From Wearable Data to Computational Notebooks: Towards an AI-assisted T1DM Data Analysis Pipeline

Anna Khristodulo, Giancarlo Mascetti, Giorgio Delzanno, Marta Bassi, Nicola Minuto

Università degli Studi di Genova
Istituto Giannina Gaslini
with the support of the benefit company ComfTech

ITADATA 2025

Motivations and Goals

- Managing Type 1 Diabetes Mellitus (T1DM) in pediatric patients presents significant challenges, particularly in monitoring and treatment.
- A major obstacle is the integration of data from diverse medical sources to form a holistic understanding of a patients' conditions
- The high volume and complexity of data generated by technologies like insulin pumps and wearable monitors can overwhelm healthcare providers
- Our research investigates whether it is feasible to design a personalized assistant that supports doctors in interpreting heterogeneous data streams from various wearable devices

Our case-study

- Data acquisition performed during a pediatric diabetes summer camp organized by the Istituto Giannina Gaslini[1]
- Participants were supervised by medical staff and wore two types of devices:
    - Tandem insulin pumps with continuous glucose monitor (CGM) communicating with the Glooko platform
    - Comftech's Howdy Senior sensorized garments which measure multiple physiological parameters such as heart rate variability, ECG, respiration rate, stress/movement index

- The goal was to monitor the physiological and metabolic responses of five pediatric patients with T1DM during daily physical activities.

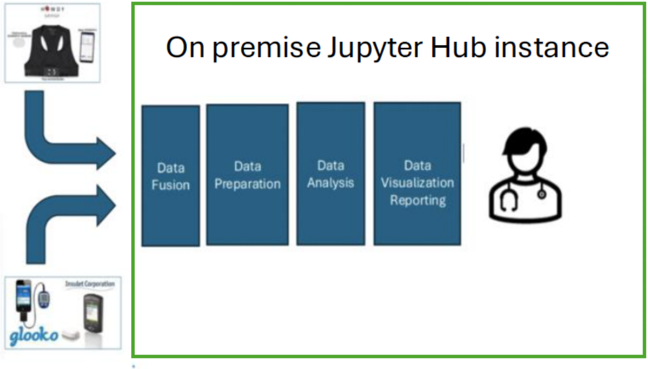
[1] A new data collection campaign with 15 patients started this week in Sarzana

Architecture

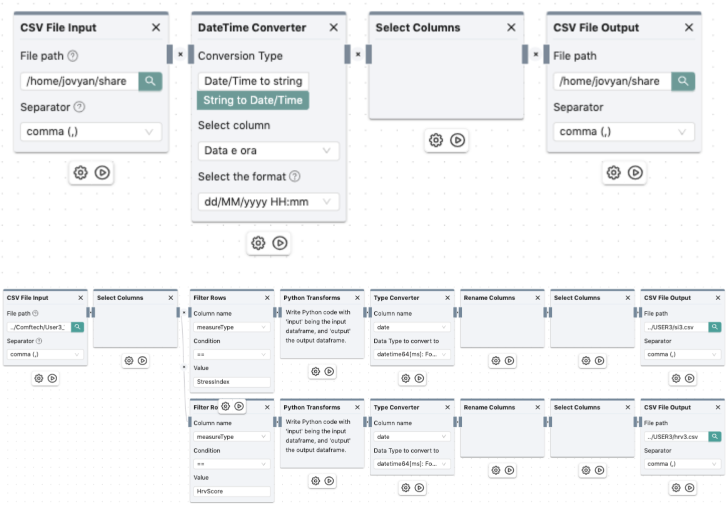
- We adopted Python-based Jupyter Hub framework as a central platform for acquiring, processing, and analyzing wearable data in a secure environment. This approach enables collaboration among researchers and clinicians while ensuring that sensitive health data remains within institutional boundaries.

- Our JupyterHub instance is deployed on a virtual machine hosted by the joint laboratory of the Gaslini Pediatric Hospital and our university.
- The system integrates data initially stored on proprietary cloud platforms managed by the device manufacturers.

