

Received 21 December 2023, accepted 31 December 2023, date of publication 3 January 2024,
date of current version 12 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3349496

RESEARCH ARTICLE

Mobile Software Development Kit for Real Time Multivariate Blood Glucose Prediction

MIRIAM K. WOLFF¹, HANS GEORG SCHAATHUN¹, ANDERS LYNGVI FOUGNER², (Member, IEEE),
MARTIN STEINERT³, (Member, IEEE), AND RUNE VOLDEN¹

¹Department of Information and Communications Technology (ICT) and Natural Sciences, Norwegian University of Science and Technology, 6009 Ålesund, Norway

²Department of Engineering Cybernetics, Norwegian University of Science and Technology, 8900 Trondheim, Norway

³Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, 8900 Trondheim, Norway

Corresponding author: Miriam K. Wolff (miriam.k.wolff@ntnu.no)

This work was supported by the Department of Information and Communications Technology (ICT) and Natural Sciences, Norwegian University of Science and Technology (NTNU).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Regional Committees for Medical and Health Research Ethics (REK) in Norway under Approval No. 527497 and SIKT—Kunnskapssektorens Tjenesteleverandør under Approval No. 393579.

ABSTRACT In the field of blood glucose prediction, the literature is abounded with algorithms that demonstrate potential in glucose management. However, these propositions face an issue common to many machine learning algorithms: the repeated reuse of datasets (overfitting) and a tendency to develop algorithms in isolation, detached from practical scenarios. Compounding these challenges is that many insulin pump vendors and continuous glucose monitor vendors use closed and proprietary protocols, restricting researchers' data access and the ability to deploy complex, multivariate optimizers. This study seeks to bridge the gap between theoretical algorithms and their real-world applications by devising a software development kit. This kit collects real-time data from continuous glucose monitors, carbohydrate intake, insulin deliveries from insulin management systems, and metrics like physical activity, stress, and sleep from wearables. Our methodology leverages the open-source insulin management system, Loop, integrated with Apple Health and various wearable devices. Although navigating through diverse communication protocols to link these devices presented challenges, we succeeded in aggregating a comprehensive dataset for blood glucose predictions. To underscore the utility of our software development kit, we executed a technical proof-of-concept on this platform, illustrating real-time, individualized, data-driven multivariate blood glucose predictions. We hope that our platform can contribute to transforming machine learning algorithms from technical developments into actionable tools with real-world benefits in blood glucose management. It provides a foundation for researchers to refine their predictive algorithms and decision support systems within a more dynamic, data-rich environment.

INDEX TERMS Blood glucose prediction, health information systems, machine learning, wearable sensors.

I. INTRODUCTION

Predicting blood glucose (BG) levels is important for managing diabetes and can also be helpful for non-diabetics [1]. BG prediction models can assist in decision-making or be part of a system that helps regulate insulin in the body [2]. Continuous glucose monitoring (CGM) devices can measure and transmit BG levels to a smartphone or insulin pump.

The associate editor coordinating the review of this manuscript and approving it for publication was Chan Hwang See.

BG dynamics are influenced by various factors such as meals, sleep, exercise, stress (MESS) [3], and the menstrual cycle [4]. BG predictions using CGM, insulin infusion data, and physiological sensor data would greatly benefit people interested in better glycemic control and researchers in developing accurate prediction models and utilizing them in products to benefit users. Researchers have attempted to create smartphone applications for predicting future BG levels, but these have been tested only on simulation data or rely heavily on user manual input.

Commercially available wearable sensors can provide raw physiological signals that can non-invasively indicate the occurrence of stress, physical activity, sleep [5], and even menstrual cycle patterns [6]. Developing lightweight physiological sensors will lead to more comfortable and less costly wearable devices to monitor various activities [7]. BG measurements, insulin injections, and carbohydrate intakes are commonly used parameters in BG prediction. This study will focus on research using additional variables from non-invasive sensors or user-reported events. Types of additional inputs can be heart rate, galvanic skin response, skin temperature, sleep events, exercise events, or accelerometer data [8], [9]. Biometric variables might contribute to enhanced glycemic control [10].

Most studies in multivariate BG prediction have only focused on comparing the prediction accuracy to other studies rather than universal requirements for prediction accuracy. To the authors' knowledge, there are no such defined requirements, and such requirements must depend on the use case of the BG predictions. The OhioT1DM dataset [9] is a common benchmark dataset to train and evaluate multivariate BG predictions because of its input feature space containing both wearable sensor data and self-reported live-event data. Researchers have proposed numerous BG prediction algorithms in the literature. However, BG prediction accuracy is statistically and clinically different between datasets [11].

Previous research has established that real-life validation for multivariate BG prediction algorithms is a remaining gap in the research [12] and that most algorithms are trained on small datasets or clinically obtained datasets from highly motivated patients. Assessments in field conditions will help develop products and evaluate and compare BG prediction models.

Researchers have attempted to create mobile systems that display BG predictions. In 2020, Krivenstov et al. developed Diabits [13], a smartphone application directly connected to CGM data. Users could input meals, insulin, and physical activity manually. They claim that most users need to provide more manual inputs; hence, CGM values are the most significant factor in the predictive models. The 280 most long-standing users observed a correlation between increased frequency of app use and improved standard BG control metrics. He et al. [14] created a smartphone application, CausalBG, that could collect CGM data and other external factors impacting BG levels. Insulin, meals, and other factors were based on manual inputs, although the smartphone-embedded sensors could automatically record sleep and calorie expenditure. The BG predictions were tested on three subjects for one day, showing promising results in terms of prediction accuracy. Li et al. deployed their BG prediction on an Android mobile phone [15]. Zhu et al. implemented a wearable wrist-worn device to run a machine-learning BG prediction model in real-time [16]. Li et al. and Zhu et al. evaluated their smartphone system on simulated data. Researchers still need to address the issue of automatically reading data from various sources of wearable

sensors and insulin management systems (IMS) in mobile BG prediction systems. It is crucial to fill this gap to help users avoid the burden of manual inputs while conducting experiments in real-life scenarios.

