Predicting REIMS from MALDI for glioblastoma sections using deep learning

Thomas Vanhemel, Melanie Nijs, Angeliki Birmpili, Tim Hendriks, Eva Cuypers, Bart De Moor
KU Leuven, Department of Electrical Engineering (ESAT-STADIUS)

Introduction

Rapid evaporative ionization mass spectrometry (REIMS) is an analytical technique that determines the **metabolic composition** of evaporating tissue during electrosurgical resection in real-time. The destructive nature of this method, i.e., burning the tissue, complicates further analysis of a tissue sample [1].

The **matrix-assisted laser desorption ionization mass spectrometry imaging (MALDI-MSI)** technique is employed to gauge the **quantity and spatial distribution** of molecules within the tissue sample. **Both techniques measure mass spectra** at specific spatial locations, albeit with different spatial resolutions and different ionization sources, but identical Quadrupole Time-of-Flight detectors.

To circumvent the **destructive nature of REIMS**, a deep learning-based approach is developed to predict REIMS data based on the results obtained from a MALDI-MSI experiment on the tissue sample.

Challenges

This work focuses on the prediction of REIMS spectra from MALDI-MSI lipidomics data of human glioblastoma stage IV samples. Due to the differences between both modalities, a robust method has to overcome following challenges:

- Preprocessing REIMS: detecting and retrieving measurement from time signal, removing background, baseline correction, limit to desired m/z range
- Preprocessing MALDI-MSI: baseline correction, interpolation, limit to desired m/z range
- Matrix effects, not identical across MALDI-MSI sample
- Detecting from which MALDI-MSI pixels the REIMS measurement was taken
- Large amount of information in a single sample
- MALDI-MSI takes information from a slim tissue slice, whereas during a REIMS experiment, information aggregates from deeper within the tissue

Method & Results

Detection of pixels
Use similarity measure to detect most similar Gaussian convoluted MALDI-MSI mass spectrum to the REIMS mass spectrum in a specific m/z value range

Model

- Reduce computational cost by performing dimensionality reduction
- Train a convolutional autoencoder on 48×48 patches (estimated size of surface of iKnife) of MALDI-MSI pixels to learn rich latent representation
- Predict REIMS mass spectrum from learned latent space, as shown in Figure 1, to improve performance with scarce amount of data [2]
- Validate model on unseen data

