Explainable Early Prediction of Gestational Diabetes Biomarkers by Combining Medical Background and Wearable Devices: A Pilot Study With a Cohort Group in South Africa

Şefki Kolozali, Sara L. White, Shane Norris, Maria Fasli, *Member, IEEE*, and Alastair van Heerden

***Abstract*—This study aims to explore the potential of Internet of Things (IoT) devices and explainable Artificial Intelligence (AI) techniques in predicting biomarker values associated with GDM when measured 13–16 weeks prior to diagnosis. We developed a system that forecasts biomarkers such as LDL, HDL, triglycerides, cholesterol, HbA1c, and results from the Oral Glucose Tolerance Test (OGTT) including fasting glucose, 1-hour, and 2-hour post-load glucose values. These biomarker values are predicted based on sensory measurements collected around week 12 of pregnancy, including continuous glucose levels, short physical movement recordings, and medical background information. To the best of our knowledge, this is the first study to forecast GDM-associated biomarker values 13 to 16 weeks prior to the GDM screening test, using continuous glucose monitoring devices, a wristband for activity detection, and medical background data. We applied machine learning models, specifically Decision Tree and Random Forest Regressors, along with Coupled-Matrix Tensor Factorisation (CMTF) and Elastic Net techniques, examining all possible combinations of these methods across different data modalities. The results demonstrated good performance for most biomarkers. On average, the models achieved Mean Squared Error (MSE) between 0.29 and 0.42 and Mean Absolute Error (MAE) between 0.23 and 0.45 for biomarkers like HDL, LDL, cholesterol, and HbA1c. For the OGTT glucose values, the average MSE ranged from 0.95 to 2.44, and the average MAE ranged from 0.72 to 0.91. Additionally, the utilisation of CMTF with Alternating Least Squares technique yielded slightly better results (0.16 MSE and 0.07 MAE on average) compared to the well-known Elastic Net feature selection technique. While our study was conducted with a limited cohort in South Africa, our findings offer promising indications regarding the potential for predicting biomarker values in pregnant women through the integration of wearable devices and medical background data in the analysis. Nevertheless, further validation on a larger, more diverse cohort is imperative to substantiate these encouraging results.**

***Index Terms*—Internet of Things healthcare, gestational diabetes mellitus, remote sensing, coupled-matrix tensor factorisation, tree-based regressors, explainable AI models.**

Manuscript received 15 November 2022; revised 29 June 2023, 29 September 2023, and 21 December 2023; accepted 27 January 2024. Date of publication 12 February 2024; date of current version 5 April 2024. The work of Şefki Kolozali and Maria Fasli was supported in part by Business and Local Government Data Research Centre under Grant ES/L011859/1, and in part by Economic and Social Research Council (ESRC), for their assistance in this work. The work of Sara L White was supported by U.K. Medical Research Council under Grant MR/W003740/1. This work was supported in part by Global Challenges Research Fund (GCRF) through A Remote Monitoring System for the Early Detection of Gestational Diabetes Project and in part by the DSI-NRF Centre of Excellence in Human Development, accelerator under Grant ACC2018003. *(Corresponding author: Şefki Kolozali.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Research Ethics Committee under Application No. M190318.

Şefki Kolozali is with the Lecturer in Embedded and Intelligent Systems, University of Essex, CO4 3SQ Colchester, U.K. (e-mail: sefki.kolozali@essex.ac.uk).

Sara L. White is with the Clinical Senior Lecturer at King's College London, WC2R 2LS London, U.K. (e-mail: sara.white@kcl.ac.uk).

Shane Norris is with the Department of Paediatrics, University of the Witwatersrand Johannesburg, Johannesburg 2050, South Africa (e-mail: shane.norris@wits.ac.za).

Maria Fasli is with the University of Essex, CO4 3SQ Colchester, U.K. (e-mail: mfasli@essex.ac.uk).

Alastair van Heerden is with the Human and Social Development Programme at the Human Sciences Research Council (HSRC), University of the Witwatersrand Johannesburg, Johannesburg 2050, South Africa (e-mail: avanheerden@hsrc.ac.za).

Digital Object Identifier 10.1109/JBHI.2024.3361505

I. Introduction

DIABETES is a disorder of carbohydrate (glucose) metabolism. Gestational diabetes mellitus (GDM) is defined as glucose intolerance with onset or first diagnosis during pregnancy. Median estimates of GDM prevalence suggest that globally it complicates between 6 and 13 percent of all pregnancies. This is of significant public health importance as GDM is a risk factor for pregnancy-related maternal and perinatal morbidity and increases the risk that both the mother and her child may develop Type 2 Diabetes. Increasing age and BMI, a family history of type 2 diabetes, ethnicity, and an unhealthy diet and sedentary lifestyle are all factors that increase the risk of developing GDM [1], [2]. The National Institute for Health and Care Excellence (NICE) guidelines recommend that all women with NICE risk factors[1]: BMI $\geq$ 30, high-risk ethnicity, previous

[1]https://www.nice.org.uk/guidance/qs22/chapter/Quality-statement-6-Risk-assessment-gestational-diabetes

GDM, family history of diabetes (i.e. first-degree relative with diabetes), a previous macrosomic baby weighing 4.5 kg or more, should have an oral glucose tolerance test (OGTT) between 24–28 gestational weeks to detect GDM. However, recent trials in [3], [4], [5] suggest that early prediction of GDM risk in the first trimester allows for timely interventions, potentially preventing the condition with moderate lifestyle changes, such as dietary adjustments and increased physical exercise. We hypothesise that it may be possible to identify risk factors in pregnant women early in pregnancy using Internet of Things (IoT) devices and advanced Artificial Intelligence (AI) techniques.

This study aims to forecast biomarker values to help identify pregnant women at an increased risk of GDM based on short Physical Activity (PA) recordings, medical backgrounds, including mental health records, and continuous glucose monitoring (CGM) data collected from participants. For this purpose, we monitored participants between weeks 12 and 14 using Freestyle Libre glucose sensors and Empatica E4 wearable sensors to record physical activities such as sleeping, sitting, eating, walking, climbing stairs, and talking on the mobile phone. These activities were monitored in a controlled environment for three minutes each. We also collected participants' medical history and physiological measurements, such as Body Mass Index (BMI), and blood biomarkers. The primary goal of our study is to investigate the feasibility of using diverse digital data to increase the accuracy of early prediction of biomarker values that are used in the diagnosis of GDM.

The paper is organised as follows: in Section II, we present related medical and telemedicine studies that discuss early detection of GDM. In Section III, we provide details about our proposed framework and the methodology used in data collection and analysis. Section IV presents our evaluations and the results of the proposed methodology for early detection of biomarker values. Finally, in Section V, we discuss the challenges faced during this research, and in Section VII, we conclude and outline our future work.

II. Background

A. Understanding Gestational Diabetes Mellitus

Urbanisation, economic growth, nutritional and food systems changes in Sub-Sahara Africa are all contributing to a rapid increase in obesity and sedentary behaviours [6]. These and other behavioural risk factors have dramatically increased the risk of cardio-metabolic disease. A recent National survey in [7], [8] suggested that 68% of South African women and 31% of men, aged 15 years and older are overweight or obese, with recently published findings suggesting a GDM prevalence of 9.1%. Several observational studies have reported an association between higher levels of Physical Activity (PA) in pre-pregnancy and early pregnancy and a lower prevalence of GDM. The impact of different intensities and duration of exercise on glucose concentration was examined in a study [9], in which an acute effect of PA (i.e. treadmill walking) on blood glucose concentrations in pregnancy in 46 women was reported. It has been reported in a meta-analysis [10] that PA in pre-pregnancy was associated with a greater than 50% reduction in risk of GDM, and early pregnancy PA was associated with a decrease of 25%. Hinkle reported in [11] that pregnant women who suffer from depression during the first trimester had a 1.72-fold higher risk of developing GDM, and a 4.62-fold increased risk of subsequent postpartum depression. In another study [12], worsening sleep patterns were also found to be associated with disturbances in maternal glucose metabolism.

In 2015, the UPBEAT study [13], [14], explored how a lifestyle intervention to improve diet and PA might improve clinical outcomes in pregnant women with obesity. The complex intervention led to a reduction in dietary glycemic load intake and an increase in physical activity which were associated with modest reductions in gestational weight gain and maternal sum-of-skin fold thicknesses. The UPBEAT study provides a baseline for this study. We measured similar biomarkers (e.g, glycated haemoglobin (HbA1c) and adiponectin). However, unlike the UPBEAT study, we also collected continuous glucose measurements in this study. While an individual patient data meta-analysis did not demonstrate an effect of lifestyle interventions on the prevalence of GDM, targeting women at higher risk may be more effective. Indeed, the ESTEEM [3], [4], [5] and RADIEL [12] trials showed that, in contrast to earlier studies where targeting was heterogeneous, the number of women developing GDM was reduced by a moderate lifestyle intervention in pregnant women identified as being at high risk for the disease. In our questionnaire data collection process, alongside survey instruments like the RAND 36-Item Health Survey (SF-36) and the Patient Health Questionnaire (PHQ-9) [15], our study also incorporated the Pregnancy Physical Activity Questionnaire (PPAQ) and the Eating Attitudes Test (EAT-26).

B. Remote Monitoring and Explainable AI for Gestational Diabetes Mellitus Management

Utilisation of AI techniques along with remote monitoring systems, such as wearable devices and smartphones, can provide a great deal of information about the behaviour and well-being of GDM patients in uncontrolled environments. In [16], a systematic review and meta-analysis of telemedicine technologies for diabetes in pregnancy showed streamlined clinical care delivery and improved maternal satisfaction. To date, some studies have focused on using either only the medical background of participants [17] or only blood glucose levels to predict GDM [18], while others have used medical background plus OGTT result to predict Type 2 diabetes after GDM [19], where they found that OGTT results and weight gain during pregnancy are significant variables in the prediction of Type 2 diabetes. Recently, a study was published [20] on the prediction of GDM from medical history and fasting glucose results, in which the authors developed a smartphone app to collect patient data. They applied seven different traditional machine learning techniques in the analysis of the data, including Decision Tree, K-nearest neighbours, Random Forest, and obtained reasonable results for GDM prediction, which varied between 65% and 78% accuracy.