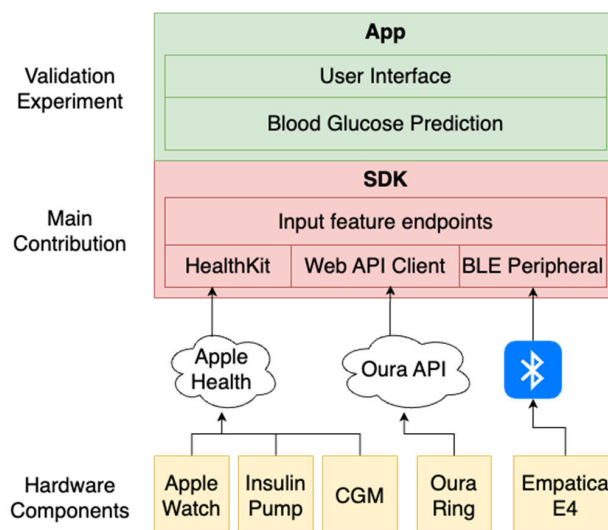


FIGURE 1. Visual representation of key components within the system, highlighting the distinction between the primary contributions and the validation experiment.

In this study, we aim to fill a gap in real-time data access from wearable devices by developing a modular software development kit (SDK) that can integrate with arbitrary blood glucose prediction algorithms. Unlike previous approaches, our SDK enables reading insulin pump data and other sensor inputs, eliminating the need for manual user inputs. We have illustrated the main components of the study in Fig. 1. Our research question focuses on whether our SDK can capture the essential factors influencing BG dynamics, including MESS and menstrual cycles. To address this, our SDK interfaces with three distinct peripherals: HealthKit, Empatica SDK, and Oura API. The SDK provides data access endpoints and can be adapted to alternative predictive methodologies. We evaluate the SDK by deploying a machine learning prediction model, a Ridge Regressor, to a smartphone application. The proof-of-concept app displays real-time BG predictions and can export experimental data for further analysis.

II. METHODS

In this study, our aim is to develop an SDK capable of interfacing with various wearable devices to provide essential inputs for BG prediction. We structure our study as follows: First, we identify the required inputs through a literature review. Following this, we design and implement the system, including the SDK. For validation, we create a proof-of-concept smartphone application utilizing the SDK to predict BG in real-time while also collecting and exporting data for analysis. In the validation experiment, one participant, an author of this study, wears the system for a continuous 24-

hour period. This validation phase adopts an observational approach. Lastly, we qualitatively analyze the results of our observational validation experiment to assess our study objectives.

In accordance with ethical guidelines, the human subject involved in this study provided their informed consent to participate.

A. INTRODUCTION OF SYSTEM

Our primary objective is to develop an SDK that facilitates access to diverse data sources for BG prediction. This achievement holds significance as current technology lacks a real-time method for predicting BG levels without relying on manual user inputs. While previous studies have explored various prediction approaches, ranging from linear models to deep neural networks [17], our software application, built using this SDK, serves as an experimental platform to address this gap. Fig. 1 provides an overview of the system developed in this study.

TABLE 1. A list of the required variables for the SDK and which factors of impact on the BG dynamics the variables can explain. Green background means that the relationship is well established in the literature. Yellow cells refer to smaller studies, and further research on how to apply the information in BG prediction might be needed.

Input feature	Meal	Exercise	Sleep	Stress	Menstrual Cycle
CGM values	[3]	[3]	[3]	[3]	[28], [29]
Insulin deliveries	[3]	[3]	[3]	[3]	[28], [29]
Carbohydrate inputs	By def.				
Energy expenditure		By def.			
Heart rate	[30]	[31]	[31]	[32]	
Heart rate variability	[33]	[34]	[34]	[32]	[6]
Galvanic skin response				[32]	
Accelerometer	[35]	[31]	[31]		
Skin temperature	[36]		[31]	[32], [37]	[6]
Sleep duration			By def.		
Workouts		By def.			

In pursuit of our research objectives, we identified the essential data input sources for our system through a review of the current literature. Table 1 presents a comprehensive list of the identified input variables, along with references to studies elaborating on the influence of these features on BG dynamics. The selection of input types was made strategically to encompass significant factors affecting BG dynamics, specifically focusing on MESS and the menstrual cycle. These input features conform to existing multi-input artificial pancreases [8] and are integral components of a benchmark dataset for predicting BG levels using multiple inputs [9].

The core purpose of our SDK is to provide a proof-of-concept for evaluating various BG prediction algorithms in real-time scenarios, while minimizing the need for manual user inputs. Therefore, while ensuring functionality and

effectiveness, our emphasis is not on achieving the high degree of robustness required for clinical-grade applications [18]. Such clinical-grade requirements represent a substantial undertaking beyond the scope of our current study.

B. OVERVIEW OF SYSTEM

1) HARDWARE COMPONENTS

The hardware components used in our system are listed below. Fig. 2 shows the experimental setup with all hardware components and a smartphone application predicting BG levels.

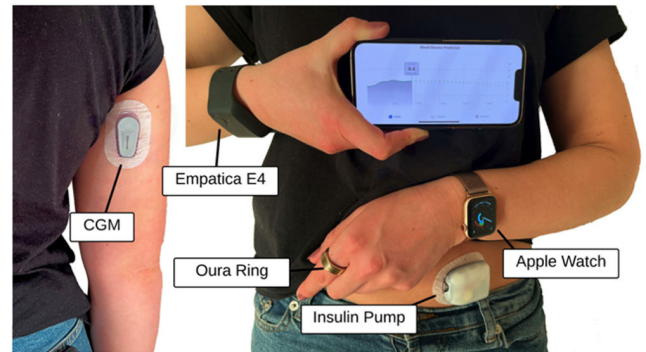


FIGURE 2. A CGM, an insulin pump and various non-invasive smart devices are connected to our SDK allowing real time BG prediction in a smartphone application.

Wireless communication unit: We use an iPhone as the wireless communication unit and for running mobile applications on top of the SDK.

Insulin management system: An IMS consists of an insulin pump and a CGM. We use a Dexcom G6 and an OmniPod. The hardware units are managed through the Loop app running on an iPhone [19]. The controller writes insulin deliveries, carbohydrate intakes, and CGM measurements to Apple Health. The CGM records measurements with a 5-minute sampling rate.

Empatica E4 wristband: Empatica E4 provides users with raw, continuous physiological sensor measurements. The communication protocol is Bluetooth low energy. A complete list of available raw sensor values can be found in the SDK documentation [20]. The sample rates are between 1 and 32 Hz.

Oura ring: Oura Ring is a smart ring focusing on sleep and activity tracking. Data can be accessed through a web API [21]. The user must open the Oura smartphone application to synchronize new sensor readings. Activities and sleep are sampled as events, and sensor measurements are available at 5-minute intervals or averaged during a sleep or activity event.

Apple Watch: The last hardware component is an Apple Watch, which automatically records workouts and biometric variables to Apple Health. Signals available in our SDK are heart rate (5-10-minute sample rate), heart rate variability (2-5-hour sample rate), energy expenditure (1-minute sample rate), and workout events.