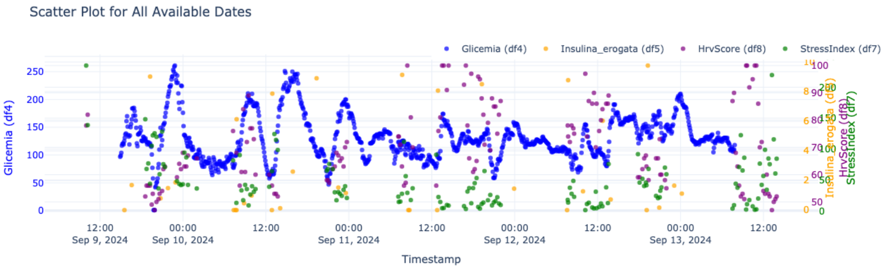
Architecture and Methodology



Data Preparation: Amphi Workflows

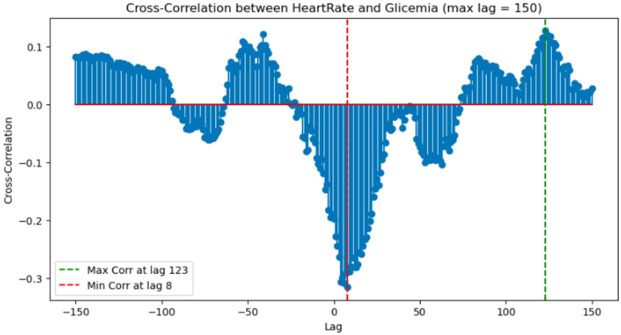


Data Fusion and Visualization



Explorative data analysis: Cross-correlation

- Among other analyses, we studied the cross-correlation between heart rate and glycemia over a range of time lags
- The goal was to identify time lags for highest positive and negative correlations to validate doctor's hypothesis on the effect of movement on glycemic level

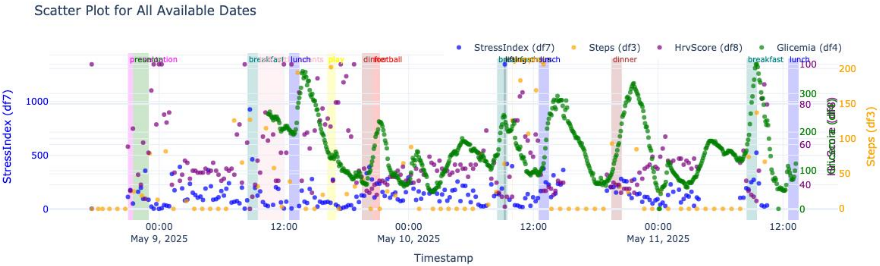



```
Maximum positive correlation at lag: 123
Maximum negative correlation at lag: 8
```

Explorative data analysis: Cross-correlation

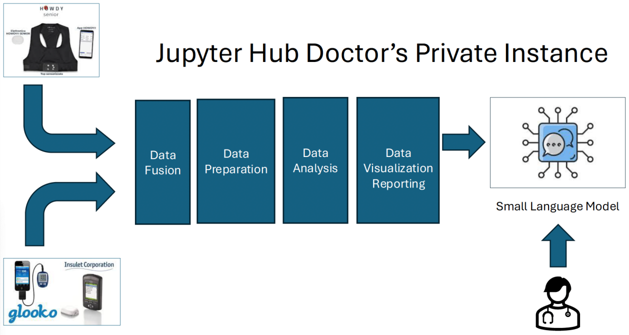
- The stress index demonstrates a strong negative correlation with glycemia (-0.72), suggesting that increased stress levels are associated with lower glycemia, potentially due to metabolic changes or increased energy expenditure under stress.
- Breathing frequency correlates positively with insulin delivery (0.59), possibly reflecting an association between metabolic demands and respiration. However, its correlation with stress is weak and negative (-0.17), suggesting that stress levels do not directly influence respiratory rate in a significant way.

- We are combining raw data with contextual information, e.g., categorization data related to physical activity



Towards an AI-assistant

- We experimented the use of local Large Language Models to support data interpretation as a possibile additional support for doctors, e.g., reports in natural language



Towards an AI-assistant

- We use the LangChain framework and Python libraries for extracting the knowledge based from patients' reports
- The knowledge base consisted on the provided biometric data.
- We created a ChromaDB vector database to generate a retriever to generate contexts associated to the questions
- During system prompting, additional contextual information was given, e.g., category for glycemic values
- We tested the resulting pipeline using a list of questions with ground truth provided by experts
- We also compared with AI services such as Google LM

Conclusions

- Exploratory study to design a pipeline for data fusion, processing and analysis in the context of a collaboration with the T1DM experts of the Gaslini Hospital supported by the benefit company ComfTech
- We are collecting more data, e.g. data collection campaign with 15 patients (this week!)
- We are considering data centric architectures such as Kafka and Spark to cope with large datasets
- We started a new collaboration with the Universitat de Girona to apply deep learning to predict the effect of movement analysis
- We need to improve the AI-assistant component extending the knowledge base and considering more complex queries posed by doctors