The OGTT screening is usually conducted between weeks 24 and 28, unless a healthcare professional determines high-risk (e.g. previous GDM) at booking in the first trimester. There have been other studies as well, but none of them aimed at forecasting biomarker values based on medical backgrounds, physical activity recordings, and CGM values captured around gestational week 12. For instance, in [21], the authors conducted causal analysis on medical records and a prerecorded dataset. Similarly, in another study [22], a conceptual framework for a telemedicine system was proposed. However, neither of these studies involved the use of sensory devices or machine learning techniques. While some generic telemedicine systems have been proposed in [23], they lack detailed research and do not incorporate the use of

sensory devices. In contrast, in [24], a diabetes dataset from the Kaggle Machine Learning repository was utilised. However, this dataset only includes a limited number of attributes, such as pregnancies, skin thickness, excessive thirst, blood pressure, glucose, smoking, insulin, body mass index, age, and diabetes pedigree function. It is not an extensive dataset and does not encompass the variables used in our study. Additionally, sensory devices were not employed in this study. Importantly, none of these studies aimed to predict GDM biomarker values, neither conceptually nor experimentally.

When selecting an approach, it is essential to strike a balance between interpretability and predictive performance [25]. Given the inherent complexity of deep learning methods and their "black box" nature, recent efforts have been directed towards developing techniques for their interpretability, including methods such as Local Interpretable Model-agnostic Explanations (LIME) [26] and Knowledge Distillation [27] and even the integration of decision trees for explanation purposes [28], [29], [30]. However, it is important to acknowledge that these interpretability approaches come at a cost, introducing additional computational overhead alongside the primary deep learning model. While deep learning methods have gained significant popularity in recent years, traditional techniques continue to demonstrate their effectiveness in extracting meaningful patterns and making predictions from large datasets. Notable among these traditional approaches are k-Nearest Neighbors (k-NN), Bayes Classifiers (BC), and Decision Trees (DT) [31], [32], [33]. Decision Trees, in particular, stand out as highly powerful and explainable models in this context [25], [32], [33], [34], making them a suitable choice for our healthcare study. Therefore, the explainable methods, such as decision trees, continue to offer valuable insights and make accurate predictions for large datasets [35], and even outperform [36] or provide competitive results compared to deep learning methods in some of the studies [37]. Furthermore, they are gaining popularity across various application domains, particularly in healthcare, where the need for explainable techniques is growing [38].

In this study, our primary objective revolves around predicting biomarker values from multimodal datasets between 13 and 16 weeks in advance of the GDM screening test. To achieve this, we have utilised Coupled-Matrix and Tensor Factorisation (CMTF) technique, which facilitates the joint factorisation of multiple datasets represented in the form of higher-order tensors and matrices [39], [40], [41]. It is particularly valuable for multi-modal data analysis and fusion, as well as for integrating information from diverse datasets, uncovering latent structures, and ensuring model explainability. We have employed it in conjunction with tree-based regression techniques, specifically Decision Tree Regression and Random Forest Regression. It's important to note that while Random Forest (RF) is a robust traditional technique still utilised in recent studies [34], [42], it is not inherently explainable. Therefore, we employed it for comparison with Decision Trees, as both are tree-based methods.

To the best of our knowledge, this is the first GDM study that aims to forecast biomarker values associated with the presence of GDM between 13 and 16 weeks in advance using physical activity data (i.e., eating, sitting, walking, walking upstairs, talking on the mobile phone, lying down; E4 Empatica wearable wristband in controlled environment), 2–4 weeks CGM, as well as detailed medical background information (diet and eating behaviour, PHQ, physical activity, and demographic questions).

III. Proposed Methodology

Fig. 1 illustrates the proposed framework, comprising several key components. We provide a detailed explanation of these components in this section. Firstly, Section III-A details the initial and essential step in training our system: data collection. Subsequently, Section III-B involves feature extraction, which focuses on capturing acute, cumulative, and magnitude changes using statistical features such as kurtosis, spectral flux, and spectral energy. Section III-C provides the details of the feature aggregation and feature selection processes utilised to summarise the extracted data and choose the most effective features for the regression models. Alternatively, as discussed in Section III-D, Data Fusion with Coupled Matrix Tensor Factorisation-Alternating Least Squares provides an alternative for this crucial phase, preserving all features in their original form within a 3 rd order tensor shape with no requirement for feature aggregation and selection. Lastly, Section III-E explains the application of Decision Tree Regression and Random Forest Regression, providing a clear demonstration of how these models are utilised to predict continuous biomarker values effectively.

A. Data Collection and Dataset Description

Fig. 2 illustrates the data collection and monitoring timeline for the participants in our study. This data was gathered during two critical time windows: between the 12th and 14th weeks and between the 24th and 26th weeks of pregnancy. It is worth noting that we only used the benchmark data acquired at the onset of the 12th week, as a significant portion of participants refrained from wearing our devices despite reminder calls. Our study's participants were recruited from the antenatal clinic associated with the Chris Hani Baragwanath Teaching Hospital in Johannesburg, South Africa. Eligible participants, defined as those between 12–14 weeks pregnant, overweight or obese, and capable of providing written consent, underwent a comprehensive assessment protocol, including a baseline health questionnaire, anthropometric measurements (weight, height, mid-upper arm circumference, and blood pressure), venous blood sampling for HbA1c, adiponectin, and cholesterol levels, wearing an E4 wristband (Empatica, Boston, USA) for real-time physiological data collection (including heart rate, actigraphy, and temperature), and the application of the Freestyle Libre 2 (Abbott Laboratories, Alameda, CA) CGM sensor to the upper arm. Freestyle Libre 2 uses enzymatic electrochemical reactions. Enzymatic electrochemical glucose sensors typically involve an enzyme, often glucose oxidase, that reacts with glucose in the blood. This reaction generates an electrical signal proportional to the glucose concentration, which the sensor detects and measures [43], [44], [45]. We opted for the upper arm for our recordings based on studies [46] that demonstrated its superior accuracy compared to the back and chest. The Freestyle Libre 2 stores 14 days' continuous data in its in-device memory, with participants receiving a second sensor for application after 14 days.

The devices were either collected by the research assistant or returned by the participant during a scheduled clinic visit. At the end of the second and beginning of the third trimester (i.e. between week 25 and 28), an OGTT was conducted to identify individuals with GDM. The study received approval by the University of the Witwatersrand [M190318] and University of Essex Ethics Committee.

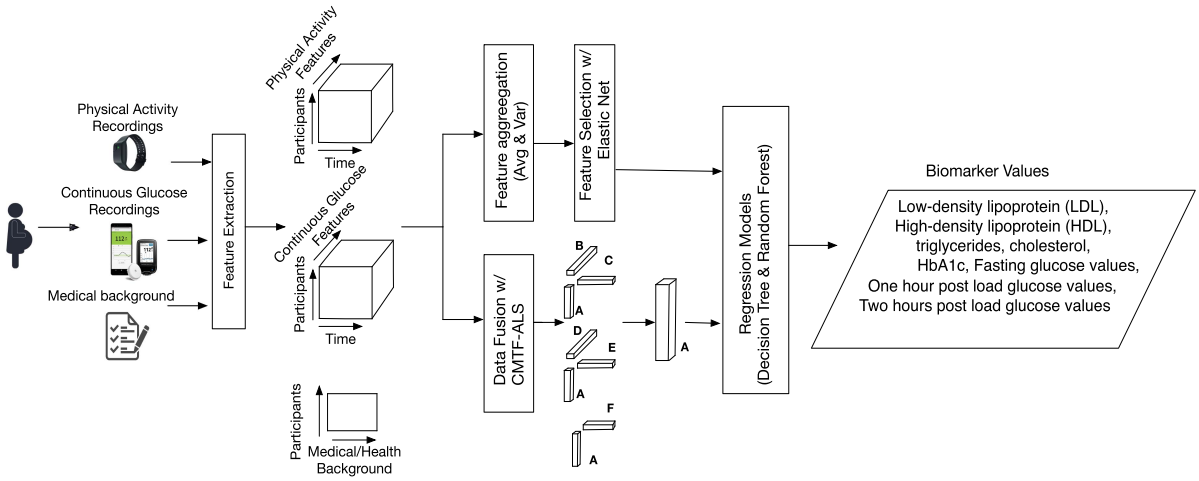


Fig. 1. Proposed framework for the early prediction of biomarker values associated with the presence of GDM involves several key components: Data collection as the initial step; Followed by feature extraction, capturing acute, cumulative, and magnitude changes; Feature aggregation and selection, or data fusion with coupled matrix tensor factorisation-alternating least squares; and concluding with tree-based regression models for effective continuous biomarker value prediction.

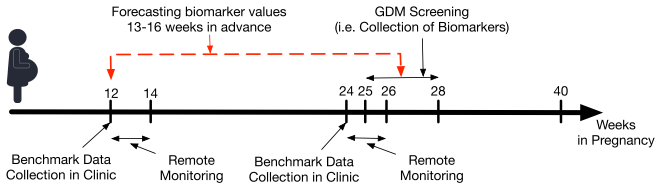


Fig. 2. Illustration of benchmark data collection at 12 and 24 weeks, remote monitoring in the first round between weeks 12–14, and in the second round between weeks 24–26, along with GDM screening between week 25 and 28 during pregnancy at a South African antenatal clinic. The dashed red line represents the forecasting period for the biomarker values collected between week 25 and 28, based on the benchmark data collected in week 12.

The dataset includes information from three sources: background health information, CGM, and Empatica E4 data. Health background data were collected using a survey consisting of multiple validated scales. These included the 1) Eating Attitudes Test (EAT-26) - used to identify the presence of eating disorder risk based on attitudes, feelings and behaviours related to eating, 2) Patient Health Questionnaire (PHQ-9) - a multipurpose instrument for screening, diagnosing, monitoring and measuring the severity of depression, 3) Pregnancy Physical Activity Questionnaire (PPAQ) - a widely used tool for the assessment and measurement of physical activity levels amongst pregnant women, 4) Global Physical Activity Questionnaire (GPAQ) - a WHO tool for the surveillance of physical activity, 5) RAND 36-Item Health Survey (SF-36) - a set of generic, coherent, and easily administered quality-of-life measures, 6) Demographics and Historical Health Questions including questions about smoking, alcohol use, family history of diabetes, and previous pregnancy history or parity. Data was collected from 17 mothers at their enrolment visit.

Continuous Glucose readings were collected using the Freestlye Libre sensor. This is placed on the upper arm and uses subcutaneous, wired enzyme glucose sensing technology to detect glucose levels in interstitial fluid [47]. A reading is taken every 15 minutes and stored on the sensor. An external reader is used to swipe at least every 8 hours by the individual to capture all the data. The data is stored on the reader. We collected data from nine women for 14 complete days, and only four of them continued to complete 28 days worth of continuous glucose data. The other eight mothers were not able to receive sensors due to COVID-19 lockdown restrictions.

The Empatica E4 reads and stores data from multiple sensors in a wrist worn device. The Photoplethysmography (PPG) sensor measures Blood Volume Pulse (BVP), from which heart rate variability can be derived. The Electrodermal activity (EDA) sensor measures the constantly fluctuating changes in certain electrical properties of the skin. The Infrared Thermopile collects peripheral skin temperature and a 3-axis Accelerometer captures motion-based activity. Sampling frequency for all sensors is 64 Hz. Three minutes of recordings were collected in a controlled environment for each of the following activities: eating, sitting, walking, walking upstairs and downstairs, talking on the phone, and lying down.