2) SDK

Our SDK interfaces with the hardware components above using three distinct communication protocols: HealthKit, Oura Web API, and Bluetooth Low Energy (BLE), as depicted in Fig. 1. Through these protocols, we access real-time data from five hardware components, comprising both components of an IMS and non-invasive wearable devices.

- **HealthKit:** Data from Apple Health is inherently integrated into an iPhone. Users will be prompted with a consent form on initial use. Subsequently, users can access the data.
- **Oura API:** The SDK requires an API key during calls to access data from the Oura Ring.
- **Empatica E4:** The Empatica E4 connects directly to the SDK via BLE. Our SDK facilitates the establishment and verification of these connections.

It's worth noting that our SDK focuses on providing endpoints for various data types. While it does not include machine learning algorithms or BG prediction capabilities, it offers comprehensive data access and integration functionalities. The rationale behind excluding machine learning algorithms is that such algorithms can vary significantly, making data preprocessing a user-dependent task.

C. IMPLEMENTATION

1) PLATFORM SDK

The SDK is written in the programming language Swift and works for iOS applications. It provides access to several input values from wearable devices and IMS. Table 1 lists all the input values accessible from our SDK. We will describe how we solved the three main technical challenges to overcome during the development: Software architecture, interoperability of input values, concurrency, and BG prediction responsiveness.

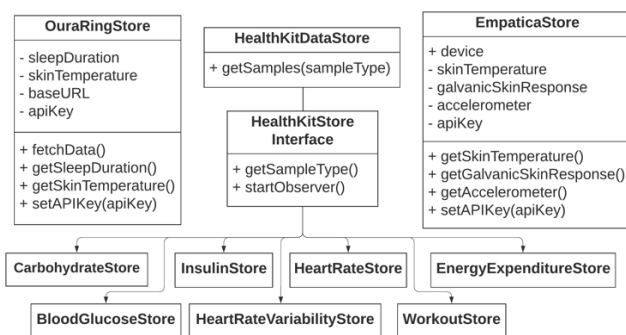


FIGURE 3. Class diagram for the proposed SDK.

2) SOFTWARE ARCHITECTURE

The proposed software architecture of the SDK aims to provide a unified interface for accessing real-time data from diverse sources, including the HealthKit framework, Empatica wristbands, and Oura Rings. Fig. 3 shows the class diagram for the proposed SDK. Specifically, access to HealthKit data is facilitated through singleton objects, which

possess a startObserver method that incorporates an update handler. This design enables applications to receive updates in real-time as data changes are detected. In contrast, the Empatica wristband and OuraRing objects provide access to all relevant data types through a single object. This design choice is consistent with the protocols to access data from these peripherals. However, this architecture only provides access to real-time data and does not incorporate any data persistence and storage mechanisms.

One key benefit of the proposed architecture is the utilization of a singleton object for HealthKit data access, which enables a centralized and efficient management of updates. Furthermore, using separate objects for Empatica wristband and OuraRing data access simplifies the development process by conforming to protocols for accessing data from these peripherals. The proposed SDK architecture is based on object-oriented programming principles and adopts a modular design, with distinct classes allocated to each data source. Additionally, the SDK class is designed to be adaptable, making it capable of accommodating new data sources and peripherals in the future. Finally, the observer pattern is utilized to deliver updates for the datatypes accessed through HealthKit, making the SDK highly responsive, which explains why real-time updates are possible.

3) INTEROPERABILITY

BG prediction algorithm inputs represent some state of an individual at a point in time. The platform SDK is connected to several sources of data inputs, which differ in sample intervals and might have some delay in their impact on BG levels. Three solutions for this problem are implemented depending on the input value:

- Aggregating values using the sum or average
- Using the most recent value
- Using physiological models to represent the delay and the relationship between the value and the BG response

Energy expenditure is an example of a value where data aggregation is used over a given time interval. In this study, sleep was recorded each morning, which is not as frequently as required for real-time prediction. Hence, our solution is to use the most recent sleep analysis value for predictions the whole day.

The effect of insulin injected subcutaneously has a slow onset and lasts for several hours. Physiological modeling of the insulin effect can improve BG prediction [22]. The amount of active insulin in the body at a given time is referred to as insulin on board (IOB). We used the insulin-on-board algorithm provided by the LoopKit library [23]. The SDK has been designed to allow developers the flexibility to select the parameters most appropriate for their specific application. This algorithm allows for user-defined insulin delay, peak activity time, and total activity time, as these vary across individuals and types of insulin. The same algorithm is used to model carbohydrate absorption. In our proof-of-concept software, the insulin model utilizes the default parameters specific to adults utilizing rapid-acting insulin provided by

the LoopKit library. The parameters utilized in the meal model were determined through trial and error, utilizing validation data to determine the optimal parameters. Table 2 lists the parameters used in our algorithms, and Fig. 4 plots the models.

TABLE 2. Parameters used in the physiological models for calculating insulin and carbohydrates on board.

	Insulin [min]	Carbohydrates [min]
Delay	10	0
Peak activity time	75	60
Total activity time	360	180

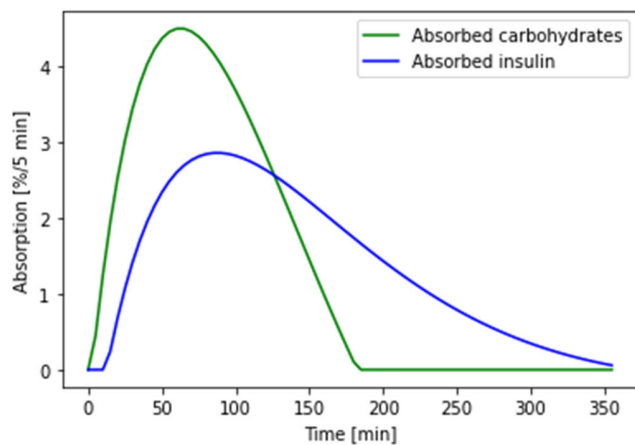


FIGURE 4. The physiological absorption models for insulin and carbohydrates.

4) CONCURRENCY

DispatchGroup is implemented in the SDK to synchronize the execution of concurrent calls and ensure that the calls for data from all the different peripherals have been completed before executing a BG prediction.

5) PREDICTION RESPONSE TIME

Two factors can impact BG prediction response times: The choice of prediction algorithm and the fetching of the input data values from wearable sensors or the IMS. For the SDK in this study, only the last factor is relevant, as users will choose their prediction approach. Three communication protocols are used, and different challenges are tied to them regarding response time optimization.

