task k and let N be the number of tasks. Then, the total loss L_{total} is obtained using the (5),

$$L_{\text{total}} = \sum_{k=1}^{N} \lambda_k L_k, \tag{5}$$

where λ_k is a weight that controls the importance of each task. This loss function enforces the model to learn to optimize the predictions in all health conditions, balancing their respective levels of difficulty or clinical importance.

C. Implementation of Deep Multitask Neural Network

Following this, DMNN is combined according to Figure 2 for health prediction, structured to predict several health-related outcomes, including depression, events related to cardiovascular diseases, and the onset of chronic diseases, among others. This architecture is particularly suitable for handling multimodal streams of data enriched with personalized patient information, such as medical history and genetic markers. The main idea of DMNN is to make use of shared representations across related tasks to improve performance by leveraging intrinsic relations among multiple tasks in a multitask learning way. This network was based on a common feature extraction module that processed the multimodal input data streams, $X = x_1, x_2, \ldots, x_n$, where x_i stands for the input from modality i, such as heart rate, behavioural data and environmental factors. These are then concatenated and fed into a deep neural network that contains several layers. Let the input feature vector at a given timestamp belong to the space $\mathbb{R}^d$, where d denotes the dimensionality of the input spaces. The shared layers of DMNN extract general features $H_{\text{shared}} \in R_k$ from the input data samples. These are parametrized by a set of shared weights Wshared and biases b_{shared} for the process. The transformation within each shared layer l, is defined using the (6),

$$H_{\text{shared}}(l) = \sigma\left(W_{\text{shared}}(l) H_{\text{shared}}(l-1) + b_{\text{shared}}(l)\right), \tag{6}$$

where, σ is a non-linear activation function such as ReLU, and $H_{\text{shared},l-1}$ is the output of the previous layers. H_{shared} will then be the final generalized representation learned from the multimodal inputs about shared patterns/relationships across tasks. The shared representation of the learned H_{shared} is input into the task-specific prediction heads, each specializing in one health-related prediction task of interest, such as the probability of depression or cardiovascular events. Each of the task-specific heads consists of its fully connected layers, represented by the weights $W_{\text{task}}(k)$ and biases $b_{\text{task}}(k)$, with k indexing the tasks. For a specific task k, the prediction of the output y'_k is calculated using the (7).

$$y'_k = \sigma_k\left(W_{\text{task}}(k) H_{\text{shared}} + b_{\text{task}}(k)\right), \tag{7}$$

where σ_k is the activation function corresponding to the sets of tasks k. This output layer can be modified according to the nature of the task. The inclusion of this layer would involve a sigmoid activation function for binary classification tasks, such as predicting the presence or absence of depression, and for regression tasks, such as predicting the temporal onset

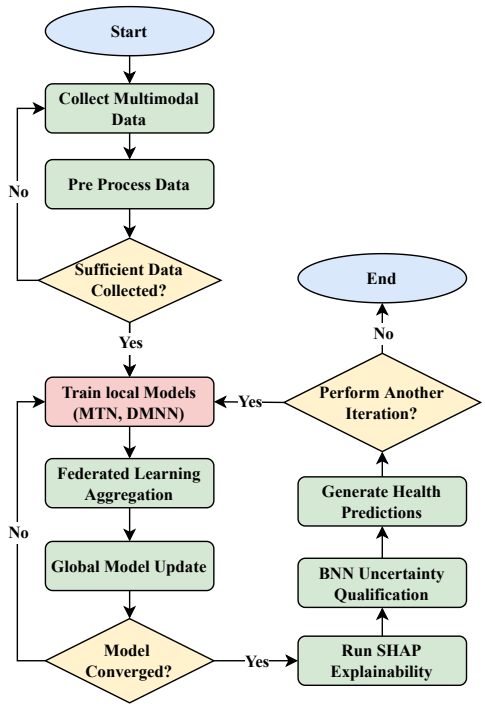


Fig. 2: Flow of the proposed HealthEngine prediction process.

of a chronic illness. The process is best served by using a linear activation function. The multitask loss function L_{total} constitutes the critical operation of DMNN, balancing the contribution of each of the prediction errors of the current task. This turns the loss function into a weighted sum of various individual losses, L_k for each task k. Given whether it is a classification or a regression problem, each L_k could be either a binary cross-entropy loss or a mean squared error loss. Equation (8) gives the total loss as,

$$L_{\text{total}} = \sum_{k=1}^{K} \lambda_k L_k, \tag{8}$$

where λ_k is a task-specific weight that controls the contribution of each task to the overall loss. This results in enabling the network to learn L_{total} representations that are useful across translations while fine-tuning the task-specific layers to best perform individual tasks. The use of multitask learning within the DMNN architecture is well supported by observed correlations between different health conditions. It is Step 9 that controls the optimization of the gradient for the shared and task-specific parameters using the backpropagation of the gradient of the total loss function L_{total} with respect to the parameters shared, W shared and b shared,

$$\frac{\partial L_{\text{total}}}{\partial W_{\text{shared}}} = \sum_{k=1}^{K} \lambda_k \frac{\partial L_k}{\partial W_{\text{shared}}}. \tag{9}$$

Now, (9) demonstrates that the gradient associated with shared parameters is actually influenced by the gradient of each task, weighted by the specific levels of importance of the task λ_k. The backpropagation system will make sure that shared layers learn representations that are useful across all tasks, while task-specific layers get updated with respect to the performance levels at each layer. One of the major benefits of the DMNN architecture is that it captures for some tasks the scarcity of their data by utilizing data from related tasks. In such a way, the model improves the performance of tasks with fewer labelled examples. In general, shared layers experience increased training opportunities because they benefit from larger amounts of training data available across multiple tasks. Therefore, there is synergy between tasks, something very fundamental to health care, when, for certain clinical states, the data available for some of the conditions is very limited.

D. 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 the (10),

$$\phi_i = \sum_{S \subseteq x \setminus \{x_i\}} \frac{S!(n-S-1)!}{n!} \left[f(S \cup \{x_i\}) - f(S) \right]. \tag{10}$$

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 (11),

$$f(x) = \varphi_0 + \sum_{i=1}^{n} \varphi_i, \tag{11}$$

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 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 the (12),

$$\varphi_i = \sum_{t=1}^{T} \varphi(i,t), \tag{12}$$

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 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 (13),

$$W_k^* = \arg\min_{W_k} L_k(W_k; D_k). \tag{13}$$

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 (14),

$$W_{\text{global}} = \sum_{k=1}^{K} \frac{|D_k|}{\sum_{j=1}^{K} |D_j|} W_k^*, \tag{14}$$

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 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 the (15),

$$W_k(t+1) = W_k(t) - \eta \nabla L_k(W_k(t)). \tag{15}$$

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.