B. Feature Extraction

We used Discrete Wavelet Transform (DWT) to decompose and reduce the resolution of the time series data collected by wearable devices and CGM sensors. Afterwards, we extracted magnitude, acute and cumulative spectral and statistical features, such as Spectral Flux, Spectral Centroid, Spectral Energy, Average, Median, Kurtosis, Variance, and Skewness. Our goal was to predict the biomarkers fasting glucose measurements, High-density lipoprotein (HDL), Low-density lipoprotein cholesterol (LDL), Triglycerides, Cholesterol, and HbA1c mmol/mol values. The window size was set to 60 seconds and step size was 30 seconds for the E4 measurements; whereas the window size was set as 4 hours for the continuous glucose measurements. Furthermore, we utilised the medical background responses from the participants, which were presented in numerical form within a predefined range. As illustrated in Fig. 1, these numerical representations were subsequently employed as vector inputs

for either the feature selection process or were directly used as inputs for the CMTF technique.

C. Feature Aggregation and Feature Selection

To simplify the data representation and reduce the tensor's order from 3 rd to 2nd, we employed feature aggregation by calculating the average and variance of the features. This transformation played a crucial role in streamlining the data. Subsequently, we employed the Elastic Net technique for feature selection [48]. Elastic Net is a powerful traditional statistical tool commonly used for feature selection that combines the L1 (Lasso) and L2 (Ridge) regularisation techniques. It balances between feature selection and handling correlated features by adding both the absolute value of the coefficients (L1) and the squared value of the coefficients (L2) to the loss function. This regularisation method helps in achieving sparse models while maintaining stability and handling multicollinearity in the data. Our objective was to identify the most relevant features to be utilised in our Decision Tree Regression and Random Forest Regression Models. The selection of these optimal features was a pivotal step in enhancing the performance and accuracy of our regression models.

D. Data Fusion With Coupled-Matrix and Tensor Factorisation-Alternating Least Squares

The coupled-matrix and tensor factorisation (CMTF) algorithms jointly factorise multiple data sets in the form of higher-order tensors and matrices by extracting a common latent structure from the shared mode [39], [40], [41]. There are two main tensor decomposition techniques used in CMTF, namely CP/PARAFAC decomposition and Tucker decomposition. While tucker decomposition is frequently used in relational data analysis, the CP/PARAFAC decomposition is a popular and effective technique to extract the true latent factors forming a tensor [49]. In our study, we chose the PARAFAC decomposition for our analysis, since it has an intuitive interpretation of its latent factors: each component can be seen as soft co-clustering of the tensor, using the high values of vectors a_k, b_k, c_k as the membership values to co-clusters. Given a tensor $\underline{\mathbf{X}}$, CP/PARAFAC decomposition can decompose $\underline{\mathbf{X}}$ as a sum of rank-one tensors. The matrix form representation of CP/PARAFAC is [**A**, **B**, **C**], where the columns of matrix **A**, **B**, **C** are a_k, b_k, c_k vectors, respectively. In the general case, a three mode tensor $\underline{\mathbf{X}}$ may be coupled with at most three matrices $\mathbf{Y}_i, i = 1 \ldots 3$, in the manner illustrated in Fig. 3 for one mode. The optimisation function that encodes this decomposition is given below:

$$\min_{A,B,C,D} ||\underline{\mathbf{X}} - \sum_{k=1}^{K} a_k \circ b_k \circ c_k||_F^2 + ||\mathrm{Z}_1 - AF^T||_F^2 \qquad (1)$$

where $||.||_F$ represents the Frobenius norm calculation and a_k is the $k-$th column of **A**. The idea behind the coupled matrix-tensor decomposition is that we seek to jointly analyse $\underline{\mathbf{X}}$ and $\mathbf{Z}_i$, decomposing them to latent factors who are coupled in the shared dimension. For instance, the first mode of $\underline{\mathbf{X}}$ shares the same low rank column subspace as $\mathbf{Z}_1$. It is expressed through the latent factor matrix **A**, which jointly provides a basis for that subspace. The CP-Alternating Least Squares is one of the most popular algorithms for fitting the CP model [41]. It aims to reduce the challenge into a linear least-squares problem by solving one

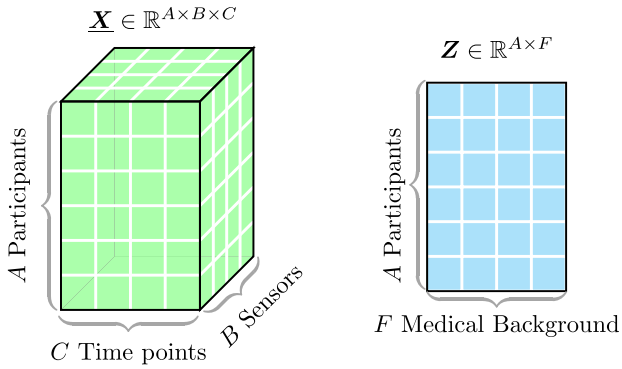


Fig. 3. Coupled matrix - tensor example: Tensors often share one or more modes (i.e. A, participants): $\underline{\mathbf{X}}$ is the activity monitoring sensor recording tensor and **Z** is the medical background matrix. As the vertical line indicates, these two datasets are coupled in the "participants" dimension.

factor matrix at a time by fixing others. While it can be computed in an unconstrained way, it can also be computed by using non-negative least-squares solvers if the tensor contains non-negative values and if one wants to improve the factors' interpretability. In this study, we apply Fast CMTF-ALS technique as proposed in [41].

In order to obtain initial estimates for matrices: **A**, **B**, **C**, we take the PARAFAC decomposition of $\underline{\mathbf{X}}$. As for matrix **D** (and similarly for the rest), it suffices to solve a simple Least Squares problem, given the PARAFAC estimate of A, we initialise as $\mathbf{F} = \mathbf{Z}_1(\mathbf{A}^{\dagger})^T$, where † denotes the Moore-Penrose Pseudoinverse. The Moore-Penrose Pseudoinverse of a matrix is computed as $X^{\dagger} = V\Sigma^{-1}UT$ given the Singular Value Decomposition of a matrix $X = V\Sigma U^T$. The Khatri-Rao product of matrices is shown by using $\odot$, which is a column-wise Kronecker product, $\otimes$, of two matrices.

We applied CMTF in two distinct multi-aspect learning scenarios. In Algorithm 1, our proposed system leverages the power of CMTF to simultaneously learn from two 3 rd order tensors and one matrix. While the matrix represents the medical background of participants, tensors were used for the representation of physical activity recordings and continuous glucose measurements of participants. The equation used in this algorithm is given below:

$$\min_{A,B,C,D,E,F} ||\underline{\mathbf{X}} - \sum_{k=1}^{K} a_k \circ b_k \circ c_k||_F^2 + ||\underline{\mathbf{Y}} - \sum_{k=1}^{K} a_k \circ d_k \circ e_k||_F^2 + ||\mathrm{Z} - AF^T||_F^2 \qquad (2)$$

where **A, B, C, D, E, F** represent the participants, physical activity features, physical activity time, continuous glucose features, continuous glucose monitoring time, medical background of participants, respectively. The input $\underline{\mathbf{X}}$ represents the physical activity tensor, $\underline{\mathbf{Y}}$ denotes the continuous glucose monitoring tensor, and **Z** corresponds to the medical background matrix utilised in the algorithms. In Algorithm 2, our proposed system similarly learns from a 3rd-order tensor and a matrix, employing CMTF to extract valuable insights from both data structures. The

Algorithm 1: CP-CMTF ALS-I.

Input: $\underline{\mathbf{X}}$ of size $I \times J \times K$, $\underline{\mathbf{Y}}$ of size $I \times L \times M$, matrix $\mathbf{Z}$ of size $I \times N$ number of factor R

Output: $\mathbf{A}$ of size $I \times R$, $\mathbf{B}$ of size $J \times R$, $\mathbf{C}$ of size $K \times R$, $\mathbf{D}$ of size $L \times R$, $\mathbf{E}$ of size $M \times R$, $\mathbf{F}$ of size $N \times R$.

Unfold $\underline{\mathbf{X}}$ into $\underline{\mathbf{X}}_{(1)}$, $\underline{\mathbf{X}}_{(2)}$, $\underline{\mathbf{X}}_{(3)}$ and unfold $\underline{\mathbf{Y}}$ into $\underline{\mathbf{Y}}_{(1)}$, $\underline{\mathbf{Y}}_{(2)}$, $\underline{\mathbf{Y}}_{(3)}$

Initialise $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ using PARAFAC of $\underline{\mathbf{X}}$, and $\mathbf{D}$, $\mathbf{E}$ using PARAFAC of $\underline{\mathbf{Y}}$.

Initialize $\mathbf{F}$ using SVD of $\mathbf{Z}$.

while *convergence criterion is not met AND maximum iteration number is not exceeded* **do**

$$\mathbf{A} = \begin{bmatrix} \underline{\mathbf{X}}_{(1)} \\ \underline{\mathbf{Y}}_{(1)} \\ \mathbf{P} \end{bmatrix}^T \left(\begin{bmatrix} \mathbf{B} \odot \mathbf{C} \\ \mathbf{D} \odot \mathbf{E} \\ \mathbf{F} \end{bmatrix}^\dagger \right)^T$$

$$\mathbf{B} = \left[\underline{\mathbf{X}}_{(2)}\right]^T \left(\left[\mathbf{C} \odot \mathbf{A}\right]^\dagger \right)^T$$

$$\mathbf{C} = \left[\underline{\mathbf{X}}_{(3)}\right]^T \left(\left[\mathbf{A} \odot \mathbf{B}\right]^\dagger \right)^T$$

$$\mathbf{D} = \left[\underline{\mathbf{Y}}_{(2)}\right]^T \left(\left[\mathbf{A} \odot \mathbf{E}\right]^\dagger \right)^T$$

$$\mathbf{E} = \left[\underline{\mathbf{Y}}_{(3)}\right]^T \left(\left[\mathbf{A} \odot \mathbf{F}\right]^\dagger \right)^T$$

$$\mathbf{F} = \mathbf{Z}(\mathbf{A}^\dagger)^T$$

Algorithm 2: CP-CMTF ALS-II.

Input: $\underline{\mathbf{X}}$ of size $I \times J \times K$, matrix $\mathbf{Z}$ of size $I \times N$ number of factor R

Output: $\mathbf{A}$ of size $I \times R$, $\mathbf{B}$ of size $J \times R$, $\mathbf{C}$ of size $K \times R$, $\mathbf{F}$ of size $I_2 \times R$.