To observe changes in the Oura Ring web API data, we have used Swift to periodically request the API to retrieve and compare the latest data to the previous data. There is no preset time of day when the API updates, so we must always assume new data might be available. The response time of the retrieval from a web API is dependent on the wireless network speed. Empatica E4 is a Bluetooth Low Energy peripheral. The SDK is the host that sends continuous requests for data, and the wristband responds with the current measurement.

Hence, if the request interval of the host is lower than the sample interval of the wristbands, measurements will be lost. Both the Oura Ring and the Empatica E4 data are accessed with requests, and the request intervals can be adjusted if the prediction run time becomes a problem.

Data from the Apple Watch and the IMS is retrieved using HealthKit's HKAnchoredObjectQuery, which uses an "anchor" to track the position of the last item returned in the previous query. The anchor allows you to retrieve data that has been added to the HealthKit store since the last query efficiently. The SDK receives notifications whenever new data is added to the HealthKit store, allowing apps to update in real-time as new data becomes available, which leads to improved response time.

D. PROOF-OF-CONCEPT SMARTPHONE APPLICATION

The primary objective of developing a smartphone application using our proposed SDK is to showcase the SDK's capability to create real-time BG prediction applications. This demonstration involves the creation of a technical proof-of-concept, resulting in a smartphone app that leverages a Ridge regression model for these predictions. Our models are trained using Python and then exported to a compatible model format for iOS development using coremltools. Please note that our proposed SDK primarily focuses on data retrieval, and we expect users to develop and deploy their predictive models.

1) RIDGE REGRESSION MODEL

Users of our SDK can employ any BG prediction approach. In our proof-of-concept, we implement a Ridge Regressor within our smartphone application. This decision is anchored in the findings of Xie and Wang [17], who demonstrated that Ridge Regression outperforms other state-of-the-art algorithms in predicting BG levels. Their analysis spanned a range of predictive algorithms, from conventional machine learning models to advanced deep neural networks. We further justify the selection of Ridge Regression by its comparative simplicity, reduced computational demands, and greater interpretability relative to alternative algorithms.

We employ a direct prediction strategy to generate future glucose level sequences, opting against a recursive approach. Direct predictions can mitigate the propagation of errors [24]. Hence, we implement two models that estimate BG levels at 30 and 60 minutes into the future. The optimization of the alpha parameter within the Ridge Regression algorithm is conducted through a grid search bolstered by a 5-fold cross-validation process.

Guided by the optimal results presented in the benchmark study by Xie and Wang [17], we add 12 time-lagged features for each feature. Each model is trained on an individual basis.

We compare two models with different feature spaces. The first model has only CGM measurements as an input, while

the other model has CGM measurements, absorbed insulin, absorbed carbohydrates, and the energy expenditure.

2) DATA PROCESSING

Our study utilized two datasets to validate the Ridge Regression model. Initially, the Ohio T1DM dataset was employed to confirm that our implementation yielded comparable performance with the benchmarks established in [17]. Subsequently, this verified implementation was applied to train and test the model using data collected from our study participant.

To align our predictive performance with [17], we engaged the same OhioT1DM dataset from 2018, which comprises data from six individuals diagnosed with Type 1 Diabetes Mellitus (T1DM) [9]. The dataset spanned eight weeks and designated approximately 20% of its data for testing. Details on the division of test and train samples are delineated in [9]. We obtained this dataset under a Data Use Agreement (DUA no. D201804) between Ohio University and the Norwegian University of Science and Technology.

For our proof-of-concept application, data sourced from an author with T1D was extracted from the Apple Health app. This dataset, encompassing three months of monitoring, served to train the personalized prediction model. The train test split approach mirrors the protocol used with the Ohio T1DM dataset, where 20% of data was reserved for testing, culminating in 20685 training samples and 5211 test samples. In order to prevent data leakage between the train and the test samples, we omitted a segment of 12*24 data points (equivalent to one day) from the sequence.

The preprocessing steps were conducted using the procedures outlined in [17]. Initially, disparate features were resampled to a unified temporal resolution of 5-minute intervals. Subsequently, any missing values were imputed using the Akima method. Imputed target values were excluded from our evaluation. Finally, feature normalization was executed, scaling values to a [0,1] range.

3) PERFORMANCE METRICS

Following the best practices for BG prediction algorithm evaluation as outlined by Jacobs et al. [25], our analysis employs recommended metrics. In accordance with *Best Practice 33*, root mean squared error (RMSE) serves as our primary measure of accuracy. RMSE is defined in (1). Additionally, adhering to *Best Practice 34*, we include a measure of relative error by the mean relative error (MRE), defined in (2).

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_1^N (\hat{y}(k + \text{PH}) - y(k + \text{PH}))^2} \quad (1)$$

$$\text{MRE} = \frac{1}{N} \sum_1^N \frac{\hat{y}(k + \text{PH}) - y(k + \text{PH})}{y(k + \text{PH})} \quad (2)$$

Here, $\hat{y}(k + \text{PH})$ denotes the predicted BG level at a future time point k plus the prediction horizon (PH), and $y(k + \text{PH})$ represents the actual measured BG level at that same time point.

III. RESULTS

A. SOFTWARE DEVELOPMENT KIT

The core objective of this study was to engineer a versatile platform SDK capable of real-time data acquisition from an IMS and non-invasive wearable devices. The data acquired encompassed critical information related to physical activity, stress levels, and sleep patterns, intending to establish endpoints for input variables suitable for multivariable BG predictions. We addressed this challenge by utilizing an open-source IMS named Loop, which facilitated data transmission to Apple Health. Concurrently, several wearable devices were employed to achieve our objectives successfully. Ultimately, our SDK seamlessly provided all the requisite feature inputs identified and detailed in Table 1.

In summary, our SDK provides the following endpoints based on the communication protocols:

- **HealthKit:**
 - o CGM values
 - o Insulin deliveries
 - o Carbohydrate inputs
 - o Energy expenditure
 - o Heart rate
 - o Heart rate variability
 - o Workouts
- **Oura API:**
 - o Sleep duration
- **BLE:**
 - o Galvanic skin response
 - o Accelerometer
 - o Skin temperature

To assess the proficiency of our newly developed SDK, we chose to implement a proof-of-concept application that harnessed the capabilities of the SDK. Our primary goal was to determine the feasibility of constructing a smartphone application leveraging our SDK, capable of presenting real-time BG predictions without necessitating manual inputs. As illustrated in Fig. 2, our mobile system established connections with our wearable devices, thus enabling the provision of real-time mobile BG predictions. Furthermore, we incorporated a feature within the smartphone application that facilitates the export of input values and BG predictions to an Excel sheet, facilitating further in-depth analysis.

