14th Conference on

DATA ANALYSIS METHODS
for Software Systems

November 30 — December 2, 2023

Druskininkai, Lithuania, Hotel "Europa Royale"
https://www.mii.lt/DAMSS

Co-Chairmen:

Prof. Gintautas Dzemyda (Vilnius University, Lithuanian Academy of Sciences)
Dr. Saulius Maskeliūnas (Lithuanian Computer Society)

Programme Committee:

Dr. Jolita Bernatavičienė (Lithuania)
Prof. Juris Borzovs (Latvia)
Prof. Robertas Damaševičius (Lithuania)
Prof. Janis Grundspenkis (Latvia)
Prof. Janusz Kacprzyk (Poland)
Prof. Ignacy Kaliszewski (Poland)
Prof. Bożena Kostek (Poland)
Prof. Tomas Krilavičius (Lithuania)
Prof. Olga Kurasova (Lithuania)
Assoc. Prof. Tatiana Tchemisova (Portugal)
Prof. Julius Žilinskas (Lithuania)

Organizing Committee:

Dr. Jolita Bernatavičienė
Prof. Olga Kurasova
Assoc. Prof. Viktor Medvedev
Laima Paliulionienė
Assoc. Prof. Martynas Sabaliauskas
Prof. Povilas Treigys

Contacts:

Dr. Jolita Bernatavičienė
jolita.bernataviciene@mif.vu.lt
Prof. Olga Kurasova
olga.kurasova@mif.vu.lt
Tel. +370 5 2109315

Detection of Pancreatic Cancer on CT Images Using Pseudo-Labelling Methods

Aušra Šubonienė[1], Olga Kurasova[1], Gintautas Dzemyda[1],
Viktor Medvedev[1], Aistė Gulla[2], Artūras Samuilis[3],
Džiugas Jagminas[3], Kęstutis Strupas[2]

[1] Institute of Data Science and Digital Technologies
Vilnius University
[2] Institute of Clinical Medicine, Faculty of Medicine
Vilnius University
[3] Institute of Biomedical Sciences, Faculty of Medicine
Vilnius University

ausra.suboniene@mif.vu.lt

This study addresses the issue of pancreatic cancer detection through the classification of computed tomography (CT) images using a semi-supervised deep learning system. Annotating medical computed tomography images is a resource-intensive process, as it requires medical experts to label each image individually within a CT scan sequence. Given the scarcity and challenges associated with acquiring labelled data, coupled with the high demand for extensive labelled datasets by contemporary machine learning techniques, the generation of pseudo-labels for unlabeled data serves to augment the volume of accessible data for machine learning applications, thereby potentially enhancing their overall performance. The aim of this research is to explore various pseudo-labelling methods to expand the labelled dataset. The study employs probability-based, entropy-based, and proximity-based pseudo-labelling methods to generate additional pseudo-labels within the unlabelled dataset. To assess the impact of pseudo-labelling on classification accuracy, both public and Vilnius University Hospital Santaros klinikos CT image datasets are utilised in this research. The convolutional neural network was employed to assess the influence of pseudo-labels on pancreatic cancer detection outcomes. The utilisation of diverse data not only enhances the reliability of the results but also allows for an evaluation of the quality of pseudo-labelling in different patient groups. By combining publicly available and private datasets, this study seeks to provide a comprehensive evaluation of the effectiveness criteria for pseudo-labelling.