Unfold $\underline{\mathbf{X}}$ into $\underline{\mathbf{X}}_{(1)}$, $\underline{\mathbf{X}}_{(2)}$, $\underline{\mathbf{X}}_{(3)}$.

Initialise $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ using PARAFAC of $\underline{\mathbf{X}}$.

Initialize $\mathbf{F}$ using SVD of $\mathbf{Z}$.

while *convergence criterion is not met AND convergence criterion is not met AND maximum iteration number is not exceeded* **do**

$$\mathbf{A} = \begin{bmatrix} \underline{\mathbf{X}}_{(1)} \\ \mathbf{P} \end{bmatrix}^T \left(\begin{bmatrix} \mathbf{B} \odot \mathbf{C} \\ \mathbf{F} \end{bmatrix}^\dagger \right)^T$$

$$\mathbf{B} = \left[\underline{\mathbf{X}}_{(2)}\right]^T \left(\left[\mathbf{C} \odot \mathbf{A}\right]^\dagger \right)^T$$

$$\mathbf{C} = \left[\underline{\mathbf{X}}_{(3)}\right]^T \left(\left[\mathbf{A} \odot \mathbf{B}\right]^\dagger \right)^T$$

$$\mathbf{F} = \mathbf{Z}(\mathbf{A}^\dagger)^T$$

optimisation equation used in this algorithm is given below:

$$\min_{A,B,C,F} ||\underline{\mathbf{X}} - \sum_{k=1}^{K} a_k \circ b_k \circ c_k||_F^2 + ||\mathbf{Z} - AF^T||_F^2 \quad (3)$$

After applying the CMTF-ALS algorithm, we utilised the component that represents the participants (referred to as matrix $\mathbf{A}$) as an input for Decision Tree and Random Forest Regressors. Our objective was to utilise these techniques to predict the biomarker values of participants collected between week 25 and 28.

E. Regression Models

We utilised Decision Tree and Random Forest Regressors to estimate biomarker values. Both of these algorithms effectively partition the data based on predictor variables, capturing complex relationships and enabling reliable predictions of biomarker values.

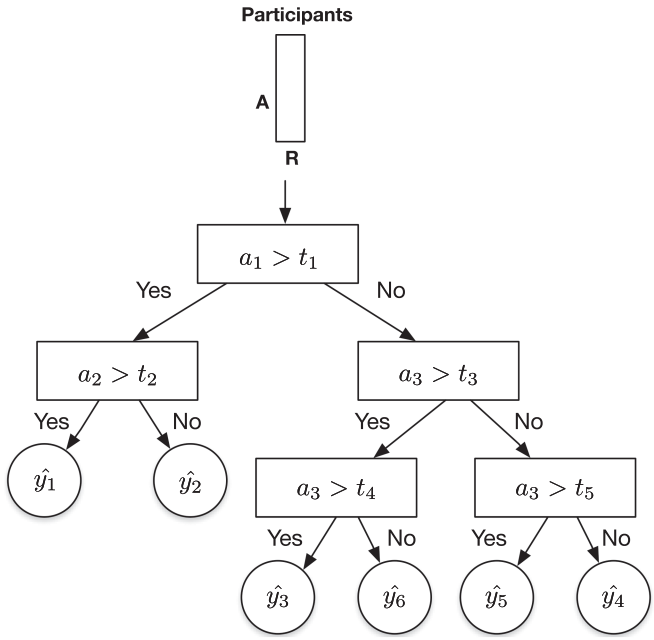


Fig. 4. Illustration of how a decision tree for regression was utilised to predict biomarker values in our experiments.

1) Decision Tree Regression: The DTR algorithm utilises if-then statements to make predictions, which are easily interpretable [50]. Fig. 4 illustrates the application of Decision Tree Regression in our experiments. We consider input features represented by $A = \{a_1, a_2, a_3\}$, which correspond to the participants component obtained from our CMTF model, and R represents the rank used in the CMTF model. The threshold values used for comparison are denoted as $T = \{t_1, t_2, t_3, t_4, t_5\}$, and the predicted biomarker values are denoted as $\hat{Y} = \{\hat{y}1, \hat{y}2, \hat{y}3, \hat{y}4, \hat{y}5, \hat{y}6\}$. The predictions are made using straightforward "if-then" rules applied to the feature vectors. This decision tree approach enables clear decision-making based on the provided thresholds and input features.

First, we divide the predictor space — that is, the set of possible values for $A_1, A_2, ..., A_p$ — into M distinct and non-overlapping regions, $S_1, S_2, \ldots, S_M$. For every observation that falls into the region S_m, we make the same prediction, which is simply the mean of the response values for the training observations in S_m. The goal is to find boxes $S_1, \ldots, S_M$ that minimise the Residual Sum Squares (RSS), given by

$$\sum_{m=1}^{|T|} \sum_{i \in S_m} (y_i - \hat{y}_{S_m})^2 + \alpha|T| \quad (4)$$

where $\hat{y}_{S_m}$ is the mean response for the training observations within the m-th box. $|T|$ indicates the number of terminal nodes of the tree T, S_m is the rectangle or box corresponding to the m-th terminal node, $\hat{y}_{S_M}$ is the predicted response associated with S_m, α controls a trade-off between the subtree's complexity and its fit to the training data. Given that this is equivalent to

constraining the value of $|T|$, we obtain the following equations:

$$\min_{\hat{y}_{S_m}} \left\{ \sum_i (y_i - \hat{y}_{S_m})^2 \right\} \text{ subject to } |T| \leq c_\alpha \quad (5)$$

$$\Delta g' = X(y_i - \hat{y}_{S_m})^2 + \lambda(|T|) \quad (6)$$

where the overall goal is to find the minimum $T, \lambda, \Delta g$, which is a discrete optimisation problem. However, since we're minimising over T and λ, this implies the location of the minimising T doesn't depend on c_α. But each c_α will imply an optimal value of λ.

2) Random Forest Regression: Random forests represent a substantial modification of the bagging technique [50]. It constructs a large ensemble of uncorrelated trees and then average their predictions. Random forests often exhibit similar performance to boosting on various problem domains, but they offer the advantage of being simpler to train and fine-tune. The core idea behind bagging is to average multiple models that are noisy yet approximately unbiased, leading to a reduction in variance. To predict a new point using the random forest algorithm, the following steps are involved: Firstly, a bootstrap sample of size N is drawn from the training data. Then, a random-forest tree is grown using the bootstrapped data by recursively selecting random subsets of variables and finding the best split points. The process continues until a minimum node size is reached. The resulting ensemble of trees is then used to make predictions. When a new point needs to be predicted, it is passed through each tree in the random forest, and the predictions from all the trees are combined to obtain the final prediction. This aggregation can involve averaging the predictions for regression problems as follows:

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B} \sum_{b=1}^{B} T_b(x) \quad (7)$$

where B stands for the number of bootstrap datasets generated from the given tree, b represents bagging estimate over a collection of bootstrap samples, T_b represents the grown random-forest tree T_b from the bootstrapped data.

F. Parameter Settings and Forecasting Biomarker Values

We applied two regression models in our analysis: Decision Tree (DT) and Random Forest (RF) Regressors to predict the biomarker values collected between week 25 and 28 based on the data collected between weeks 12–14. Our experiments were designed to compute the biomarker values of participants collected between week 25 and 28 for the combinations of the following factors:

- Regression Models: Optimised DT and RF Regressors that are tuned for hyper-parameters;
- Preprocessing: CMTF-ALS and Elastic Net;
- The multimodal input variations used as an input in the preprocessing phase are as follows:
  - E4: Physical activity recordings; eating, sitting, walking, walking upstairs, talking on the mobile phone, and lying down (i.e. 3 minutes recording for each activity);
  - CG: Continuous Glucose Monitoring Values (i.e. 14 days of recording per participant);
  - HB: Medical/Health Background
  - E4-HB: Physical Activities and Health Background;
  - CG-HB: Continuous Glucose Monitoring Values and Health Background;
  - E4-CG-HB: Physical Activities, Continuous Glucose Monitoring Values and Health Background;
- Two sets of biomarker values:
  - Low-density lipoprotein cholesterol (LDL), High-density lipoprotein (HDL), Triglycerides, Cholesterol, HbA1c
  - OGTT biomarkers: Fasted Glucose, 1-hour Glucose, 2-hour Glucose

By combining DT and RF with a random search based on the preprocessed data obtained with Elastic Net and CMTF-ALS techniques, we have boosted the performance of RF and DT models. We have used the following parameters in the random search algorithm to find the optimum parameters for the RF model. The number of trees in the DT and RF: 20 random trees in a range from 10 to 200; number of features to consider at every split was automatic and square root functions; 5 different maximum number of levels used in a range from 2 and 100; minimum number of samples required to split a node: 2, 5, 8, and 10; 4 different minimum number of samples required for each leaf node: 2, 6, 8, 10; we activated the bootstrap feature of the algorithm. The parameters used for elastic net were as follows: $\alpha = 1, l1 - ratio = 0.8$. We computed the Mean Squared Error (MSE) and Mean Absolute Error (MAE) for each and overall of the biomarker predictions.

For each regression model, we conducted ten experiments for each set of biomarkers including combinations of the three data modalities as input; sensory recordings collected by Empatica E4 for six different physical activities in a clinical environment (i.e. 3 minutes recording of per activity): CGM data for 2 weeks; Medical/Health background collected in the recruitment process. We obtained the first set of biomarkers (i.e. LDL, HDL, Triglycerides, Cholesterol, HbA1c) for 9 participants, but only collected OGTT biomarkers from 5 out of 9 participants. More importantly, there was complete physical activity recordings and fasting biomarker data for only 4 participants. This was due to COVID-19 affecting our recruitment and data collection process. Thus, to reduce bias in training and testing, we randomly selected participants for training and testing (i.e. seed was fixed) and ran each regression model 10 times, measuring the Mean Squared Error (MSE) and Mean Absolute Error (MAE) between the predicted and actual biomarker values. In all our experiments, we used 60% of the dataset for training and 40% for testing. We used Python programming language and conducted these experiments on an Apple MacBook Pro with the following specifications: 32 GB of RAM, an Apple M1 Max chip featuring 10 cores (8 performance cores and 2 efficiency cores), and 2 TB of memory.

IV. Evaluations

The results are presented in Table I for the predictions of blood test results, and in Table II for the predictions of OGTT results based on the data collected in week 12. We will evaluate our results in three categories: regression models, preprocessing techniques, and biomarkers.