B. PREDICTION ALGORITHM PERFORMANCE

We benchmarked our Ridge Regression model against [17] using the OhioT1DM dataset, conducting individualized training for each subject and then computing the mean performance. Table 3 presents the RMSE results across different PHs and input features, comparing the OhioT1DM dataset with our data. Including additional features consistently yielded superior RMSE values compared to using only CGM inputs, both in the benchmark dataset and our participant's data.

Relative to the findings reported in Table 4 in [17], our model—incorporating CGM, insulin, carbohydrates, and heart rate with 12 time-lagged features over a 30-minute

TABLE 3. Accuracy in RMSE for models using a benchmark dataset compared to the data from our validation experiment, with respect to the input features and prediction horizons. All models use 12 time-lagged features for each input feature.

		RMSE (mg/dL)	
Dataset	Input features	PH = 30	PH = 30
Ohio T1DM	CGM	19,99 ($\pm 2,62$)	32,42 ($\pm 2,94$)
	CGM, insulin, carbohydrates, and heart rate	18,87 ($\pm 2,91$)	30,82 ($\pm 3,35$)
Validation experiment participant	CGM	18,32	27,61
	CGM, insulin, carbohydrates, and energy expenditure	17,01	26,28

TABLE 4. Accuracy in MRE for models using a benchmark dataset compared to the data from our validation experiment, with respect to the input features and prediction horizons. All models use 12 time-lagged features for each input feature.

		MRE (mg/dL)	
Dataset	Input features	PH = 30	PH = 60
Ohio T1DM	CGM	0,01 ($\pm 0,01$)	0,03 ($\pm 0,03$)
	CGM, insulin, carbohydrates, and heart rate	0,02 ($\pm 0,02$)	0,06 ($\pm 0,06$)
Validation experiment participant	CGM	0,02	0,05
	CGM, insulin, carbohydrates, and energy expenditure	0,02	0,05

PH—achieved a RMSE of 18.87 (± 2.91), compared to the benchmark’s 19.52 (± 2.90). This enhancement is likely attributable to our physiological modeling of insulin and carbohydrates. Table 4 illustrates the MRE results, indicating balanced predictions without significant skewness.

C. VALIDATION EXPERIMENT

To validate our SDK, one of the authors participated in a 24-hour experiment using the proof-of-concept smartphone application. The experiment encompassed various scenarios, resulting in the collection of a diverse set of data, including metrics related to absorbed carbohydrates, absorbed insulin, and calorie expenditure, as well as measured and predicted BG levels. The subject of the experiment was one of the authors, who has diabetes. The experiment was structured around three primary scenarios:

- **Carbohydrate intake:** The subject added carbohydrates to the IMS in conjunction with meals.
- **Insulin administration:** The subject administered insulin in response to meals or to correct elevated BG levels.
- **Physical activity:** Thirty minutes of high-intensity physical activity was included as a part of the experiment.

As the smartphone application was utilized throughout the experiment, real-time changes in BG predictions were observed in response to all the varying scenarios. As depicted in Fig. 5, the consumption of a carbohydrate-rich meal corresponded with an increase in predicted BG levels, while insulin administration decreased predicted BG levels.

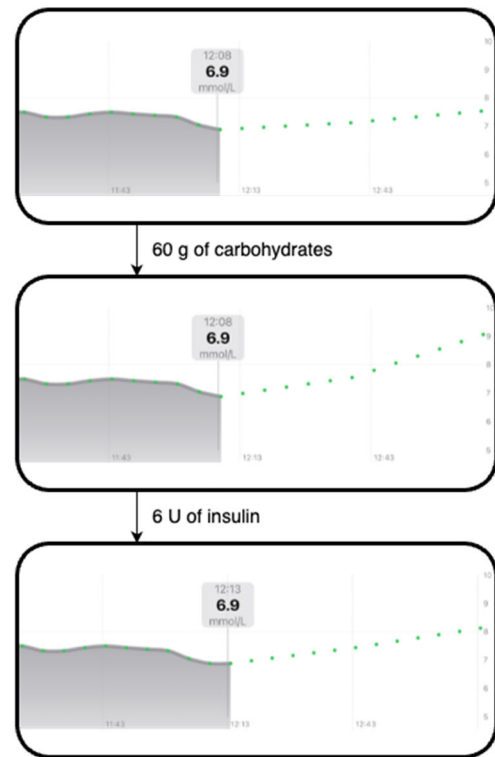


FIGURE 5. Screenshots from mobile predictions in different scenarios. Upper picture shows the initial BG prediction. The middle one shows the updated prediction after adding 60 g of carbohydrate to the insulin management system, and the lower picture is after 6 U of insulin is injected.

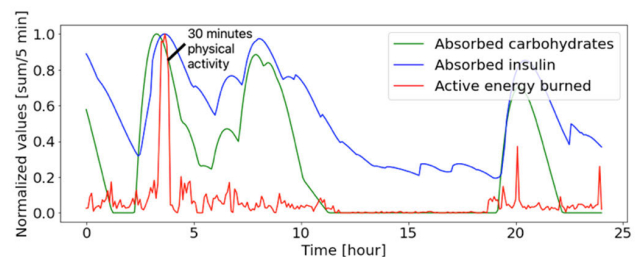


FIGURE 6. Absorbed carbohydrates, insulin and active energy burned.

During the 24-hour duration of the experiment, an array of data was tracked, including absorbed carbohydrates, absorbed insulin, and active calorie expenditure. These metrics were standardized and plotted in Fig. 6. The graphical representation allowed us to discern how the levels of absorbed carbohydrates and insulin evolved over the course of the day, along with fluctuations in calorie expenditure. Notably, there was a rise in calorie expenditure during

periods of high-intensity physical activity, accompanied by corresponding declines in BG levels.

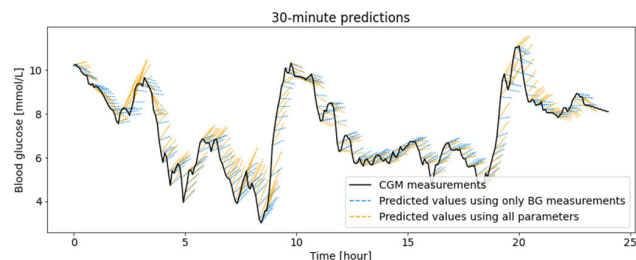


FIGURE 7. Comparison of predictions of a model trained on only BG measurements and another model using all the parameters plotted in Fig. 6.