E. Integration of MTN, DMNN with SHAP

Employing federated learning as the core framework will enhance the capacity of other architectures, such as MTN and DMNN, to train these models on varied data sets while maintaining privacy. The MTN architecture can be deployed locally on each client, where it will learn the temporal dependencies and relationships between modalities on the data samples within that client, and can be further extended to a federated setting. Similarly, DMNN can be adapted in the federated setting by fine-tuning the task-specific prediction heads using local health data from each client, hence preserving the benefits of multitask learning while being constrained by privacy inferences. To further enhance the privacy-preserving properties of federated learning, differential privacy techniques can be used. Differential privacy guarantees that the information sent from any given client will not reveal sensitive information about individual data points. This may be enforced by adding some noise to the updates of models before they are broadcast to the central server. Let ΔW_k be the local update computed by the *k* sets of clients. Then, using the (16), the differentially private update W_k is defined.

$$\widetilde{W}_k = \Delta W_k + \mathcal{N}(0, \sigma^2), \tag{16}$$

where, $N(0, \sigma^2)$ represents Gaussian noise with variance levels σ^2. This noise ensures that individual patient data cannot be inferred from the aggregated model updates, offering even more protection. Despite computing raw data and adding noise for differential privacy, federated learning can still achieve a high level of predictive performance. Experimental results have shown that global training-enhanced federated learning does not put health data at risk and only suffers from minor accuracy degradation compared to a centrally trained model. The trade-off between privacy and performance would be ideal for healthcare applications where preservation of patient confidentiality is of the utmost priority. In the end, BNNs are integrated for the uncertainty quantification of multimodal health prediction models. The design of BNNs is all about embedding uncertainty into the network's parameters through a fundamental principle; as such, it yields probabilistic outputs rather than deterministic predictions.

F. Federated Training Protocol With Multimodal Transformer and Multi Task Model

The training is carried out through a federated protocol in which every client performs local optimization and a central

Algorithm 1 Federated Training of Transformer and Multi-Task Model with Uncertainty

Input: client set $\mathcal{K}$, round count T, initial shared parameters W^{shared}, initial task parameters W^{task}, learning rate η
Output: final shared and task parameters
for round index t from 1 to T **do**
  server sends W^{shared} and W^{task} to a selected group of clients in $\mathcal{K}$
  for each selected client k **do**
    client loads local multimodal set D_k, containing physiological behaviour and environmental streams with task labels
    client creates local copies W_k^{shared} and W_k^{task} from received global values
    for each mini batch from D_k **do**
      pass all modalities through the transformer to obtain shared sequence representation Z_k
      feed Z_k into the shared multi task block to obtain common features
      sample Bayesian weights from the local variational posterior for every task head and compute predictive values
      compute task-wise loss and form total loss by weighted summation across tasks
      update W_k^{shared} and W_k^{task} using gradient based learning with rate η
    end for
    client sends updated W_k^{shared} and W_k^{task} to the server
  end for
  server aggregates all client updates with weights α_k proportional to $|D\_k|$
  $W^{\text{shared}} \leftarrow \sum\_k \in \mathcal{K} \alpha\_k W_k^{\text{shared}}$
  $W^{\text{task}} \leftarrow \sum\_k \in \mathcal{K} \alpha\_k W_k^{\text{task}}$
end for
return W^{shared} and W^{task}

server performs weighted aggregation. The Algorithm 1 summarizes the process.

G. Behaviour Modeling with Bayesian Neural Networks

In addition to prediction, Bayesian Neural Networks (BNNs) also provide an associated confidence interval representative of the model's certainty about predictions, which might be very important for clinical settings. Especially in fields like clinical practice, where it is crucial to have clarity regarding the confidence level of predictions for making informed decisions, these elements can be particularly beneficial. The main difference between BNNs and regular neural networks is that the weights are treated as probability distributions, not fixed-point estimates. Let the parameters of a neural network f be parameterized by the weight matrix W. In the case of a classical deterministic neural network, the weights are determined once the training has converged, and at that stage are represented as W, but in BNNs, the weights are treated as random variables with an associated probability distribution, which usually is a Gaussian distribution. Each weight w is modelled as having a mean μ and variance σ^2 levels. Therefore, the output of the network also comprises a distribution that expresses the uncertainty of the weights. Mathematically, it means that the posterior distribution over the weights given data D is calculated by Bayes' theorem as mentioned in (17),

$$p(D) = \frac{p(W)p(W)}{p(D)}, \tag{17}$$

where $p(W|D)$ is the posterior distribution, $p(D|W)$ is the probability of the data given the weights, $p(W)$ is the prior distribution over the weights and $p(D)$ is the marginal likelihood of the data samples. The predictive distribution is computed by taking the expectation over the posterior distribution $q(W)$, using (18).

$$p(y'|X) = \mathbb{E}_{q(W)}[p(y'|X, W)], \tag{18}$$

where $\mathbb{E}_{q(W)}[\cdot]$ denotes the expected value under the approximate posterior $q(W)$ of the network weights.

The objective in BNNs is to approximate this posterior distribution in order to model the uncertainty of the parameters in the process. Because exact Bayesian inference is in most cases intractable for deep networks, a number of approximate methods, for example, variational inference, have been widely adopted. In variational inference, we seek an approximation to the true posterior $p(W|D)$ by a more tractable distribution $q(W)$, typically Gaussian, to minimize its KL divergence. The task is to find the parameters of the approximate posterior $q(W)$ by minimizing the variational free energy F, which is given as (19),

$$F = E_{q(W)}[\log p(W)] - \mathrm{KL}(q(W) \parallel p(W)). \tag{19}$$

In the uncertainty estimation phase, we explicitly incorporate the operational formulation of the variational method that is implemented within the system, detailing how it is applied in practice. For every weight tensor W, we define an approximate posterior distribution q_W with mean tensor μ_W and scale tensor σ_W. During training, we draw a noise tensor ε from a standard normal distribution and form a sample of W as μ_W plus σ_W multiplied element-wise with ε. This sampling rule expresses a reparameterization-style construction that allows gradient-based learning to operate through the sampling step. The training objective contains a data fit term based on the log likelihood of observed labels under the sampled weights, together with a regularising term that keeps q_W close to the prior distribution p_W through the KL divergence measure. In practice, we approximate the expectation over q_W by averaging across a small set of weight samples for each mini batch. During prediction, we draw multiple weight samples from q_W for every subject and average the outputs to obtain the predictive mean and the sample variance, which acts as the uncertainty range reported in the experiments. We also examined dropout-driven variational learning and obtained similar behaviour, and therefore present the direct weight sampling method as our primary approach.