1) Regression Models: When we applied the models on the HbA1c, the error rates for Decision Tree were 0.34 ($\mp$0.06) MSE and 0.44 ($\mp$0.04) MAE, and for Random Forest were 0.36 ($\mp$0.0.09) MSE and 0.43 ($\mp$0.04) MAE. On the other

TABLE I

AVERAGE AND VARIANCE MEAN SQUARED ERROR AND MEAN ABSOLUTE ERROR VALUES OF EACH REGRESSION AND PREPROCESSING MODEL OBTAINED IN THE FORECASTING OF THE BIOMARKER MEASUREMENTS OF LDL, HDL, TRIGLYCERIDES, CHOLESTEROL, AND HBA1C, BASED ON THE SENSOR RECORDINGS COLLECTED BETWEEN WEEK 12–14

Monitoring of Week 12-14														
Classifier	Preprocessing	LDL mmol/L		HDL mmol/L		Triglycerides		Cholesterol		HbA1c mmol/mol		Overall (MSE)		
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	n
DT	CMTF (CG-E4-HB)	0.23 (0.13)	0.39 (0.12)	0.38 (0.29)	0.48 (0.21)	0.49 (0.65)	0.51 (0.36)	0.31 (0.19)	0.45 (0.17)	0.2 (0.21)	0.33 (0.19)	0.3 (0.17)	0.42 (0.12)	9
	CMTF (CG-HB)	0.18 (0.09)	0.35 (0.12)	0.56 (0.69)	0.55 (0.4)	0.63 (0.74)	0.58 (0.4)	0.46 (0.51)	0.49 (0.34)	0.22 (0.15)	0.35 (0.14)	0.38 (0.23)	0.45 (0.14)	9
	CMTF (E4-HB)	0.28 (0.24)	0.42 (0.17)	0.4 (0.65)	0.45 (0.34)	0.33 (0.27)	0.44 (0.19)	0.28 (0.23)	0.4 (0.21)	0.36 (0.37)	0.44 (0.24)	0.34 (0.14)	0.43 (0.09)	9
	E-Net (CG-E4-HB)	0.34 (0.17)	0.5 (0.15)	0.3 (0.23)	0.4 (0.19)	0.4 (0.31)	0.51 (0.2)	0.44 (0.42)	0.51 (0.25)	0.48 (0.45)	0.53 (0.24)	0.4 (0.09)	0.49 (0.05)	9
	E-Net (CG-HB)	0.53 (0.4)	0.6 (0.27)	0.54 (0.55)	0.55 (0.31)	0.27 (0.23)	0.4 (0.17)	0.29 (0.38)	0.4 (0.28)	0.28 (0.2)	0.41 (0.16)	0.36 (0.11)	0.46 (0.07)	9
	E-Net (E4-HB)	0.45 (0.29)	0.55 (0.21)	0.31 (0.22)	0.45 (0.18)	0.38 (0.19)	0.49 (0.14)	0.31 (0.25)	0.44 (0.16)	0.64 (0.54)	0.61 (0.31)	0.44 (0.17)	0.52 (0.1)	9
	E-Net (CG)	0.33 (0.23)	0.47 (0.19)	0.28 (0.2)	0.41 (0.19)	0.29 (0.33)	0.39 (0.24)	0.47 (0.42)	0.55 (0.3)	0.31 (0.37)	0.39 (0.28)	0.33 (0.14)	0.43 (0.1)	9
	E-Net (E4)	0.34 (0.25)	0.48 (0.19)	0.2 (0.08)	0.35 (0.09)	0.39 (0.38)	0.47 (0.22)	0.43 (0.23)	0.53 (0.16)	0.37 (0.3)	0.47 (0.2)	0.35 (0.13)	0.46 (0.08)	9
	E-Net (HB)	0.16 (0.11)	0.32 (0.1)	0.16 (0.11)	0.31 (0.13)	0.15 (0.12)	0.33 (0.15)	0.3 (0.19)	0.44 (0.16)	0.24 (0.36)	0.33 (0.24)	**0.21 (0.12)**	**0.35 (0.08)**	9
RF	CMTF (CG-E4-HB)	0.39 (0.34)	0.49 (0.24)	0.37 (0.54)	0.42 (0.31)	0.18 (0.11)	0.33 (0.11)	0.29 (0.21)	0.43 (0.2)	0.35 (0.28)	0.42 (0.23)	0.32 (0.12)	0.42 (0.09)	9
	CMTF (CG-HB)	0.15 (0.08)	0.32 (0.08)	0.18 (0.2)	0.3 (0.2)	0.56 (1.13)	0.47 (0.47)	0.37 (0.35)	0.48 (0.23)	0.28 (0.22)	0.37 (0.17)	0.3 (0.24)	0.38 (0.12)	9
	CMTF (E4-HB)	0.39 (0.46)	0.48 (0.31)	0.4 (0.6)	0.46 (0.34)	0.28 (0.2)	0.39 (0.14)	0.17 (0.16)	0.32 (0.14)	1.08 (1.57)	0.75 (0.54)	0.55 (0.44)	0.52 (0.16)	9
	E-Net (CG-E4-HB)	0.26 (0.23)	0.4 (0.21)	0.22 (0.23)	0.33 (0.2)	0.28 (0.33)	0.39 (0.19)	0.27 (0.18)	0.4 (0.16)	0.91 (1.54)	0.66 (0.53)	0.46 (0.44)	0.47 (0.15)	9
	E-Net (CG-HB)	0.17 (0.1)	0.33 (0.12)	0.38 (0.5)	0.46 (0.29)	0.45 (0.5)	0.48 (0.28)	0.16 (0.15)	0.31 (0.15)	0.59 (0.63)	0.55 (0.36)	0.38 (0.21)	0.44 (0.13)	9
	E-Net (E4-HB)	0.19 (0.13)	0.33 (0.11)	0.35 (0.5)	0.41 (0.3)	0.22 (0.12)	0.37 (0.12)	0.32 (0.47)	0.44 (0.31)	0.34 (0.28)	0.45 (0.19)	0.29 (0.13)	0.41 (0.09)	9
	E-Net (CG)	0.35 (0.37)	0.47 (0.26)	0.26 (0.2)	0.4 (0.16)	0.31 (0.29)	0.41 (0.2)	0.3 (0.21)	0.42 (0.15)	0.36 (0.52)	0.39 (0.31)	0.31 (0.17)	0.41 (0.11)	9
	E-Net (E4)	0.19 (0.11)	0.35 (0.13)	0.19 (0.16)	0.34 (0.16)	0.2 (0.12)	0.37 (0.1)	0.52 (0.51)	0.55 (0.3)	0.27 (0.31)	0.36 (0.23)	0.27 (0.08)	0.39 (0.05)	9
	E-Net (HB)	0.45 (0.5)	0.52 (0.33)	0.39 (0.3)	0.48 (0.24)	0.34 (0.36)	0.43 (0.25)	0.31 (0.27)	0.43 (0.2)	0.35 (0.39)	0.42 (0.29)	0.36 (0.18)	0.45 (0.13)	9

The average values are presented outside the parenthesis and variance values are presented in the parenthesis. Corresponding meanings of the abbreviations are as follows: DT: Decision Tree, RF: Random Forest; CG: Continuous Glucose; HB: Health/Medical Background; E4: Physical Activity Recordings; E-Net: Elastic Net; CMTF-ALS: Coupled-Matrix Tensor Factorisation with Alternating Least Squares; LDL: Low-density lipoprotein; HDL: High-density lipoprotein. The results are considered to be better when the value of MSE is close to zero. The values indicated in bold font show the best performance obtained for each biomarker and regression model. Old HBA1C values.

TABLE II

MEAN SQUARED ERROR VALUES OF EACH REGRESSION AND PREPROCESSING MODEL OBTAINED IN THE FORECASTING OF THE BIOMARKER MEASUREMENTS OF FASTING GLUCOSE, FASTING GLUCOSE FOR 1 HOUR, AND FASTING GLUCOSE FOR 2 HOURS

Monitoring of Week 12–14										
Classifier	Preprocessing	Fasted Glucose (mmol/L)		1 hour Glucose (mmol/L)		2 hours Glucose (mmol/L)		Overall (MSE)		
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	n
DT	CMTF (CG-E4-HB)	1.2 (2.68)	0.53 (0.75)	1 (1.85)	0.64 (0.61)	1.19 (1.99)	0.72 (0.61)	1.13 (2.16)	0.63 (0.64)	6
	CMTF (CG-HB)	0.93 (1.59)	0.58 (0.53)	0.76 (0.89)	0.63 (0.45)	0.65 (0.64)	0.63 (0.31)	0.78 (0.74)	**0.61 (0.32)**	6
	CMTF (E4-HB)	1.7 (2.92)	0.75 (0.79)	1.14 (1.84)	0.75 (0.64)	1.01 (1.96)	0.7 (0.6)	1.28 (2.13)	0.73 (0.62)	6
	E-Net (CG-E4-HB)	1.05 (1.56)	0.72 (0.46)	0.63 (0.48)	0.63 (0.26)	0.59 (0.32)	0.66 (0.19)	**0.76 (0.46)**	0.67 (0.14)	6
	E-Net (CG-HB)	1.86 (2.87)	0.8 (0.79)	1.08 (1.86)	0.69 (0.64)	1.17 (2)	0.71 (0.63)	1.37 (2.12)	0.73 (0.63)	6
	E-Net (CG-HB)	4.41 (8.88)	1.3 (1.29)	0.67 (0.6)	0.64 (0.3)	0.54 (0.44)	0.58 (0.24)	1.87 (2.91)	0.84 (0.43)	6
	E-Net (E4-HB)	4.41 (8.88)	1.3 (1.29)	0.67 (0.6)	0.64 (0.3)	0.54 (0.44)	0.58 (0.24)	1.87 (2.91)	0.84 (0.43)	6
	E-Net (CG)	3.71 (8.4)	1.05 (1.33)	0.87 (0.45)	0.77 (0.25)	0.69 (0.65)	0.68 (0.28)	1.76 (2.68)	0.84 (0.38)	6
	E-Net (E4)	8.32 (11.48)	1.95 (1.68)	1 (1.56)	0.72 (0.54)	1 (1.65)	0.68 (0.57)	3.44 (3.94)	1.12 (0.69)	6
	E-Net (HB)	1.47 (2.59)	0.79 (0.64)	1.48 (1.88)	0.92 (0.67)	1.24 (1.97)	0.77 (0.62)	1.4 (2.11)	0.82 (0.61)	6
RF	CMTF (CG-E4-HB)	3.67 (8.42)	1.03 (1.34)	0.5 (0.49)	0.55 (0.32)	0.32 (0.19)	0.49 (0.16)	1.5 (2.76)	0.69 (0.43)	6
	CMTF (CG-HB)	1.06 (1.56)	0.65 (0.53)	0.88 (0.84)	0.71 (0.46)	1.57 (3.08)	0.8 (0.77)	1.17 (1.25)	0.72 (0.41)	6
	CMTF (E4-HB)	1.85 (2.85)	0.83 (0.74)	1.13 (1.89)	0.72 (0.69)	0.93 (1.98)	0.63 (0.61)	1.3 (2.13)	0.73 (0.62)	6
	E-Net (CG-E4-HB)	2.73 (3.25)	1.16 (0.81)	1.05 (1.6)	0.7 (0.6)	1.14 (1.58)	0.75 (0.56)	1.64 (1.98)	0.87 (0.6)	6
	E-Net (CG-HB)	1.66 (2.57)	0.86 (0.65)	1.78 (1.85)	1.06 (0.63)	2.43 (3.4)	1.09 (0.86)	1.96 (2.23)	1.01 (0.62)	6
	E-Net (E4-HB)	2.07 (3.24)	0.93 (0.8)	0.91 (1.56)	0.61 (0.59)	0.82 (1.5)	0.63 (0.55)	1.27 (2.01)	0.72 (0.59)	6
	E-Net (CG)	1.47 (1.97)	0.72 (0.64)	0.52 (0.42)	0.53 (0.25)	1.6 (3.11)	0.83 (0.77)	1.19 (1.24)	0.69 (0.38)	6
	E-Net (E4)	3.46 (8.88)	0.99 (1.33)	0.69 (0.47)	0.64 (0.3)	0.49 (0.31)	0.59 (0.26)	1.54 (2.87)	0.74 (0.4)	6
	E-Net (HB)	1.45 (1.41)	0.85 (0.45)	1.1 (0.97)	0.82 (0.44)	2.23 (3.02)	1.05 (0.76)	1.59 (1.19)	0.91 (0.39)	6