In Fig. 7, we saw that the Ridge regression model using multiple parameters resulted in better predictions of glycemic responses to food, insulin, and physical activity compared to the model using only one parameter. However, these models were implemented to illustrate the potential benefits of providing access to more parameters in such prediction models, but the model approach still has room for improvement.

Overall, our experiment was a success, showing us that our SDK can provide data inputs in real-time without requiring manual inputs from the user. As we can see in Fig. 5, our app was able to make predictions in various situations, and the predictions were updated without requiring manual inputs. We observed that predictions increased after we registered carbohydrates in the IMS and decreased after we injected insulin and performed physical activity.

IV. DISCUSSION

One of the major challenges we encountered in implementing the platform SDK was the need to use multiple communication protocols, such as HealthKit, Bluetooth Low Energy, and a Web API, to connect to different wearable devices and the IMS. It required significant effort to ensure compatibility and interoperability but ultimately allowed us to collect a wide range of data types, as outlined in Table 1.

There has been a missing gap in research to evaluate multivariate BG predictions in real-life scenarios [12]. Some attempts have been made, but the solutions have required manual user inputs [13], [14]. In prior work, Kriventsov et al. created Diabits, an app that displayed BG predictions. They found a statistical connection between better glucose control and more frequent app use [13]. However, a drawback was that the predictions strongly relied on manual inputs from the users because most users failed to provide them. A significant challenge for developing a system that does not require manual inputs from the user is that insulin pump and CGM vendors usually have closed protocols so that researchers cannot access the data or the insulin delivery algorithms. Closed protocols are also the case for many wearable devices; they do not provide any raw data for users or researchers because they want to protect their algorithms. This might slow research in many health-related projects because it might

require research groups to develop their own devices or go with unpractical, non-mobile solutions, making it challenging to run long-term experiments.

The potential applications of the platform SDK are numerous and varied. For example, the platform could be used to develop decision support systems for people with diabetes, providing real-time BG predictions and personalized recommendations for insulin dosing and lifestyle changes. The platform could also be integrated with other wearable devices, such as smart scales and smart clothing, to provide a more comprehensive view of an individual's health and well-being. Additionally, the platform could be used to evaluate different multivariate BG prediction algorithms in real-life scenarios. Various predictive methodologies have been explored, including traditional machine learning techniques, advanced deep neural networks, and physiological modeling [22]. Transformer models represent a novel and growing area in glucose level prediction [26]. Our SDK provides a more realistic testbed for developing new and improved BG prediction models.

A limitation of this study is that the developed platform SDK depends on whether the IMS can write data to Apple Health in real-time. We solved this by using Loop, an open-source system without medical approval. Hence, doctors can not recommend that patients use this system. Also, it would be beneficial to know more about the insulin delivery algorithms of the IMS to create proper machine-learning models. For example, the interval between insulin deliveries for basal rates is unknown. Another consideration is the system's limitation regarding robustness, which falls short of the stringent clinical regulation requirements. Robustness pertains to the system's resilience in managing errors, which may arise from communication breakdowns between units or intermittent signal loss. Predictive adjustments to channel fluctuations can facilitate stable connectivity and data fidelity in remote sensor-smartphone communications [27]. However, the primary focus of this study was to establish data endpoints. We intended to develop a system capable of supporting real-life exploration of BG prediction algorithms, irrespective of the stringent regulatory demands. In this context, our system aligns with our intended objectives and scope.

V. CONCLUSION

Overall, our study provides proof of concept for a platform SDK that could be used to develop tools and decision support systems for people with diabetes or non-diabetics who want to optimize their health. This platform might contribute to a broader understanding of multivariate BG prediction algorithms by allowing for testing algorithms in real-life scenarios. Previous work has usually evaluated the algorithms on small, benchmark datasets from controlled environments. Currently, available simulators do not incorporate other multiple inputs, although this would generally be a good way to evaluate prediction algorithms. Further research is needed to explore the platform's potential applications in a broader range of BG prediction approaches, settings, and populations.