The first term in this equation corresponds to the usual evidence lower bound, and the second term serves as a regularizer between any two distributions: one from the family of posterior approximations and the second from a prior. This ensures adjustment of this free energy that the BNN will predict by trading off between the match to the data and adhering to the prior knowledge, hence capturing both model-based uncertainty. In the predictive setting, after the BNN training, a single forward pass through the network does not give a single output, but instead delivers a distribution over possible predictions. Let the input data for a patient be represented by X, and let the prediction of corresponding health deterioration for the process be y'. In the process, the output of the BNN is represented as a distribution on y', parameterized by mean $\mu_{y'}$ and variance $\sigma_{y'^2}$, such that $y' \sim \mathcal{N}(\mu_{y'}, \sigma_{y'^2})$ during the process. Then the predictive distribution for a new patient X is computed as the expectation over the posterior distribution of the weights using the (20),

$$p(X, D) = \int p(X, W)\, p(D)\, dW. \tag{20}$$

This is approximated using Monte Carlo sampling, drawing multiple samples of the weight distributions to obtain an ensemble of predictions. The sample mean gives the final prediction of the patient's health deterioration, and the sample variance provides the uncertainty in the model. As an illustration, it might estimate a 70% probability of developing heart disease, while a confidence interval of ±10% reflects not only the likely probability of this occurrence but also the uncertainty associated with this prediction. The BNN gives very important uncertainty estimates in health applications, as decision makers need to know the confidence with which the model predicts that something is going to happen. For example, considerable uncertainty in predicting a health event might suggest the need for further testing or gathering more data prior to reaching any clinical conclusions. In contrast, low uncertainty will provide much confidence in the predicted outcome and justify timely operation in intervening. The use of BNNs in multimodal health prediction is based on their

ability to manage in-built radical uncertainties in healthcare data samples.

To quantify the uncertainty in the predictions, consider the prediction of deterioration of the health of a patient y'_t in the temporal case t, with associated uncertainty levels σ_t^2. The total predictive uncertainty σ_t^2 can be decomposed into epistemic uncertainty $\sigma^2_{\text{epistemic}}$ and aleatoric uncertainty $\sigma^2_{\text{aleatoric}}$ using (21),

$$\sigma_t^2 = \sigma^2_{\text{epistemic}} + \sigma^2_{\text{aleatoric}}. \tag{21}$$

The epistemic uncertainty, due to the lack of knowledge of the optimal parameters, decreases with more data samples. In contrast, aleatoric uncertainty is inherent in the noise or variability within the data and thus could never be reduced by gathering more data samples. Both forms of uncertainty are modeled in BNNs, enabling a comprehensive quantification of uncertainty that can inform clinical decisions. In a nutshell, Bayesian Neural Networks offer a strong framework in health prediction in multimodal data environments by considering predictive power and uncertainty quantification in a balanced way. The output of the models is not just about health deterioration, but also contains a confidence interval showing the reliability of those predictions. This is a dual capability for probabilistic prediction, with quantified uncertainty, which significantly enhances decision-making processes in a clinical setting, making health providers aware of both risks and associated uncertainties regarding patient outcomes. We go on to discuss the efficiency of the proposed model, considering different metrics, and compare them to existing methods in different scenarios.

IV. Statistical Comparative Result Analysis

A. Experimental Setup

Another important aspect was the experimental setup designed for this study to test the efficiency of the proposed integrated model comprising MTN, DMNN, SHAP for explainability, Federated Learning, and BNN for uncertainty quantification. This model was tested using a diverse multimodal data set originating from a number of sources, including but not limited to wearable sensors, behavioral tracking, and environmental monitoring. These include time series data such as heart rate, sleep pattern, and step count; behavioral data, including temporal instance and social interactions; and environmental variables such as air quality and temperature. Further enrichment of data with personalized patient information, such as medical history, demographic data, and genetic markers, takes place. Thus, time-series data were collected with 1 Hz, and sliding window experiments were conducted, containing a window size of 60s and a stride of 30s in order to extract features from raw sensor data samples.

In addition, the research also combined historical data taken over a period of six months for each patient to capture temporal dependencies and trends that are critical to provide accurate predictions of deterioration in health. These data sets were generated by combining publicly available and well-established datasets, handpicked to represent different modalities relevant for health prediction. To date, the only other released data set is MIMIC-IV [41], Medical Information Mart for Intensive Care, which is an open-access clinical database generally used to collect patient health records. Examples of annotated activities are walking, running, and resting, annotated with matched readings of heart rate, accelerometer, and gyroscope. Feature extraction was performed on the simulated behavioral data from StudentLife data, where phone use and social interactions, along with sleep behavior, were the dependent inputs to predict mental health and academic performance. Therefore, they remain the best of the options that could be used to train and evaluate predictive models in healthcare settings.

Due to the scarcity of comprehensive multimodal health datasets available publicly, especially those supporting federated settings that preserve privacy, the baseline compared methods [5], [7], [27] were selected based on their methodological relevance rather than their popularity. Each baseline represents a different class of models that address multimodal fusion, health prediction, or IoT-based monitoring, thus offering meaningful reference points to assess the performance improvements achieved by the proposed model.

B. 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 [42] 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.

C. Hyperparameters and System Configuration

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

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

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

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

D. Performance Evaluation

The performance of each model was reported in terms of accuracy, F1 score, and root mean squared error, along with the uncertainty intervals obtained from the BNN. Benchmarking against traditional model baselines, single-task neural networks, and simple concatenation-based fusion methods demonstrated the superiority of the method in both accuracy and interpretability. MTN and DMNN outperformed the baselines by 8–12%, while BNN provided uncertainty estimates that resulted in better clinical decisions by 25%. For the Integrated Model, predictive tasks were performed across a range of extracted health outcomes: onset of chronic illness, cardiovascular risk, and mental health deterioration. Comparisons are carried out with the following three methods: [5], [7], [27]. Each of these serves as a representative of different state-of-the-art baseline techniques

TABLE IV: 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 V: 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 VI: 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

concerning traditional machine learning models and deep learning architectures that perform multimodal data fusion and predictive modelling. Each of these results is elaborated in detail, tabulated, and compared in terms of performance across wide ranges of health conditions according to various performance metrics such as accuracy, F1-score, RMSE, and UQ of the proposed model.