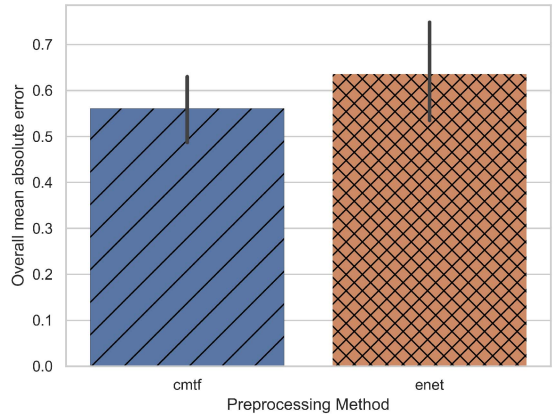
Corresponding meanings of the abbreviations are as follows: DT: Decision Tree, RF: Random Forest; CG: Continuous Glucose; HB: Health/Medical Background; E4: Physical Activity Recordings; E-Net: Elastic Net; CMTF-ALS: Coupled-Matrix Tensor Factorisation with Alternating Least Squares. The results are considered to be better when the value of MSE is close to zero. The values indicated in bold font show the best performance obtained for each biomarker and regression model.

hand, when we applied the models to the OGTT biomarkers, which include glucose measurements obtained before and after a glucose load (i.e. fasting glucose, one-hour glucose, two-hour glucose), the Decision Tree regressor outperformed the Random Forest Regressor with 0.69 MSE and 0.01 MAE. The error rates were 1.46 (∓0.25) MSE and 0.78 (∓0.11) MAE for Random Forest, and 0.77 (∓1.11) MSE and 1.20 (∓0.15) MAE for Decision Tree. Overall, the error rates on average were 0.78 (∓0.51) MSE and 0.58 (∓0.13) MAE for Decision Tree, and 0.92 (∓0.60) MSE and 0.61 (∓0.20) MAE for the Random Forest Regression Model for the overall biomarkers of OGTT and HbA1c predictions. There was only a 0.14 MSE and 0.03 MAE difference between the two models. While the difference was very small, the DT model outperformed the RF model in our experiments.

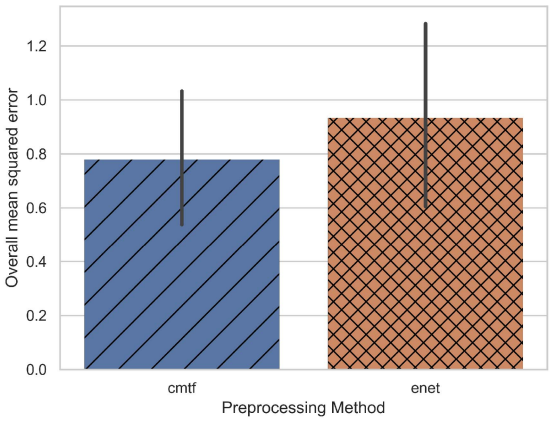
2) Preprocessing Techniques: When we take into account multimodal models for preprocessing approaches, namely, E4-CG-HB, E4-HB, CG-HB, for the HbA1c biomarkers collected between week 25 and 28, the error rates were 1.19 (∓1.86) MSE and 0.68 (∓0.50) MAE for CMTF-ALS and were 1.47 (∓1.95) MSE and 0.80 (∓0.50) MAE for Elastic Net. Similarly, CMTF-ALS slightly outperformed Elastic Net for the OGTT biomarkers as well (see Table II). The error rates were 0.36 (∓0.22) MSE and 0.43 (0.12) MAE for CMTF-ALS and 0.38 (∓0.19) MSE and 0.46 (0.09) MAE for Elastic Net.

Fig. 5 depicts the overall results obtained for CMTF-ALS and Elastic Net. Overall, when we compared the results obtained with CMTF-ALS and Elastic-Net, we found that CMTF-ALS outperformed Elastic Net with an average error rate of 0.16 MSE and 0.07 MAE for all biomarkers. The error rate on average was 0.77 (∓1.04) MSE and 0.56 (∓0.31) MAE for the models involving CMTF-ALS, and 0.93 (∓1.07) MSE and 0.63 (∓0.30) MAE on average for the models involving Elastic Net. However, it is worth pointing out that while we used 10 participants for the models involving CG-HB, we had only 4 participants with physical activity recordings and fasting biomarkers.

3) Biomarkers: Amongst all the biomarkers, the best results were obtained using CG-E4-heath data to predict MSE for LDL and MAE for HDL at the time of the OGTT at 25 weeks, which were 0.29 (∓ 0.23) MSE and 0.43 (∓ 0.18) MAE on average for LDL and 0.32 (∓ 0.34) MSE and 0.41 (∓ 0.23) MAE on



(a) Mean Absolute Error obtained in the predictions with CMTF and Elastic Net



(b) Mean Squared Error obtained in the predictions with CMTF and Elastic Net

Fig. 5. Overall mean absolute error and mean squared error obtained in the predictions when we used CMTF-ALS and Elastic Net approaches.

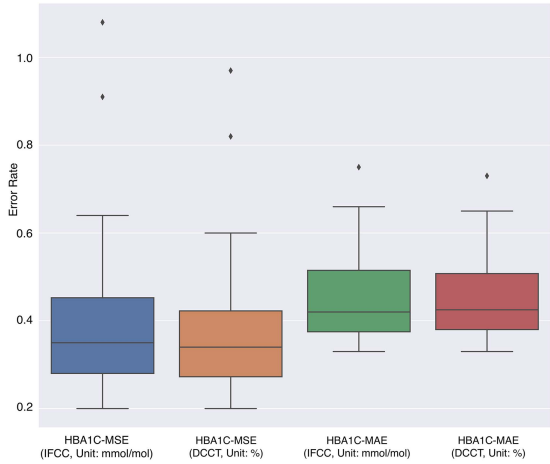


Fig. 6. Error rates in the prediction of IFCC and DCTT units of HbA1c biomarker values.

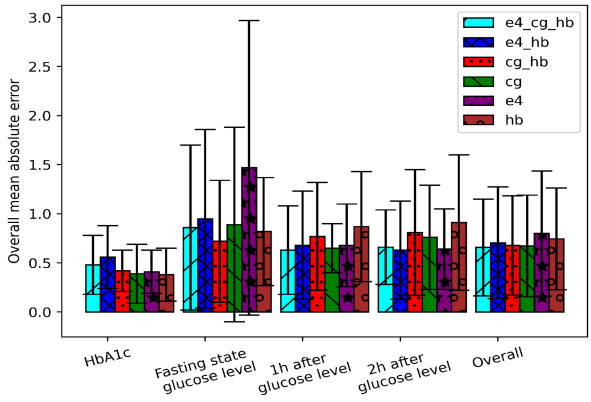


Fig. 7. OGTT and hba1c results for different modalities. The graph represents the summary of predictions for the OGTT and HbA1c values for different combination of data modalities.

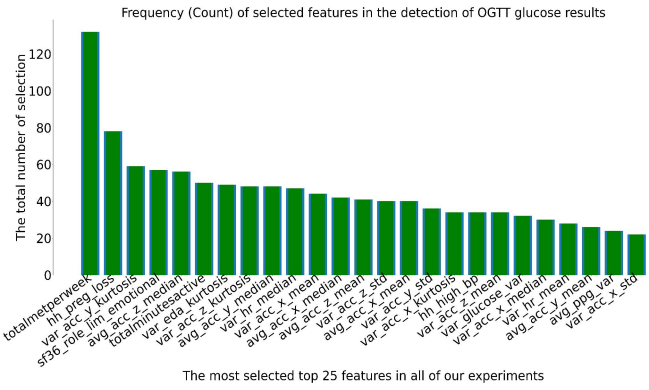


Fig. 8. Most frequently selected features in the feature selection process.

average for HDL. Triglycerides had the third-best results with an average MSE of 0.34 (∓ 0.35) and an average MAE of 0.43 (∓ 0.21). The error rate for Cholestrol was 0.33 (∓ 0.29) MSE and 0.44 (∓ 0.21) MAE on average, and for HBA1C, it was 0.42 (∓ 0.48) MSE and 0.45 (∓ 0.26) MAE on average. The error rate varied for the prediction of the fasting glucose values with an average of 2.44 (∓ 4.28) MSE and an average of 0.91 (∓ 0.86) MAE. In contrast, the error rates were 0.95 (∓1.19) MSE and 0.70 (∓ 0.48) MAE for 1-hour post-load glucose values, and 1.08 (∓1.65) MSE and 0.72 (∓ 0.51) MAE for 2-hour post-load glucose values.

Overall, we obtained the best results for the first set of biomarkers with 0.07 (∓ 0.10) MSE and 0.04 (∓ 0.03) MAE, while the error rates for the OGTT glucose biomarkers were 1.49 (∓ 2.05) MSE and 0.78 (∓ 0.49) MAE in total. Additionally, we predicted both Diabetes Control and Complications Trial (DCCT) units (i.e., %) and International Federation of Clinical Chemistry (IFCC) units (i.e. mmol/mol) for HbA1c. Fig. 6 displays the results obtained for both DCCT and IFCC units. We observed comparable outcomes for HbA1c measurements using both DCCT and IFCC units in our experiments. This similarity could be attributed to the utilisation of min-max scaling for both input and target values in our analysis.