REFERENCES

- [1] H. Hall, D. Perelman, A. Breschi, P. Limcaoco, R. Kellogg, T. McLaughlin, and M. Snyder, "Glucotypes reveal new patterns of glucose dysregulation," *PLOS Biol.*, vol. 16, no. 7, Jul. 2018, Art. no. e2005143, doi: 10.1371/journal.pbio.2005143.
- [2] A. Z. Woldaregay, E. Årsand, S. Walderhaug, D. Albers, L. Mamykina, T. Botsis, and G. Hartvigsen, "Data-driven modeling and prediction of blood glucose dynamics: Machine learning applications in type 1 diabetes," *Artif. Intell. Med.*, vol. 98, pp. 109–134, Jul. 2019, doi: 10.1016/j.artmed.2019.07.007.
- [3] A. Cinar and K. Turksoy, *Advances in Artificial Pancreas Systems Adaptive and Multivariate Predictive Control*. Cham, Switzerland: Springer, 2018, pp. 1–7. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-319-72245-0_1
- [4] L. Herranz, L. Saez-de-Ibarra, N. Hillman, R. Gaspar, and L. F. Pallardo, "Cambios glucémicos durante el ciclo menstrual en mujeres con diabetes mellitus tipo 1," *Medicina Clínica*, vol. 146, no. 7, pp. 287–291, Apr. 2016, doi: 10.1016/j.medcli.2015.11.044.
- [5] S. Saganowski, P. Kazienko, M. Dziezyc, P. Jakimow, J. Komoszynska, W. Michalska, A. Dutkowiak, A. Polak, A. Dziadek, and M. Ujma, "Consumer wearables and affective computing for wellbeing support," in *Proc. 17th EAI Int. Conf. Mobile Ubiquitous Syst., Comput., Netw. Services*, Dec. 2020, pp. 482–487, doi: 10.1145/3448891.3450332.
- [6] A. D. Grant, M. Newman, and L. J. Kriegsfeld, "Uladian rhythms in heart rate variability and distal body temperature anticipate the luteinizing hormone surge onset," *Physiology*, vol. 2023, pp. 1–57, Jul. 2020, doi: 10.1101/2020.07.15.205450.
- [7] S. C. Mukhopadhyay, "Wearable sensors for human activity monitoring: A review," *IEEE Sensors J.*, vol. 15, no. 3, pp. 1321–1330, Mar. 2015, doi: 10.1109/JSEN.2014.2370945.
- [8] C. Hettiarachchi, E. Daskalaki, J. Desborough, C. J. Nolan, D. O'Neal, and H. Suominen, "Integrating multiple inputs into an artificial pancreas system: Narrative literature review," *JMIR Diabetes*, vol. 7, no. 1, Feb. 2022, Art. no. e28861, doi: 10.2196/28861.
- [9] C. Marling and R. Bunescu, "The OhioT1DM dataset for blood glucose level prediction: Update 2020," in *Proc. CEUR Workshop*, 2020, p. 5.
- [10] K. Turksoy, C. Monforti, M. Park, G. Griffith, L. Quinn, and A. Cinar, "Use of wearable sensors and biometric variables in an artificial pancreas system," *Sensors*, vol. 17, no. 3, p. 532, Mar. 2017, doi: 10.3390/s17030532.
- [11] H. Hameed and S. Kleinberg, "Comparing machine learning techniques for blood glucose forecasting using free-living and patient generated data," *Proc. Mach. Learn. Res.*, vol. 126, pp. 871–894, Aug. 2020.
- [12] O. Diouri, M. Cigler, M. Vettoretti, J. K. Mader, P. Choudhary, E. Renard, and H. Consortium, "Hypoglycaemia detection and prediction techniques: A systematic review on the latest developments," *Diabetes/Metabolism Res. Rev.*, vol. 37, no. 7, p. e3449, Oct. 2021, doi: 10.1002/dmrr.3449.
- [13] S. Kriventsov, A. Lindsey, and A. Hayeri, "The diabits app for smartphone-assisted predictive monitoring of glycemia in patients with diabetes: Retrospective observational study," *JMIR Diabetes*, vol. 5, no. 3, Sep. 2020, Art. no. e18660, doi: 10.2196/18660.
- [14] M. He, W. Gu, Y. Kong, L. Zhang, C. J. Spanos, and K. M. Mosalam, "CausalBG: Causal recurrent neural network for the blood glucose inference with IoT platform," *IEEE Internet Things J.*, vol. 7, no. 1, pp. 598–610, Jan. 2020, doi: 10.1109/JIOT.2019.2946693.
- [15] K. Li, J. Daniels, C. Liu, P. Herrero, and P. Georgiou, "Convolutional recurrent neural networks for glucose prediction," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 2, pp. 603–613, Feb. 2020, doi: 10.1109/JBHI.2019.2908488.
- [16] T. Zhu, L. Kuang, J. Daniels, P. Herrero, K. Li, and P. Georgiou, "IoMT-enabled real-time blood glucose prediction with deep learning and edge computing," *IEEE Internet Things J.*, vol. 10, no. 5, pp. 3706–3719, Mar. 2023, doi: 10.1109/JIOT.2022.3143375.
- [17] J. Xie and Q. Wang, "Benchmarking machine learning algorithms on blood glucose prediction for type 1 diabetes in comparison with classical time-series models," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 11, pp. 3101–3124, Nov. 2020, doi: 10.1109/TBME.2020.2975959.
- [18] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi, and G. Fichtinger, "PLUS: Open-source toolkit for ultrasound-guided intervention systems," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 10, pp. 2527–2537, Oct. 2014, doi: 10.1109/TBME.2014.2322864.
- [19] *LoopDocs*. Accessed: Oct. 20, 2022. [Online]. Available: <https://loopkit.github.io/loopdocs/>
- [20] *E4link iOS SDK API Reference*. Accessed: Dec. 4, 2022. [Online]. Available: <https://developer.empatica.com/ios-sdk-api-reference-100.html#didReceiveGSR:withTimestamp>
- [21] *Oura API*. Accessed: Dec. 4, 2022. [Online]. Available: <https://cloud.ouraring.com/docs/>
- [22] A. Bertachi, L. Biagi, I. Contreras, N. Luo, and J. Vehí, "Prediction of blood glucose levels and nocturnal hypoglycemia using physiological models and artificial neural networks," in *Proc. KHD@IJCAI*, 2018, p. 6.
- [23] (Nov. 27, 2022). *LoopKit*. Accessed: Dec. 5, 2022. [Online]. Available: <https://github.com/LoopKit/LoopKit>
- [24] S. Bengio, O. Vinyals, N. Jaitly, and N. Shazeer, "Scheduled sampling for sequence prediction with recurrent neural networks," 2015, *arXiv:1506.03099*.
- [25] P. G. Jacobs, P. Herrero, A. Facchinetti, J. Vehí, B. Kovatchev, M. Breton, A. Cinar, K. Nikita, F. Doyle, J. Bondia, T. Battelino, J. R. Castle, K. Zarkogianni, R. Narayan, and C. Mosquera-Lopez, "Artificial intelligence and machine learning for improving glyemic control in diabetes: Best practices, pitfalls and opportunities," *IEEE Rev. Biomed. Eng.*, pp. 1–19, Nov. 2023. [Online]. Available: <https://ieeexplore.ieee.org/document/10313965>, doi: 10.1109/RBME.2023.3331297.
- [26] R. A. Laursen, P. Alo, M. Goodwin, and P.-A. Andersen, "Distinct sequential models for inference boosting," in *Artificial Intelligence XL (Lecture Notes in Computer Science)*, M. Bramer and F. Stahl, Eds. Cham, Switzerland: Springer, 2023, pp. 198–203, doi: 10.1007/978-3-031-47994-6_15.
- [27] W. Xu, J. An, Y. Xu, C. Huang, L. Gan, and C. Yuen, "Time-varying channel prediction for RIS-assisted MU-MISO networks via deep learning," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 4, pp. 1802–1815, Dec. 2022, doi: 10.1109/TCCN.2022.3188153.
- [28] K. A. MacGregor, I. J. Gallagher, and C. N. Moran, "Relationship between insulin sensitivity and menstrual cycle is modified by BMI, fitness, and physical activity in NHANES," *J. Clin. Endocrinol. Metabolism*, vol. 106, no. 10, pp. 2979–2990, Sep. 2021, doi: 10.1210/clinem/dgab415.
- [29] S. A. Brown, B. Jiang, M. McElwee-Malloy, C. Wakeman, and M. D. Breton, "Fluctuations of hyperglycemia and insulin sensitivity are linked to menstrual cycle phases in women with T1D," *J. Diabetes Sci. Technol.*, vol. 9, no. 6, pp. 1192–1199, Oct. 2015, doi: 10.1177/1932296815608400.
- [30] J. D. Armstrong, "Heart rate as an indicator of activity, metabolic rate, food intake and digestion in pike, *Esox lucius*," *J. Fish Biol.*, vol. 29, pp. 207–221, Dec. 1986, doi: 10.1111/j.1095-8649.1986.tb05012.x.
- [31] S. D. Mamdiwar, Z. Shakruwala, U. Chadha, K. Srinivasan, and C.-Y. Chang, "Recent advances on IoT-assisted wearable sensor systems for healthcare monitoring," *Biosensors*, vol. 11, no. 10, p. 372, Oct. 2021, doi: 10.3390/bios11100372.
- [32] S. Gedam and S. Paul, "A review on mental stress detection using wearable sensors and machine learning techniques," *IEEE Access*, vol. 9, pp. 84045–84066, 2021, doi: 10.1109/ACCESS.2021.3085502.
- [33] S. Päßler, A. Noack, R. Poll, and W.-J. Fischer, "Validation of the use of heart rate variability measurements during meal intake in humans," in *Proc. Comput. Cardiol.*, Sep. 2013, pp. 999–1002.
- [34] O. Faust, W. Hong, H. W. Loh, S. Xu, R.-S. Tan, S. Chakraborty, P. D. Barua, F. Molinari, and U. R. Acharya, "Heart rate variability for medical decision support systems: A review," *Comput. Biol. Med.*, vol. 145, Jun. 2022, Art. no. 105407, doi: 10.1016/j.compbiomed.2022.105407.
- [35] S. Stankoski, M. Jordan, H. Gjoreski, and M. Luštrek, "Smartwatch-based eating detection: Data selection for machine learning from imbalanced data with imperfect labels," *Sensors*, vol. 21, no. 5, p. 1902, Mar. 2021, doi: 10.3390/s21051902.
- [36] B. Martinez-Tellez, L. Ortiz-Alvarez, G. Sanchez-Delgado, H. Xu, F. M. Acosta, E. Merchan-Ramirez, V. Muñoz-Hernandez, W. D. Martinez-Avila, M. A. Contreras-Gomez, A. Gil, I. Labayen, and J. R. Ruiz, "Skin temperature response to a liquid meal intake is different in men than in women," *Clin. Nutrition*, vol. 38, no. 3, pp. 1339–1347, Jun. 2019, doi: 10.1016/j.clnu.2018.05.026.
- [37] B. L. Smarr, K. Aschbacher, S. M. Fisher, A. Chowdhary, S. Dilchert, K. Pulton, A. Rao, F. M. Hecht, and A. E. Mason, "Feasibility of continuous fever monitoring using wearable devices," *Sci. Rep.*, vol. 10, no. 1, Dec. 2020, Art. no. 1, doi: 10.1038/s41598-020-78355-6.