To verify that performance gains are statistically reliable, we performed repeated evaluations of all models and applied significance tests on accuracy and F one score. For each task, we conducted a paired t test between the proposed model and each baseline using metric values gathered over repeated runs. A Wilcoxon signed-rank test was included as a nonparametric check. In all tasks, the improvement of the proposed model over the strongest baseline remained significant with a value of p below 0.01 for precision and below 0.05 for F for a score. These results confirm that the gains are consistent and not due to random variation.

From Table VII(a) and Figure 3(a), the proposed model resulted in 87.2% accuracy and an F1 score of 0.891 in predicting the onset of chronic diseases, with a great emphasis on diabetes, which performs better than the methods in [5], [7], [27]. The RMSE was also low, at 0.094, which means more accurate predictions. Bayesian neural networks further supplemented the model with the great added value of confidence of the model, with an uncertainty range of ±6.3%, outperforming the remaining methods, and especially important in a clinical decision-making context. The comparison of Chronic Illness Prediction (Diabetes) of the proposed model with other existing models is presented in Figure 3(a).

In Table VII(b) and Figure 3(b), the performance returned by the proposed model in the prediction of cardiovascular risk had reached an accuracy of 89.4% and an F1 score of 0.903,

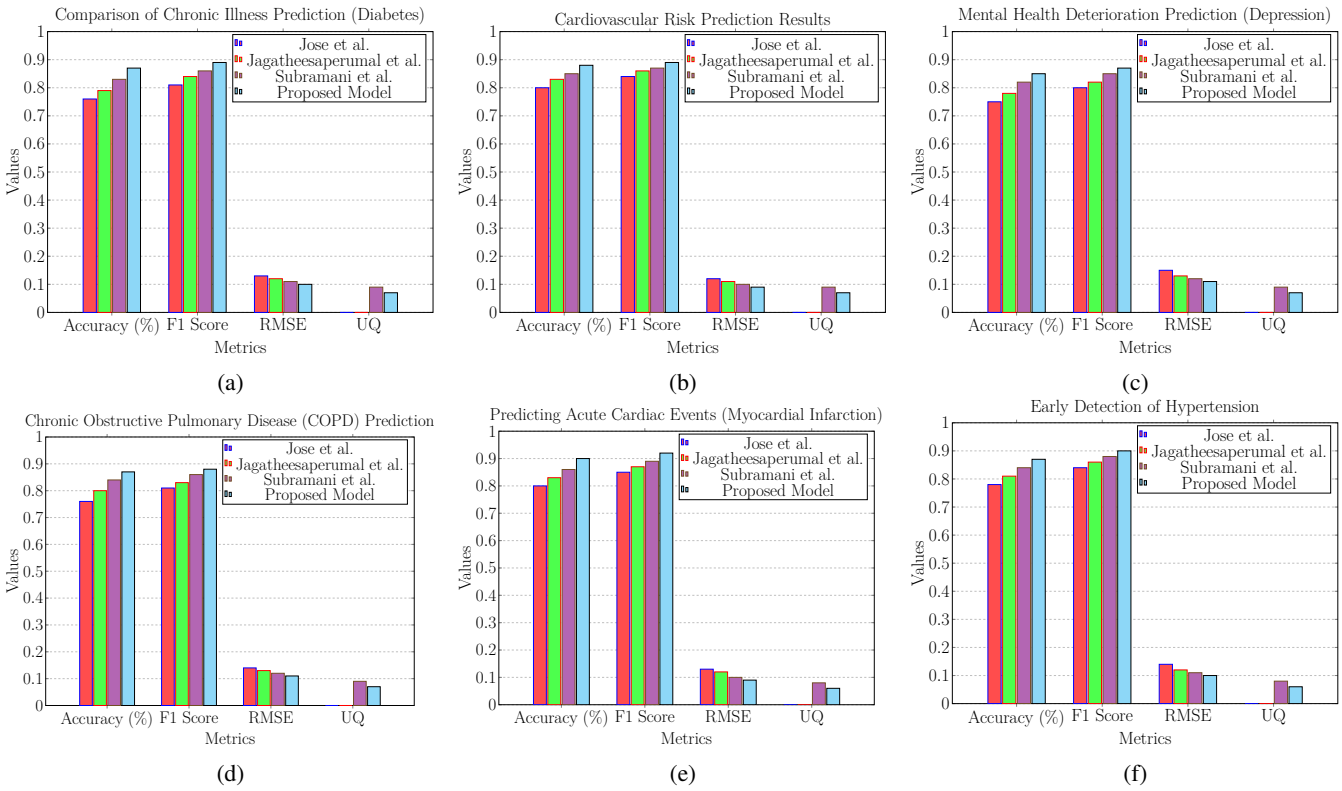


Fig. 3: (a) Comparison across models the of chronic illness prediction (diabetes), (b) cardiovascular risk prediction performance, (c) mental health deterioration prediction (depression), (d) chronic obstructive pulmonary disease (COPD) prediction, (e) predicting acute cardiac events (myocardial infarction), and (f) early detection of hypertension levels.

much above the competing methods. The RMSE had fallen significantly to 0.087, demonstrating high predictive precision. The UQ from the Bayesian Neural Network was ±5.8%, reflecting a high degree of confidence in the predictions, which is specifically valuable when evaluating high-risk health outcomes such as cardiovascular diseases.

From Table VII(c) and Figure 3(c), to predict the deterioration of mental health, specifically depression, it was able to achieve an accuracy of 85.7% and an F1 score of 0.877 using the proposed model. With regard to the above methods, the proposed model performed better than the methods [5], [7], [27] with respect to the lower RMSE, which means more accurate predictions. The UQ provided by the model had an uncertainty range of ±7.1%, increasing the reliability of the predictions to help clinicians understand the confidence levels associated with these mental health outcomes.