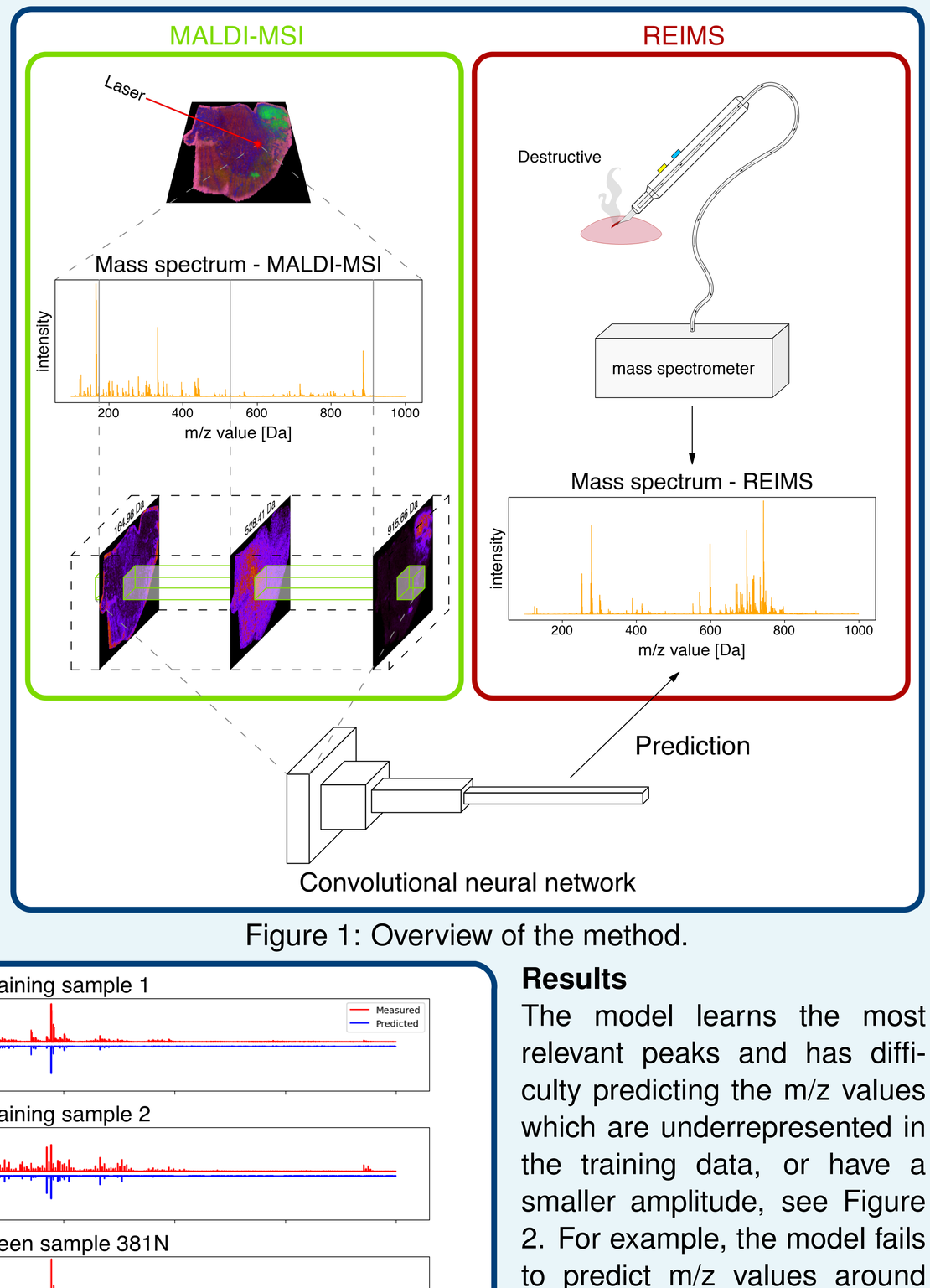

Figure 1: Overview of the method.

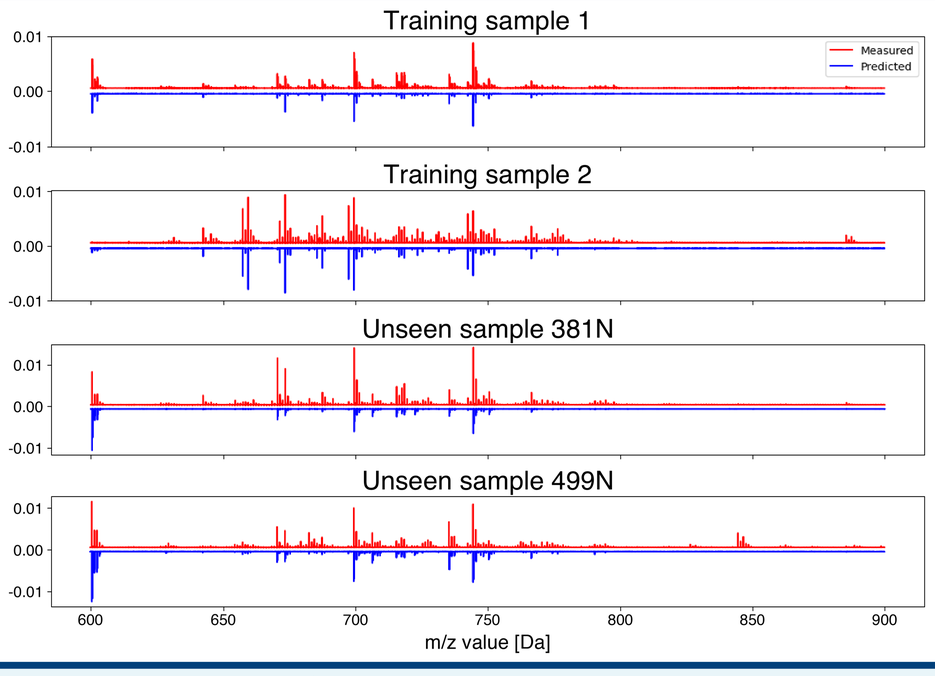

Figure 2: Predictions of the model on training and unseen samples.