MIRIAM K. WOLFF was born in Ålesund, Norway, in 1997. She received the M.Sc. degree in engineering and ICT from the Department of Mechanical and Industrial Engineering, Norwegian University of Science and Technology, Trondheim, in 2021, where she is currently pursuing the Ph.D. degree with the Department of ICT and Natural Sciences.

She was an Exchange Student with Valencia Polytechnic University, Valencia, Spain, from 2019 to 2020. During the master's degree, she had a summer internship with the Offshore Simulator Centre AS, in Summer 2018, and Holte Consulting AS, in 2019 and 2020. She was a Software Developer with Schibsted, in 2021. Her research interests include medical technology for enhanced glycemic control, sensor applications, rapid prototyping, human-centered design, and machine learning.



HANS GEORG SCHAATHUN was born in Bergen, Norway, in 1975. He received the Cand.Mag. degree in mathematics, economics, and informatics, in 1996, the Cand.Scient. degree in industrial and applied mathematics and informatics, in 1999, and the Dr.Scient. degree in informatics—coding theory from the University of Bergen, Norway, in 2002.

He also has Postgraduate Certificate of Academic Practice (PGCAP) from the University of Surrey, England, from 2006 to 2008. He was a Lecturer in coding and cryptography with the University of Bergen, in 2002, where he was a Postdoctoral Researcher, from 2003 to 2006. He was a Lecturer and a Senior Lecturer with the University of Surrey, England, from 2006 to 2010. Since 2011, he has been a Professor in computing with the Ålesund University College, which merged into NTNU—Norwegian University of Science and Technology, since 2016. He has been an Excellent Teaching Practitioner with NTNU, since 2018. He also teaches artificial intelligence and machine vision and does research within machine learning, philosophy of education, and design methodology. He has authored a book *Machine Learning in Image Steganalysis* (Wiley 2012).



ANDERS LYNGVI FOUGNER (Member, IEEE) was born in Namsos, Norway, in 1981. He received the M.Sc. and Ph.D. degrees in engineering cybernetics from the Department of Engineering Cybernetics, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, in 2008 and 2013, respectively.

His M.Sc. and Ph.D. degrees were in robust, coordinated, and proportional myoelectric control of upper-limb prostheses. Since 2014, he has been the Coordinator of the Artificial Pancreas Trondheim Research Group (APT), NTNU, where he has been an Associate Professor with the Department of Engineering Cybernetics, since 2018. His current research interests include artificial pancreas, closed-loop glucose control, glucose sensing, mathematical modeling, system identification, model-based control, pattern recognition, machine learning, sensor fusion, and acoustics.



MARTIN STEINERT (Member, IEEE) was born in Dresden, Germany. He received the B.A., M.A., and Ph.D. (Dr.rer.pol) degrees in technology management from the University of Fribourg, Switzerland.

He has been an Assistant Professor with the University of Fribourg, Switzerland, and a Visiting Scholar with MIT and Stanford University, before changing full time to the Center for Design Research (CDR), Stanford University, as the Deputy Director, where he was an Assistant Professor (Acting) in mechanical engineering with Stanford University. Since 2013, he has been a Full Professor in engineering design with the Department of Mechanical and Industrial Engineering (MTP), Norwegian University of Science and Technology (NTNU). His research interests include the fuzzy front end of new product development and design: optimizing the intersection of engineering design thinking and new product development, mechatronics/sensors, and computer sciences (especially machine learning). A special focus is on conceptual development and alpha prototype generation of high-performance requirements and on experimental tools and setups. In January 2023, he has more than 240 publications registered in Google Scholar. He has several prizes in both teaching and research. He has been a member of the Norwegian Academy of Technological Sciences (NTVA), since 2015.



RUNE VOLDEN was born in Oslo, Norway, in 1966. He received the Bachelor of Science degree in electrical engineering from the Institute of Science and Technology, The University of Manchester, England, and the Ph.D. degree in computer vision from the Department of Engineering Cybernetics, Norwegian University of Science and Technology, in 1995.

Currently, he is the Head of the Department of ICT and Natural Sciences, Norwegian University of Science and Technology. With 25 years of industrial research and development experience in manufacturing, maritime, and marine control systems, he has also served as the Chairperson of the Board for the Norwegian Association for Automatic Control.

...