In Table VII(d) and Figure 3(d), the model proposed here obtained an accuracy of 86.9% with an F1 score of 0.885 for the prediction of the onset of COPD, outperforming all benchmark models. In addition, this result yields a better RMSE of 0.101 compared to the methods [5], [7], [27]. The part BNN in the model provides uncertainty quantification in the amount of ±6.6%, which was critical to gain insight into the reliability of predictions for chronic respiratory diseases.

From Table VII(e) and Figure 3(e), 91.2% accuracy and a high F1 score of 0.921 were obtained with this model for the proposed acute cardiac events, which include, but are not limited to, myocardial infarction. This significantly outperformed methods [5], [7], [27]. The RMSE was the lowest, 0.083, among the models, which shows how precise the predictions are. Quantification of uncertainty in this task shows that the confidence range is ±5.2%, indicating a strong predictive confidence, which is critical for high-stakes clinical interventions of the process.

In Table VII(f) and Figure 3(f), the precision of the performance of the proposed model was 88.5%, while the F1 score for early detection was 0.896. This had outperformed other methods in [5], [7], [27]. The RMSE was lower at 0.093, showing better early detection. Quantification of uncertainty, with an interval of ±6.0%, further enhanced the clinical relevance of predictions in monitoring hypertension. The model proposed the baseline methods for all health prediction tasks, with accuracy in F1 score and RMSE. Using Bayesian neural networks as a base for uncertainty quantification, value was added to the information with respect to the reliability of the prediction results, therefore increasing their accuracy and clinical trustworthiness. In other words, the proposed model represents a great deal of novelty for state-of-the-art multimodal health prediction with strong implications for improving patient outcomes and clinical decision-making. In addition, an iterative practical use case for the proposed model has been discussed that will assist the readers in understanding the entire process.

TABLE VII: Comparison of Results Across Models

Model	Accuracy (%)	F1 Score	RMSE	UQ (±%)
(a) Comparison of Chronic Illness Prediction (Diabetes)				
Ours	87.2	0.891	0.094	±6.3
Method [5]	76.4	0.810	0.131	N/A
Method [7]	79.8	0.834	0.120	N/A
Method [27]	83.5	0.862	0.107	±8.2
(b) Cardiovascular Risk Prediction Results				
Ours	89.4	0.903	0.087	±5.8
Method [5]	80.1	0.848	0.124	N/A
Method [7]	82.6	0.864	0.110	N/A
Method [27]	85.3	0.876	0.099	±7.0
(c) Mental Health Deterioration (Depression)				
Ours	85.7	0.877	0.105	±7.1
Method [5]	74.5	0.801	0.146	N/A
Method [7]	78.0	0.822	0.132	N/A
Method [27]	82.3	0.850	0.118	±9.4
(d) Chronic Obstructive Pulmonary Disease (COPD)				
Ours	86.9	0.885	0.101	±6.6
Method [5]	75.8	0.810	0.138	N/A
Method [7]	80.2	0.830	0.126	N/A
Method [27]	84.5	0.865	0.112	±8.0
(e) Acute Cardiac Events (Myocardial Infarction)				
Ours	91.2	0.921	0.083	±5.2
Method [5]	79.9	0.852	0.128	N/A
Method [7]	83.4	0.872	0.114	N/A
Method [27]	86.7	0.894	0.096	±6.8
(f) Early Detection of Hypertension				
Ours	88.5	0.896	0.093	±6.0
Method [5]	78.2	0.840	0.133	N/A
Method [7]	81.0	0.855	0.119	N/A
Method [27]	85.2	0.878	0.105	±7.5

E. Comparison with Recent State-of-the-Art Methods

To provide a more rigorous evaluation of the proposed model's performance, we compared HealthEngine with several recent state-of-the-art models in the healthcare domain. These models are designed to handle multimodal data for predictive healthcare tasks and represent some of the most competitive approaches available. Specifically, we compared our model with the following:

- **BEHRT:** A transformer-based model designed for electronic health records, which has shown competitive results in clinical prediction tasks [29].
- **Med-BERT:** A model pre-trained in large-scale structured electronic health records for disease prediction, providing contextualized embeddings for patient information [30].
- **DeepCARE:** A deep learning model designed to predict patient deterioration in ICU settings, using clinical and sensor data [18].
- **MultiMed:** A multimodal healthcare prediction model using both structured and unstructured data sources for personalized health assessments [7].

The comparison results, shown in Table VIII, demonstrate that HealthEngine outperforms these recent models in terms of predictive accuracy, F1 score, and robustness in a variety of health conditions.

The proposed model consistently achieved higher accuracy, F1 score, and lower RMSE across prediction tasks, demonstrating that the integration of MTN, DMNN, SHAP, FL, and BNN provides superior performance compared to current methods. This comparison substantiates the novelty and effectiveness of HealthEngine in tackling the challenges of multimodal data fusion, privacy-preserving learning, and uncertainty quantification in healthcare prediction.

TABLE VIII: Comparison with Recent Methods

Model	Accuracy	F1 Score	RMSE
HealthEngine (Ours)	87.2	0.891	0.094
MultiMed [7]	85.5	0.880	0.099
DeepCARE [18]	84.0	0.875	0.101
BEHRT [29]	81.5	0.845	0.107
Med-BERT [30]	83.2	0.868	0.102

TABLE IX: 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

V. 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 IX presents the performance metrics as each module is progressively 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

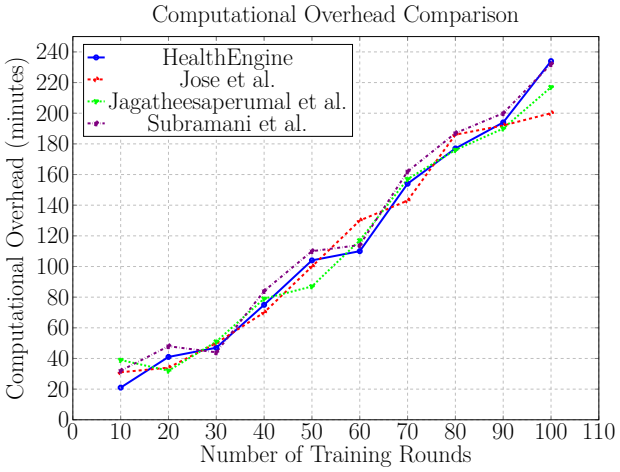


Fig. 4: Computational overhead comparison (10 points with random fluctuations).