Results
The model learns the most relevant peaks and has difficulty predicting the m/z values which are underrepresented in the training data, or have a smaller amplitude, see Figure 2. For example, the model fails to predict m/z values around 850 Da. The targets are TIC normalized, but the predictions are not. As a consequence, absolute intensities are not always predicted correctly, whereas relative intensities are.

Next steps

Preprocessing

- Use pictures of tissue samples to more accurately detect location of REIMS measurements
- Improve detection of pixels by considering m/z ranges with a minimal amount of matrix effects
- Removing matrix effects from MALDI-MSI measurements

Data

- Including more diverse tissue samples
- Including more REIMS measurements
- Data augmentation, i.e., flipping or rotating the patches

Model selection

- Further tuning of hyperparameters
- Experimenting with different architectures
- Incorporate TIC normalization in the cost function

Explainability

- Include SHAP values [3] in analysis to understand significance of specific m/z values and corroborate findings with experts

Conclusion

The proposed convolutional neural network is **able to capture** the majority of **important features** by leveraging the powerful abstract representations of the trained autoencoder to infer the REIMS measurement data from the MALDI-MSI results. However, the model fails to predict several m/z values and the results are not TIC normalized.

The model enables us to acquire REIMS measurement data without damaging the tissue, preserving them for further experimentation.

Once satisfactory results are obtained, they are fed to an existing REIMS model [4] to improve the model for in surgery, real-time tissue-type predictions during brain surgery.

REFERENCES

[1] E. R. St John et al., “Rapid evaporative ionisation mass spectrometry of electrosurgical vapours for the identification of breast pathology: towards an intelligent knife for breast cancer surgery,” Breast Cancer Research, vol. 19, no. 1, p. 59, May 2017, doi: 10.1186/s13058-017-0845-2.

[2] D. Soekhoe, P. van der Putten, and A. Plaat, “On the Impact of Data Set Size in Transfer Learning Using Deep Neural Networks,” in Advances in Intelligent Data Analysis XV, H. Boström, A. Knobbe, C. Soares, and P. Papapetrou, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016, pp. 50–60. doi: 10.1007/978-3-319-46349-0_5.

[3] L. S. Shapley, “17. A Value for n-Person Games,” in Contributions to the Theory of Games (AM-28), Volume II, H. W. Kuhn and A. W. Tucker, Eds., Princeton University Press, 1953, pp. 307–318. doi: 10.1515/9781400881970-018.

[4] Van Hese L, De Vleeschouwer S, Theys T, Larivière E, Solie L, Sciot R, Siegel TP, Rex S, Heeren RMA, Cuypers E. "Towards real-time intraoperative tissue interrogation for REIMS-guided glioma surgery," J Mass Spectrom Adv Clin Lab. 2022 Apr 29;24:80-89. doi: 10.1016/j.jmsacl.2022.04.004. PMID: 35572786; PMCID: PMC9095887.

ACKNOWLEDGMENTS

This work was supported in part by the KU Leuven: Research Fund (iBOF/23/064, C3/20/117, C3I-21-00316), Industrial Research Fund and several Leuven Research and Development bilateral industrial projects; in part by Flemish Government agencies: FWO (S005319N, T001919N, 11K5623N), EWI and VLAIO (HBC.2019.2204, HBC.2021.0076); in part by the European Commission (ERC Adv. Grant under grant 885682); and in part by the NWO STEM (Project Number 19013 to E.C.) which is financed by the Netherlands Organization for Scientific Research. This research is part of the LINK program which is financially supported by the Dutch Province of Limburg. The authors are thankful to the Department of Neurosurgery of KU Leuven for providing the GBM samples. The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government.