V. Discussion

Overall, none of our participants developed GDM. We obtained promising results when predicting fasting glucose, lipids, and HbA1c biomarkers, one-hour and two-hour post-load glucose values. Unlike the forecasting model applied to health questionnaire data in [20], our model made predictions based on

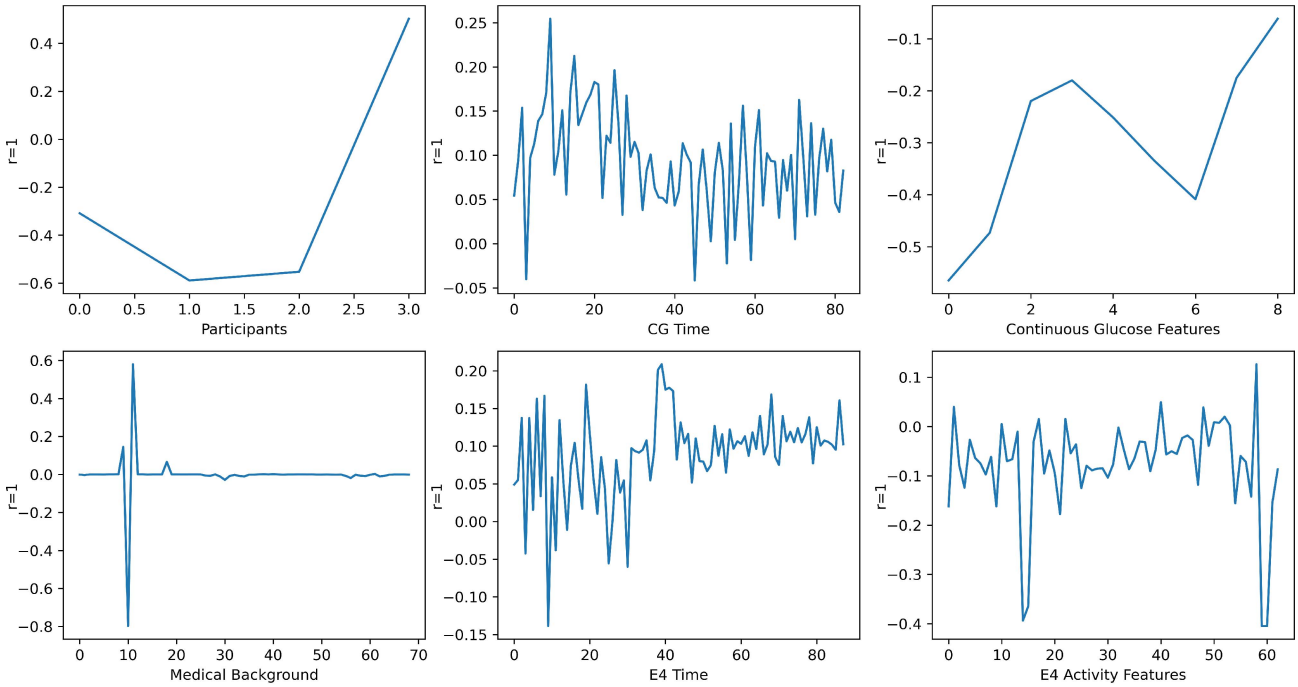


Fig. 9. Visualization of the coupled tensor, E4-CG-HB, was used in the testing phase of HbA1c biomarkers for the decision tree representing participants 2, 3, 5, and 7 in the 9th fold.

Fig. 10. Comparison of the top four results based on mean absolute error for the OGTT biomarkers.

not only health background of participants but also included their physical activity and continuous glucose monitoring recordings.

Fig. 7 shows the results obtained for OGTT and HbA1c biomarker values. The diagnostic parameters for GDM are fasting glucose values, and 1-hour and 2-hours post-load glucose values from the OGTT. While HbA1c is not used in the same way for diagnosis of GDM at 24–28 weeks, it provides information about glycaemic status. When we only focus on these biomarkers, we can see that while CMTF outperformed Elastic Net in the OGTT and HbA1c biomarkers, there was slightly lower performance in the prediction of the OGTT biomarker values. This could be explained by the fact that we had a very small dataset for the glucose biomarkers, whereas there was larger data set for the LDL, HDL, Triglycerides, Cholesterol, and HbA1c measurements. While we will continue investigating the performance of CMTF on larger GDM datasets once we obtain further project funding, it may be possible to argue that traditional statistical techniques perform as well as sophisticated techniques. While we were expecting to see significantly higher results when we used multi-modal data sources simultaneously, our results indicated in Fig 7 that utilising only rich medical background can produce competitive results to multi-modal data sources and sensory recordings, including continuous glucose recordings.

Fig. 8 depicts the most important 25 features selected by Elastic Net, and Fig. 9 shows an example of tensor used in the testing phase, which helps interpret the model used in the predictions. It is worth pointing out that some of the notable frequently selected features in our preprocessing stage with Elastic Net were *totalmetperweek* – total physical activity MET[2] minutes (totalmetperweek: sum of the total MET minutes of activity computed for each setting) – and *total_mins_sedentary*, which reports the mean or median amount of sedentary activity in minutes. In addition, *totalminutesactive* was another variable that was frequently selected, indicating self-reported activity levels of participants. Thus, this small data set shows that self reported sedentary lifestyle information may provide important data for GDM prediction. Similarly, it is also possible to see some spikes for *totalminutesactive* and *totalminutessedentary* in the 10th and 11th position of x-axis in the Medical/Health Background component of the decomposed tensor in Fig. 9.

[2]METs (Metabolic Equivalents) are commonly used to express the intensity of physical activities and are also used for the analysis of GPAQ data. MET is the ratio of a person's working metabolic rate relative to the resting metabolic rate.

Furthermore, similar spikes are noticeable for *continuous glucose monitoring values - flux* and *continuous glucose monitoring values - variance* in the 3 rd and 4th position of x-axis as well as *continuous glucose monitoring values - variance* in the 8th position of the x-axis for Continuous Glucose Monitoring features. Moreover, the highest peaks in the E4 Activity features are represented by *temperature flux* and *temperature kurtosis*, which are not visible in the selected features of Elastic Net. However, it's important to note that this particular visualisation represents only one of the testing datasets used out of the 10-fold experiments.

Fig. 8 demonstrates that other notable selected features were sensory recordings from the accelerometer, PPG, and continuous glucose sensors. It was interesting to see that *sf36_role_lim_emotional* was also a notable feature in our predictions, indicating the importance of participants' emotional state and mental health in the prediction of biomarker values associated with the presence of GDM. This finding is in line with the study in [11] that pregnant women who suffer from depression during their first trimester had a 1.72 fold higher risk of developing GDM.

Our study collected data at week 12 and produced highly predictive models for OGTT values, as well as LDL, HDL, Triglycerides, Cholesterol, and HbA1c biomarkers. Fig. 10 presents the top four results based on mean absolute error for the OGTT biomarkers obtained using our multimodal systems. Among these, the DT-CMTF-(E4-CG-HB) model achieved the best results compared to DT-CMTF-(CG-HB), DT-E-Net-(E4-CG-HB), RF-E-Net-(CG). However, it is important to note that all systems had a mean absolute error of less than 0.83 for predicting all biomarker values, underscoring the robust predictive capability of our models.

Currently, GDM is diagnosed through a fasted glucose tolerance test administered between 24 and 28 weeks of pregnancy. This leaves little time for lifestyle or behavioural interventions to have an effect. It is worth pointing out that the ESTEEM [3], [4], [5] and RADIEL [12] trials showed that, in contrast to earlier studies where targeting was heterogeneous, GDM can be prevented by a moderate lifestyle intervention in pregnant women, who are identified as being at high risk for the disease. Contrary to diagnosis made at the end of the second trimester, predictions made for OGTT values in the first trimester would mean that the women identified as at risk have more time to implement activities and dietary changes [3], [4], [5], [12] that could prevent GDM in the remaining period of pregnancy.

VI. Limitations

It should be noted that our original goal was to monitor physical activities of participants in their daily lives in addition to the recordings obtained in clinical conditions. However, most of the participants, with an exception of one, did not use the E4 wristband wearable devices in their daily lives even though they received notification to actively use these devices. This issue remains one of the biggest challenges in remote monitoring of participants in Internet of Things Healthcare projects, as the goal is to monitor participants in "uncontrolled environments". Instead, we were only able to use the physical activity recordings collected in the clinical environment.

The significant limitations of our study include racial and ethnic disparities and the small sample size, which resulted from COVID-19 lockdown restrictions in South Africa during the data collection period. To address these limitations, future steps will involve replicating these results in a larger sample and assessing whether the same predictive signal emerges earlier than week 12. Utilising this data as an early trigger for health and lifestyle adjustments could potentially contribute to reducing the prevalence of gestational diabetes mellitus.

VII. Conclusion

The early prediction of GDM is an important and active area of research. The use of OGTT and blood based biomarkers, medical history, and IoT data collected at the end of the first trimester appear to be a promising approach towards this goal. In this study, we present our results using remote monitoring and forecasting of biomarker values to help in prediction of GDM. The forecasting was made for biomarker values collected between week 25 and week 28 based on the sensory recordings and medical/health background information collected in week 12 of pregnancy. In general, we obtained promising results for all biomarkers. While the mean squared error and mean absolute error were very low for almost all of the biomarkers, namely, LDL, HDL, Triglycerides, Cholesterol, fasting glucose, and HbA1c values, the error rate was slightly higher for the prediction of the fasting glucose values, an area for future investigation. We found that there were no large differences between DT and RF regression models or between CMTF-ALS and Elastic Net preprocessing models. It should be noted that it is difficult to monitor physical activity of participants in a free living cohort such as ours; participants frequently chose not to use the wristbands. Nonetheless, we utilised physical activity recordings obtained in the clinical environment. We are aware of the fact that our biomarker value predictions are likely to be affected by over-fitting, since they are trained with a relatively small data set per participant, compared to other studies. Nonetheless, we believe that the results are encouraging and first of their kind in the prediction of GDM-associated biomarker values.

Acknowledgment

The authors would like to thank Professor Lucilla Poston, King's College London, for sharing her expertise and providing invaluable guidance throughout the project.

References

[1] C. Chi et al., "Impact of adopting the 2013 World Health Organization criteria for diagnosis of gestational diabetes in a multi-ethnic Asian cohort: A prospective study," *BMC Pregnancy Childbirth*, vol. 18, 2018, Art. no. 152.

[2] W. T. Teh, H. J. Teede, E. Paul, C. L. Harrison, E. M. Wallace, and C. Allan, "Risk factors for gestational diabetes mellitus: Implications for theapplication of screening guidelines," *Australian New Zealand J. Obstet. Gynaecol.*, vol. 51, pp. 26–30, 2011.