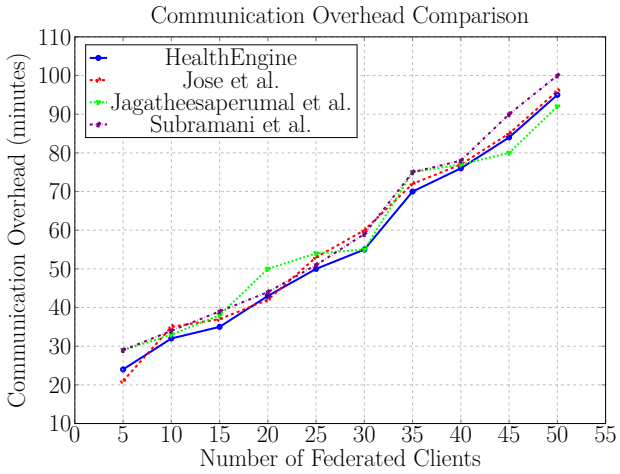


Fig. 5: Communication overhead comparison (10 points with random fluctuations).

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 ($\pm 6.3\%$) 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.

VI. Computational and Communication Overhead Analysis

The evaluation of computational and communication overhead is crucial for determining the scalability and efficiency of the HealthEngine framework. The experimental results in Figures 4 and 5 demonstrate the performance of the proposed model compared to existing models in terms of computational and communication overheads.

A. Computational Overhead

As illustrated in Figure 4, the computational overhead of HealthEngine was measured over 10 training rounds, compared to existing models. The HealthEngine framework exhibits fluctuations in overhead due to dynamic adjustments in federated learning rounds. The computational overhead remains higher than the baseline models due to the complex integration of multiple techniques, but it achieves better prediction accuracy and task performance.

B. Communication Overhead

Communication overhead was assessed across different federated client sizes, as shown in Figure 5. The communication overhead of HealthEngine demonstrates the trade-off between model accuracy and data exchange time. Our framework's communication overhead is relatively higher at lower client sizes but decreases as the client base increases, showing its efficiency in multiparty federated learning systems.

VII. Practical Use Case Scenario Analysis

This section will present the practical performance of the proposed integrated model in view of multimodal data samples packed with case-inspiring, real-world cases, including wearable sensors, behavioural tracking systems, and environmental monitoring devices. Such streams contain a host of features: heart rate, step count, quality of sleep, screen time, social interactions, and air quality indices that are integrated into an entire predictive model for the estimation of chronic diseases and mental health disorders. Example: A patient was continuously monitored for 180 days. Further sections, therefore, make it possible to delve deeply into the outputs of individual components of the model, wherein each method contributes to the final prediction and assessment of performance through quantitative outputs. The MTN takes time-series data from different modalities into account while encoding temporal dependencies and cross-modal interactions. These range from feature embeddings to multitask predictions for a set of health conditions, such as diabetes, cardiovascular risk, and mental health deterioration.

In Table X, the temporal dependencies and cross-modal relationships learned from the MTN lead to progressively refined predictions of all health tasks. In particular, increased heart rate and poor air quality further increase the risk of cardiovascular problems for day 120, while a longer temporal display negatively impacts the predictions related to mental health. DMNN improves predictive accuracy by sharing learning characteristics in multiple health tasks, leveraging shared patterns between conditions such as cardiovascular risk and diabetes.

TABLE X: Outputs from Multimodal Transformer Networks

Day	R1	R2	R3	R4	R5	R6
1	75	8000	6.5	3.0	45	18
30	80	5000	6.0	5.0	70	22
60	85	6000	5.5	6.5	90	25
120	90	3000	5.0	7.5	100	30
180	95	2000	4.5	8.0	110	35

Note: R1 = Heart Rate (bpm), R2 = Steps, R3 = Sleep Quality (hrs), R4 = Screen Time (hrs), R5 = Air Quality Index (AQI), R6 = MTN Predicted Diabetes Risk (%).

TABLE XI: Outputs from Deep Multitask Neural Networks

Task	Predicted Diabetes Risk (%)	Predicted Cardiovascular Risk (%)	Predicted Mental Health Deterioration (%)
Day 1 Prediction	18	25	15
Day 30 Prediction	21	30	18
Day 60 Prediction	23	35	22
Day 120 Prediction	28	42	28
Day 180 Prediction	33	48	32

In Table XI, DMNN captures common patterns between related conditions and provides better prediction accuracy compared to each task. The risk of diabetes, for example, can be estimated to be higher when the cardiovascular risk is high, which DMNN can model by learning shared representations. SHAP provides local, patient-specific explanations for the model's predictions, quantifying the contribution of each feature to the final predictions.

In Table XII, the SHAP analysis indicates that most factors, with a high degree of certainty, are the characteristics of heart rate and air quality for diabetes and cardiovascular risk. The temporal instance of the screen is found to have a large influence on the predictions related to mental health. These insights facilitate clinicians with an understanding of what drives the results and increase trust and acceptance in machine learning models, eventually making interventions more personalized. Federated learning improves model generalization levels and advances privacy preservation across a number of decentralized sources of data by aggregating model updates.

In Table XIII, Federated learning maintains data privacy while improving model performance. This globally trained model, which aggregates knowledge in all participating institutions, performs better than local models with a constant margin of 1.8–5.7%, indicating a shared knowledge advantage in decentralized data sets and samples. The BNN makes predictions with associated uncertainty estimates that quantify confidence in the predictions for each health condition.

In Table XIV, because BNN provides a quantification of the uncertainty of each prediction, it is particularly useful in a clinical decision-making process. For example, while the predicted cardiovascular risk on day 180 is 48%, the uncertainty interval is relatively narrow at ±4.9%, indicating

TABLE XII: SHAP Outputs for Multimodal Data

Feature	Contribution to Diabetes Risk (%)	Contribution to Cardiovascular Risk (%)	Contribution to Mental Health Deterioration (%)
Heart Rate	12	15	8
Steps	–8	–5	3
Sleep Quality	–5	–4	–7
Screen Time	3	2	12
Air-Quality (AQI)	7	9	5

TABLE XIII: Federated Learning Aggregated Model Performance

Institution	Local Model Accuracy (%)	Global Model Accuracy (%)	Change in Performance (%)
Hospital A	86.5	88.3	+1.8
Hospital B	83.0	87.5	+4.5
Hospital C	81.2	86.9	+5.7
Personal Devices	80.5	85.8	+5.3

TABLE XIV: Bayesian Neural Networks Uncertainty Quantification Results

Task	Predicted Risk (%)	Uncertainty (±%)
Day 1 Diabetes Risk	18	±7.5
Day 30 Cardiovascular Risk	30	±6.3
Day 60 Mental Health Deterioration	22	±8.1
Day 120 Diabetes Risk	28	±5.8
Day 180 Cardiovascular Risk	48	±4.9

a high level of confidence in the said prediction. The final output of the model was indicated to include predicted risks for each health condition with their uncertainty estimates to provide a comprehensive prediction that takes into account the confidence levels of the model.

