



## POSTER 29

### **A one-dimensional convolutional neural network model for classifying breast cancer subtypes and evaluating biochemical content using micro-FTIR hyperspectral images**

Matheus del Valle<sup>1</sup>, Denise Maria Zezell<sup>1</sup>

<sup>1</sup> Center for Lasers and Applications, Nuclear and Energy Research Institute IPEN-CNEN, Sao Paulo-SP- Brazil

#### **Introduction:**

Breast cancer, comprising 11.7% of new cancer cases and causing 6.9% of cancer-related deaths in 2020, presents a formidable global health challenge. Despite advancements in classification, treatment remains complex. Molecular subtyping—luminal A, luminal B, HER2, and TNBC—guides therapy selection based on biomarker expression. However, histology and immunohistochemistry, the gold standards for classification, suffer from inconsistencies and interobserver variability, complicating accurate diagnosis. Fourier Transform Infrared (FTIR) spectroscopy offers promising biochemical insights, including lipid and protein content, enhancing assessment quality. Machine learning, applied to micro-FTIR hyperspectral images, aids in predicting crucial parameters, yet interpretability, especially with deep learning, remains challenging. This study pioneers a one-dimensional deep learning approach to detect breast cancer, classify subtypes, and explore biochemical influences. Such automated insights have the potential to improve pathology reports and deepen our understanding of breast cancer biology.

#### **Methods:**

The BR804b breast cancer microarray, ordered from Biomax, Inc., USA, comprised formalin-fixed and paraffin-embedded (FFPE) histological sections on calcium fluoride (CaF<sub>2</sub>) slides from Crystran, UK. Sixty cores of 1.5 mm each were imaged, with one Cancer and one Adjacent Tissue (AT) core per patient. Molecular subtypes, based on St. Gallen International Expert Consensus guidelines, were classified as Luminal A (LA), Luminal B (LB), HER2, and Triple-negative Breast Cancer (TNBC). Image acquisition utilized a Cary Series 600 system (Agilent Technologies, USA), generating 320x320 pixel hyperspectral images with 5.5 μm spatial resolution for each core, resulting in 6,144,000 raw spectra. Data preprocessing included tissue and paraffin region selection using two-step K-means clustering, biofingerprint truncation, outlier removal, and Extended Multiplicative Signal Correction (EMSC). A novel 1D convolutional neural network (CNN) named CaReNet-V1 was developed for classification, with performance evaluation using accuracy, specificity, sensitivity metrics, and Gradient-weighted Class Activation Mapping (Grad-CAM). All analyses were conducted using Python, primarily Tensorflow and Keras libraries, on a GeForce GTX 1080 GPU

#### **Result:**

Type classification employed a single output (a sigmoid-activated neuron), yielding one Grad-CAM for CA class activation. Conversely, subtype classification, with four output probabilities, generated four Grad-CAMs. These align with vibrational spectroscopy's molecular footprints, aiding composition impact understanding. Main regions contributing to cancer type activation included adenine vibrations and Amide I, Amide II, and Amide III bands. Distinctive spectral regions contributed to LA, LB, HER2, and TNBC classifications,



reflecting biochemical composition variations. 1D deep learning with Grad-CAM offers label-free insights into breast cancer composition, aiding diagnosis, treatment assessment, and therapeutic development. Models like CaReNet-V1, with fewer parameters and efficient memory usage, show promise for classification tasks. Challenges like heterogeneous tissue representation and dataset augmentation exist but optimizing spectral analysis techniques and model architectures can improve accuracy and clinical applicability. Further studies should expand datasets and evaluate individual receptor expressions to enhance model performance.

#### **Discussion:**

Two K-Means clustering identified tissue and paraffin spectra, streamlining data processing. Employing a 1D approach expanded the dataset significantly. CaReNet-V1 effectively classified breast cancer tissue with minimal errors.

Acknowledgement: This work was supported by CNPq (INCT-INTERAS 406761/2022-1), INCT-INFO (465763/2014-6); Sisfoton (440228/2021-2); PQ (314517/2021-9); CAPES Finance code 001 and FAPESP (21/00633-0).