[3] H. A. Bassel et al., "Mediterranean-style diet in pregnant women with metabolic risk factors (ESTEEM): A pragmatic multicentre randomised trial," *PLoS Med.*, vol. 16, no. 7, pp. 1–20, 2019.

[4] A. Bolou et al., "Acceptability and adherence to a mediterranean diet in the postnatal period to prevent type 2 diabetes in women with gestational diabetes in the U.K.: A protocol for a single-arm feasibility study (MERIT)," *BMJ Open*, vol. 11, no. 12, 2021, Art. no. e050099.

[5] H. A. Bassel et al., "Effect of simple, targeted diet in pregnant women with metabolic risk factors on maternal and fetal outcomes (ESTEEM): Study protocol for a pragmatic multicentre randomised trial," *Brit. Med. J.*, vol. 6, no. 10, 2016, Art. no. e013495.

[6] S. A. Norris, S. Wrottesley, R. S. Mohamed, and L. K. Micklesfield, "Africa in transition: Growth trends in children and implications for nutrition," *Ann. Nutr. Metab.*, vol. 64, pp. 8–13, 2014.

[7] "South Africa demographic and health survey 2016: Key Indicators," National Department of Health Pretoria, South Africa; South African Medical Research Council, Cape Town, South Africa, and the DHS Program ICF Rockville, MD, USA, Tech. Rep., May 2017. [Online]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjdytbej-CEAxWdSEEAHR40CmgQFnoECBMQAQ&url=https%3A%2F%2Fdhsprogram.com%2Fpubs%2Fpdf%2FFR337%2FFR337.pdf&usg=AOvVaw06YDT4mg1b9tr5vnCZ8lWg&opi=89978449

[8] S. Macaulay, M. Ngobenia, D. B. Dunger, and S. A. Norris, "The prevalence of gestational diabetes mellitus amongst black south African women is a public health concern," *Diabetes Res. Clin. Pract.*, vol. 139, pp. 278–287, May 2018.

[9] S. M. Ruchat et al., "Effect of exercise intensity and duration on capillary glucose responses in pregnant women at low and high risk for gestational diabetes," *Diabetes/Metab. Res. Rev.*, vol. 28, no. 8, pp. 669–678, 2012.

[10] D. K. Tobias, C. Zhang, R. M. V. Dam, K. Bowers, and F. B. Hu, "Physical activity before and during pregnancy and risk of gestational diabetes mellitus," *Diabetes Care*, vol. 43, no. 1, pp. 223–229, Jan. 2011.

[11] S. N. Hinkle, G. M. B. Louis, S. Rawal, Y. Zhu, P. S. Albert, and C. Zhang, "A longitudinal study of depression and gestational diabetes in pregnancy and the postpartum period," *Diabetologia*, vol. 59, no. 12, pp. 2594–2602, Dec. 2016.

[12] E. Huvinen et al., "Effects of a lifestyle intervention during pregnancy and first postpartum year: Findings from the RADIEL study," *J. Clin. Endocrinol. Metab.*, vol. 103, no. 4, pp. 1669–1677, 2018.

[13] L. Poston et al., "Improving pregnancy outcome in obese women: The UK pregnancies better eating and activity randomised controlled trial," Southampton, U.K.: NIHR Journals Library, Apr. 2017. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/28671801/

[14] L. Poston et al., "Effect of a behavioural intervention in obese pregnant women (the UPBEAT study): A multicentre, randomised controlled trial," *Lancet Diabetes Endocrinol.*, vol. 3, pp. 767–777, 2017.

[15] A. Tylee et al., "UPBEAT-U.K.: A programme of research into the relationship between coronary heart disease and depression in primary care patients," *Programme Grants Appl. Res.*, vol. 4, no. 8, pp. 1–172, 2016.

[16] W. K. Ming et al., "Telemedicine technologies for diabetes in pregnancy: A systematic review and meta-analysis," *J. Med. Internet Res.*, vol. 18, no. 11, pp. 1–12, 2016.

[17] S. Polak and A. Mendyk, "Artificial intelligence technology as a tool for initial GDM screening," *Expert Syst. Appl.*, vol. 26, no. 4, pp. 455–460, 2004.

[18] J. J. Liszka-Hackzell, "Prediction of blood glucose levels in diabetic patients using a hybrid AI technique," *Comput. Biomed. Res.*, vol. 32, no. 2, pp. 132–144, 1999.

[19] K. J. Wang, A. M. Adrian, K. H. Chen, and K. M. Wang, "An improved electromagnetism-like mechanism algorithm and its application to the prediction of diabetes mellitus," *J. Biomed. Inform.*, vol. 54, pp. 220–229, 2015.

[20] J. Shen et al., "An innovative artificial intelligence-based app for the diagnosis of gestational diabetes mellitus (GDM-AI): Development study," *J. Med. Internet Res.*, vol. 22, no. 9, pp. 1–11, 2020.

[21] M. R. Neves et al., "Causal dynamic Bayesian networks for the management of glucose control in gestational diabetes," in *Proc. - IEEE 9th Int. Conf. Healthcare Inform.*, 2021, pp. 31–40.

[22] X. Li, X. Gao, Y. Wang, Y. Zhang, Y. Wang, and H. Hu, "Ideas on the construction of the telemedicine system for the gestational diabetes mellitus based on the clinical decision support system," in *Proc. - Int. Conf. Public Health Data Sci.*, 2021, pp. 96–100.

[23] J. Mizera-Pietraszko, "Computer-assisted clinical diagnosis in the official European union languages," in *Proc. IEEE 18th Int. Conf. e-Health Netw., Appl. Serv.*, 2016, pp. 1–5.

[24] R. S. Shankar, V. V. S. Raju, K. Murthy, and D. Ravibabu, "Optimized model for predicting gestational diabetes using ML techniques," in *Proc. 5th Int. Conf. Electron., Commun. Aerosp. Technol.*, 2021, pp. 1623–1629.

[25] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Mach. Intell.*, vol. 1, no. 5, pp. 206–215, 2019.

[26] M. T. Ribeiro and C. Guestrin, ""Why should I trust you?" Explaining the predictions of any classifier," in *Proc. Conf. Knowl. Data Discov.*, 2016, pp. 1135–1144.

[27] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.

[28] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, pp. 56–67, 2020.

[29] X. H. Li et al., "A survey of data-driven and knowledge-aware explainable AI," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 1, pp. 29–49, Jan. 2022.

[30] T. Hickling, N. Aouf, and P. Spencer, "Robust adversarial attacks detection based on explainable deep reinforcement learning for UAV guidance and planning," *IEEE Trans. Intell. Veh.*, vol. 8, no. 10, pp. 4381–4394, Oct. 2023.

[31] A. Swetapadma and A. Yadav, "A novel decision tree regression-based fault distance estimation scheme for transmission lines," *IEEE Trans. Power Del.*, vol. 32, no. 1, pp. 234–245, Feb. 2017.

[32] Z. Chen, S. Tan, U. Chajewska, C. Rudin, and R. Caruana, "Missing values and imputation in healthcare data: Can interpretable machine learning help?," in *Proc. Conf. Health, Inference, Learn.*, 2023, pp. 86–99.

[33] C. Rudin and Y. Shaposhnik, "Globally-consistent rule-based summary-explanations for machine learning models : Application to credit-risk evaluation," *J. Mach. Learn. Res.*, vol. 24, pp. 1–44, 2023.

[34] F. Cheng, Y. Ming, and H. Qu, "DECE: Decision explorer with counterfactual explanations for machine learning models," *IEEE Trans. Vis. Comput. Graph.*, vol. 27, no. 2, pp. 1438–1447, Feb. 2021.

[35] Z. Zhang and C. Jung, "GBDT-MO: Gradient-boosted decision trees for multiple outputs," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 3156–3167, Jul. 2021.

[36] W. Wu, Y. Xia, and W. Jin, "Predicting bus passenger flow and prioritizing influential factors using multi-source data: Scaled stacking gradient boosting decision trees," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 4, pp. 2510–2523, Apr. 2021.

[37] N. Kazemi, M. Abdolrazzaghi, and P. Musilek, "Comparative analysis of machine learning techniques for temperature compensation in microwave sensors," *IEEE Trans. Microw. Theory Techn.*, vol. 69, no. 9, pp. 4223–4236, Sep. 2021.

[38] M. Marsden et al., "Intraoperative margin assessment in oral and oropharyngeal cancer using label-free fluorescence lifetime imaging and machine learning," *IEEE Trans. Biomed. Eng.*, vol. 68, no. 3, pp. 857–868, Mar. 2021.

[39] E. Acar, A. J. Lawaetz, M. A. Rasmussen, and R. Bro, "Structure-revealing data fusion model with applications in metabolomics," in *Proc. IEEE Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2013, pp. 6023–6026.

[40] E. Acar, R. Bro, and A. K. Smilde, "Data fusion in metabolomics using coupled matrix and tensor factorizations," *Proc. IEEE*, vol. 103, no. 9, pp. 1602–1620, Sep. 2015.

[41] E. Evangelos et al., "Turbo-SMT: Parallel coupled sparse matrix-tensor factorizations and applications," *Stat. Anal. Data Mining*, vol. 9, pp. 269–290, 2016.

[42] M. S. H. Lipu et al., "Real-time state of charge estimation of lithium-ion batteries using optimized random forest regression algorithm," *IEEE Trans. Intell. Veh.*, vol. 8, no. 1, pp. 639–648, Jan. 2023.

[43] S. Ollmar et al., "A battery-less implantable glucose sensor based on electrical impedance spectroscopy," *Sci. Rep.*, vol. 13, no. 1, pp. 1–8, 2023.

[44] I. Ahmed et al., "Recent advances in optical sensors for continuous glucose monitoring," *Sensors Diagnostics*, vol. 1, no. 6, pp. 1098–1125, 2022.

[45] S. K. Das, K. K. Nayak, P. R. Krishnaswamy, V. Kumar, and N. Bhat, "Review-electrochemistry and other emerging technologies for continuous glucose monitoring devices," *ECS Sensors Plus*, vol. 1, no. 3, 2022, Art. no. 031601.

[46] R. M. Hall, S. Dyhrberg, A. McTavish, L. McTavish, B. Corley, and J. D. Krebs, "Where can you wear your Libre? Using the freestyle libre continuous glucose monitor on alternative sites," *Diabetes, Obesity Metab.*, vol. 24, no. 4, pp. 675–683, 2021.

[47] A. Heller and B. Feldman, "Electrochemistry in diabetes management," *Accounts Chem. Res.*, vol. 43, no. 7, pp. 963–973, 2010.

[48] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc. Ser. B: Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.

[49] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2008.

[50] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, 2nd ed. Berlin, Germany: Springer, 2009.