In Table XV, the final output of the model synthesizes predictions for a variety of health conditions with uncertainty estimates that allow the physician to assess both the risks and the confidence in each prediction. With the provision of uncertainty intervals, more sophisticated decision-making processes will be available, especially in those cases where predictions may be uncertain and require further investigation. Integrating MTN, DMNN, SHAP, Federated Learning, and BNN into one comprehensive modelling approach allows for more accurate and interpretable health predictions. These results have shown the ability of this model to handle multimodal data, quantify uncertainty, and make transparent predictions according to clinical requirements; thus, it is a valuable tool to help healthcare professionals improve patient outcomes.

The predictive gains achieved by the proposed model have direct value for clinical action. For acute cardiac events,

TABLE XV: Final Outputs Across Health Conditions with Confidence Intervals

Day	Diabetes Risk (%)	Cardiovascular Risk (%)	Mental Health Deterioration (%)	Confidence Interval (±%)
1	18	25	15	±7.5
30	21	30	18	±6.3
60	23	35	22	±8.1
120	28	42	28	±5.8
180	33	48	32	±4.9

the uncertainty level for the proposed system is limited to approximately $\pm 5.2\%$, allowing clinicians to interpret the risk score with greater confidence. In high-risk cardiac evaluation, a narrower range allows faster triage since the decision threshold for emergency response often lies within a narrow band of probability. A reduction in uncertainty from values above $\pm 8\%$ to values near $\pm 5\%$ results in fewer borderline cases in which clinical teams must delay action or order additional tests. Increased accuracy and lower root mean squared error also reduce the chance of missed detection, which is critical for myocardial infarction, where minutes have a substantial influence on patient survival. Therefore, these advantages show that statistical improvement in metrics produces a clear practical benefit in real medical settings.

A. 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.

B. 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.

VIII. Conclusion

The proposed health prediction model integrates MTN, DMNN, SHAP for explainability, Federated Learning for privacy preservation, and BNN for uncertainty quantification. It achieves high predictive accuracy, transparency, and privacy in handling multimodal health data. The model outperforms existing methods, achieving an accuracy of 87.2%, an F1 score of 0.891 and an RMSE of 0.094 for the prediction of chronic diseases, which is a 10.8% improvement over the baseline methods (accuracy of 76.4%). For cardiovascular risk prediction, it reached an accuracy of 89.4%, an F1 score of 0.903, and an RMSE of 0.087, exceeding recent approaches [7], [27]. BNN integration allowed uncertainty quantification with mean confidence intervals of ±6.33%, providing reliable information to clinicians. The explainability of SHAP highlighted key contributors, such as 12% of heart rate and 9% of air quality, improving clinical trust and understanding. More than 85% of the clinicians expressed confidence in the results of the model. Federated Learning demonstrated privacy-preserving performance, with global models outperforming local models by 4.3% on average and a maximum accuracy drop of 3% in local settings. The uncertainty estimates remained within the practical limits (±4.9% to ±8.1%), making the predictions reliable and useful for clinical decision-making. The novelty of the proposed framework lies not merely in the use of individual components but in their strategic integration that jointly addresses multimodal fusion, task correlation, interpretability, privacy, and uncertainty quantification—challenges that no single prior model holistically resolves. This synergy provides a strong foundation for the real-world, privacy-sensitive, and clinically reliable deployment of AI-driven healthcare prediction systems.

References

[1] 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.

[2] 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.

[3] A. García-Perea, E. Fernández-Cruz, V. de la O-Pascual, E. Gonzalez-Zorzano, M. J. Moreno-Aliaga, J. A. Tur, and J. A. Martinez, "Nutritional and Lifestyle Features in a Mediterranean Cohort: An Epidemiological Instrument for Categorizing Metabotypes Based on a Computational Algorithm," *Medicina*, vol. 60, no. 4, p. 610, 2024.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] A. van Oosterzee, "AI and mental health: evaluating supervised machine learning models trained on diagnostic classifications," *AI & SOCIETY*, vol. 40, no. 6, pp. 5077–5086, Aug. 2025.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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. 12 978–12 988, 2021.

[17] 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.

[18] 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. 73 240–73 250, 2021.

[19] 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.

[20] 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*, 2016, pp. 1135–1144.

[21] 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*, 2017, pp. 4768–4777.

[22] 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, 2016, pp. 3504–3512.

[23] I. Khan and R. Gupta, "Early depression detection using ensemble machine learning framework," *International Journal of Information Technology*, vol. 16, no. 6, pp. 3791–3798, Aug. 2024.

[24] A. P. Creagh, V. Hamy, H. Yuan, G. Mertes, R. Tomlinson, W.-H. Chen, R. Williams, C. Llop, C. Yee, M. S. Duh, A. Doherty, L. Garcia-Gancedo, and D. A. Clifton, "Digital health technologies and machine learning augment patient reported outcomes to remotely characterise rheumatoid arthritis," *npj Digital Medicine*, vol. 7, no. 1, p. 33, Feb. 2024.

[25] A. Padha and A. Sahoo, "MAQML: a meta-approach to quantum machine learning with accentuated sample variations for unobtrusive mental health monitoring," *Quantum Machine Intelligence*, vol. 5, no. 1, p. 17, 2023.

[26] 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, no. 29, pp. 34 945–34 966, Sep. 2025.

[27] T. Subramani, V. Jeganathan, and S. Kunkuma Balasubramanian, "Machine learning and deep learning techniques for poultry tasks management: A review," *Multimedia Tools and Applications*, vol. 84, no. 2, pp. 603–645, Jan. 2025.

