Machine Learning Approaches to Identify Volatile Organic Compounds

Daniel Kerber

Supervisor(s): Prof. Dr. Lilian Witthauer, Prof. Dr. José Garcia-Tirado
Institution(s): University of Bern and University Hospital Bern (Inselspital),
Department of Diabetes, Endocrinology, Nutritional Medicine and Metabolism (UDEM)
Examiners: Prof. Dr. Lilian Witthauer, Prof. Dr. José García-Tirado

Introduction

Diabetes mellitus, a chronic metabolic disorder, poses significant global health challenges. Managing hypoglycemia, especially in patients with Type 1 Diabetes (PwT1D), is a complex task with current monitoring technologies. However, the potential of non-invasive glucose monitoring through the analysis of volatile organic compounds (VOCs) in exhaled breath offers a promising alternative. This study, which investigates the use of a Gas Chromatography -Ion Mobility Spectroscopy (GC-IMS) dataset from PwT1D patients, introduces a novel Convolutional Neural Network (CNN) approach. This approach is a novel development aimed at improving traditional methods, bypassing part of the extensive preprocessing, and effectively identifying VOC patterns for distinguishing glycemic states, with significant implications for the future of diabetes management.

Materials and Methods

A dataset of 330 GC-IMS breath samples was collected from ten PwT1D patients during induced hypoglycemia. The study employed a preprocessing pipeline that included binning, alignment, baseline correction, and scaling to facilitate effective non-targeted analysis (NTA). This pipeline was used to compare the traditional Partial Least Squares Discriminant Analysis (PLS-DA) method, utilizing Variable Importance in Projection (VIP), with a Convolutional Neural Network (CNN) integrated with Gradient-weighted Class Activation Mapping (Grad-CAM). These methods were evaluated for their ability to perform feature selection and classify samples into fasting, hypoglycemic, and postprandial states.

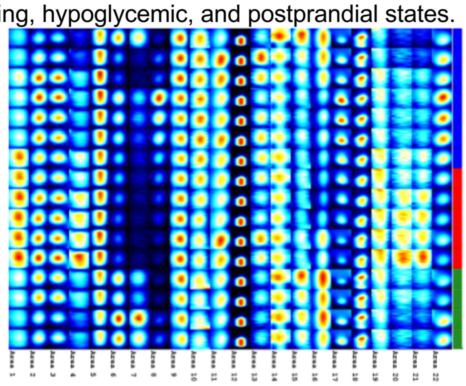

Fig. 1 Variation of VOC peaks across a patient visit. Rows represent a whole spectrum, and columns display the same VOC peak across all spectra in the visit. The rightmost column shows the experimental phases: blue for fasting, red for hypoglycemia, and green for postprandial.

Results

PLS-DA and CNN identified similar key VOCs, including acetone, isoprene, and the monomeric and dimeric forms of 2-propanol and ethanol, as significant markers for differentiating between glycemic states, all of which have been investigated in the literature. While PLS-DA showed slightly better overall accuracy in the preprocessed pipeline, CNN with Grad-CAM demonstrated the best accuracy, sensitivity, and specificity when removing scaling and baseline correction steps. The CNN approach effectively captured relevant VOC patterns used in the feature selection process for validation.

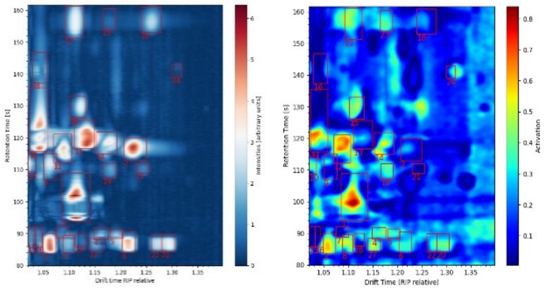

Fig. 2 Comparison of Mean VIP plot (left) and overall Grad-CAM activation plot (right) highlighting key VOCs across the spectra for distinguishing the glycemic states.

Discussion

The novel CNN with Grad-CAM approach offers a promising new technique for NTA of breath VOC spectra in non-invasive diabetes monitoring. It showed comparable performance to traditional methods while requiring fewer preprocessing steps that are time-consuming and subjective to implement. Future research should explore the generalizability of this approach across different datasets and refine the CNN model to enhance its applicability in clinical settings

References

Nicolier C, Künzler J, Lizoain A, Kerber D, Hossmann S, Rothenbühler M, Laimer M, Witthauer L. “Detection of hypoglycemia in Type 1 Diabetes through breath VOC profiling using Gas Chromatography-Ion Mobility Spectrometry.” Diabetes Technology Therapeutics. 2024.

Acknowledgments

I sincerely thank my supervisor, Prof. Dr. Lilian Witthauer, for the opportunity and guidance. I also thank Cléo, Camilo, and the Samlab team for their valuable support.