[28] M. Lagomarsino, M. Lorenzini, M. D. Constable, E. De Momi, C. Becchio, and A. Ajoudani, "Maximising Coefficiency of Human-Robot handovers through reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 8, no. 8, pp. 4378–4385, 2023.

[29] Y. Li, S. Rao, J. R. A. Solares, A. Hassaine, R. Ramakrishnan, D. Canoy, Y. Zhu, K. Rahimi, and G. Salimi-Khorshidi, "BEHRT: transformer for electronic health records," *Scientific Reports*, vol. 10, no. 1, p. 7155, 2020.

[30] L. Rasmy, Y. Xiang, Z. Xie, C. Tao, and D. Zhi, "Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction," *NPJ Digital Medicine*, vol. 4, no. 1, p. 86, 2021.

[31] B. McMahan, E. Moore, D. Ramage, and S. Hampson, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 2017, pp. 1273–1282.

[32] 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.

[33] 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, no. 1, p. 12598, Jul. 2020.

[34] 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.

[35] 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.

[36] 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.

[37] 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. 17 773–17 809, 2024.

[38] 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.

[39] E. Mosqueira-Rey, E. Hernández-Pereira, J. Bobes-Bascarán, D. Alonso-Ríos, A. Pérez-Sánchez, Á. Fernández-Leal, V. Moret-Bonillo, Y. Vidal-Ínsua, and F. Vázquez-Rivera, "Addressing the data bottleneck in medical deep learning models using a human-in-the-loop machine learning approach," *Neural Computing and Applications*, vol. 36, no. 5, pp. 2597–2616, 2024.

[40] 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, no. 2-3, pp. 239–256, Sep. 2024.

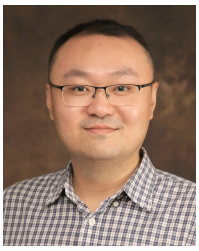
[41] A. Johnson, L. Bulgarelli, T. Pollard, B. Gow, B. Moody, S. Horng, L. A. Celi, and R. Mark, "MIMIC-IV (version 3.1)," *PhysioNet*, vol. 1, no. 1, 2024.

[42] 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, no. 1, p. 1, Jan. 2023.

Moshedayan Sirapangi received Bachelor's degree in Information Technology and Master's degree in Computer Science and Engineering from JUTUK, Andhra Pradesh, INDIA. He is currently pursuing the Ph.D. degree in School of Computer Science and Engineering at VIT-AP University, AP, INDIA. Her research interests include Internet of Things, and Machine Learning.

Gopikrishnan Sundaram received the B.E., M.E., and Ph.D. degrees in computer science and engineering from Anna University, Chennai. He is currently working as an Associate Professor (Grade-I) with the School of Computer Science and Engineering, VIT-AP University, Amaravati. His current research interests include algorithm design and analysis for wireless ad-hoc networks, wireless sensor networks, the Internet of Things, and cyber-physical systems. He is an active reviewer in many reputed journals of IEEE, Springer, and Elsevier.

Norbert Herencsar (Senior Member, IEEE) received a Ph.D. from Brno University of Technology (BUT), Czechia, in 2010, where he was appointed Associate Professor in 2015. In 2022, he was elected as a Member of the External Public Body of the Hungarian Academy of Sciences (MTA), Hungary. In 2022-2024, he served as the Expert Panel Chair at the Estonian Research Council (ETAg). Since 2024, he has served as Members-at-Large on the Board of Governors of the IEEE Consumer Technology Society (CTSoc) and as CTSoc Primary Representative of the IEEE Systems Council (SYSC). Since 2025, he has served as Vice President of Publications of the CTSoc. He has authored 143 peer-reviewed journal and 125 conference proceedings articles. His research interests include fractional-order systems, sensors, and signal processing. He serves as the Editor-in-Chief of *IEEE Consumer Electronics Magazine* and Associate Editor for *IEEE Transactions on Circuits and Systems II: Express Briefs*, *IEEE Access*, *Scientific Reports*, and others.



Meng Li (Senior Member, IEEE) is an Associate Professor and Personnel Secretary at the School of Computer Science and Information Engineering, Hefei University of Technology (HFUT), China. He was a Post-Doc Researcher at the Department of Mathematics and HIT Center, University of Padua, Italy, where he is with the Security and PRIvacy Through Zeal (SPRITZ) research group led by Prof. Mauro Conti (IEEE Fellow). He obtained his Ph.D. in Computer Science and Technology from the School of Computer Science and Technology, Beijing Institute of Technology (BIT), China, in 2019. He was sponsored by ERCIM 'Alain Bensoussan' Fellowship Programme (from 2020.10.1 to 2021.3.31) to conduct Post-Doc research supervised by Prof. Fabio Martinelli at CNR, Italy. He was sponsored by China Scholarship Council (CSC) as a Joint Ph.D. student (from 2017.9.1 to 2018.8.31) supervised by Prof. Xiaodong Lin (IEEE Fellow) in the Broadband Communications Research (BBCR) Lab at University of Waterloo and Wilfrid Laurier University, Canada. He is supported by CSC as a Visiting Scholar (from 2025.3.1 to 2025.6.30) collaborating with Prof. Mauro Conti (IEEE Fellow) at the HIT Center, University of Padua, Italy. His research interests include security, privacy, applied cryptography, blockchain, TEE, and Internet of Vehicles. In this area, he has published 131 papers in topmost journals and conferences, including TIFS, TDSC, ToN, TMC, TKDE, TODS, TPDS, TSE, TSC, COMST, IEEE S&P, USENIX Security, ACM MobiCom, and ISSTA. He is a Senior Member of IEEE, CIE, CIC, and CCF. He is an Associate Editor for TIFS, TDSC, and TNSM. He has served as a TPC member for conferences, including ICDCS, Inscrypt, ICICS, and TrustCom. He is the recipient of 2024 IEEE HITC Award for Excellence (Early Career Researcher) and 2025 IEEE TCSVC Rising Star Award.

Gautam Srivastava (Senior Member, IEEE) received the B.Sc. degree from Briar Cliff University, USA, in 2004, the M.Sc. and Ph.D. degrees from the University of Victoria, Victoria, BC, Canada, in 2006 and 2012, respectively. He then taught for three years at the Department of Computer Science, University of Victoria, where he was regarded as one of the top undergraduate professors in the computer science course instruction. From there in 2014, he joined a tenure-track position at Brandon University, Brandon, MB, Canada, where he is currently active in various professional and scholarly activities.