b
UNIVERSITÄT
BERN

Master Artificial Intelligence in Medicine

Annual Report 2024

INSELGRUPPE

Introduction

The MSc in Artificial Intelligence in Medicine (AIM) at University of Bern is a unique, end-to-end program that provides an excellent and innovative educational scheme to train experts in the domain.

The program covers topics from artificial intelligence, machine learning, and deep learning to their practical applications in the medical field. In addition to AI focused courses, the curriculum includes courses related to medical and clinical practice, providing students with valuable insights into the daily operations of the clinical world and guiding them in the effective integration of AI into the medical domain. Students acquire a solid background in AI and knowledge of the most advanced techniques in the field. They apply their skills to design, develop and evaluate intelligent systems across the medical and clinical world.

By working closely with physicians from Bern University Hospital (Inselspital) and internationally acknowledged researchers in AI, students experience the benefits of AI and learn to recognise the extent to which AI offers intelligent answers to open questions and are able to identify areas in clinical practice that have research and development potential. Graduates are trained to become a new generation of experts in the field of AI, capable of driving entrepreneurial, technical and practical innovation for tomorrow’s healthcare.

With the master's thesis, students expand their knowledge in a selected field, preparing themselves for further academic programs or a career in industry.

Program Management

Prof. Dr. Raphael Sznitman
Director CAIM, Co-Responsible Research Project Fund

Prof. Dr. Stavroula Mougiakakou
Co-responsible Digitalisation & AI Education

Study Coordination

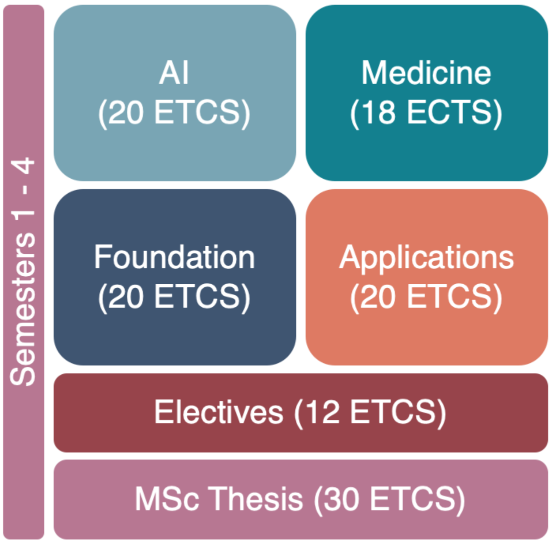
Julia Spyra, M.A.
Program Coordinator, Event Organisation

Program

The MSc Artificial Intelligence in Medicine (AIM) is a two-year, full-time master's program in English. The master program consists of 120 ECTS credits, where one ECTS corresponds to approximately 25-30 hours of work.

Students who work alongside their master's may officially request to extend their studies to six semesters.

Students are awarded 88 ECTS credits for completing mandatory courses, 12 ECTS credits for completing elective courses and 30 ECTS credits for their master's thesis.The outline of the study structure can be seen below.



Module AI

Students gain deep knowledge in AI and machine learning through the madatory courses:
- Introduction to Artificial Intelligence
- Machine Learning
- Deep Learning
- Reinforcement Learning

Module Medicine

The module provides students with an introduction to the clinical environment. In addition, they spend time exploring fundamental aspects of physiology and the crucial topic of omics. The “Clinical Implementations in AI” course allows students to experience different medical specialties through a system of rotations in different hospital departments.

Module Applications

Several mandatory courses will allow students to explore and understand the existing applications of AI in domains like medical image processing, computer vision, natural language processing and medical decision support systems.

Module Foundation / Electives

Students can select from a wide range of courses in computer science, mathematics, biomedical engineering, bioinformatics, and computational biology. This choice enables them to
- bridge the gap between their bachelor's and MSc AIM studies by tailoring their course selection to their individual scientific background (Foundation).
- strengthen their skills in specific areas by choosing courses according to their personal interests (Electives).

Master Thesis

The MSc thesis project will be conducted in collaboration with renowned physicians from Bern University Hospital (Inselspital), internationally acknowledged researchers in the field of AI from the University of Bern and leading Swiss and international companies.

New Course: Trustworthy AI in Medicine

Machine learning in medicine has the potential to transform patient care in coming years. In medical image analysis, ML algorithms already reach excellent performance in many tasks, sometimes even surpassing expert human performance. Despite these advances, using ML in patient care holds a potential for harm: when an algorithm makes mistakes this can lead to wrong treatment decisions. ML systems in healthcare must therefore be trustworthy and compliant with regulatory requirements. The course covers: regulatory context, explainability, reliability, robustness, fairness, privacy and clinical validation.

Student testimonials

Isabella, MSc AIM student (second year)

> "The opportunity to learn how these disciplines are applied to real-world roblems is a great advantage of the program, thanks to courses on topics such as computer vision, natural language processing, medical time series data, computer-assisted surgery, and medical robotics."

Could you describe your academic and professional background before starting your AIM Master's studies?
I have a bachelor's in biomedical engineering and a specialization in artificial intelligence. I have also worked for a year and a half as a machine learning engineer at a software company.

What has been your experience studying AIM at the University of Bern?
My experience at the University of Bern has been magnificent. The program is very innovative and interesting, designed to prepare professionals for the future, and that is precisely what I have perceived during my studies. The structure of the program offers an appropriate combination of topics related to medicine and computer science, as well as artificial intelligence. Both disciplines are taught through in-depth and detailed courses. Additionally, the opportunity to learn how these disciplines are applied to real-world problems is a great advantage of the program, thanks to courses on topics such as computer vision, natural language processing, medical time series data, computer-assisted surgery, and medical robotics, among others.

I can affirm that I have acquired a great amount of new knowledge and that my expectations have been met so far. Moreover, I have not only acquired intellectual knowledge but also learned about Swiss culture, its language, and its customs. The city of Bern is a multicultural place that allows you to develop as a well-rounded person, enjoy the beautiful nature, and savor delicious cafés and restaurants.

If you worked part-time during your studies, how was that experience?
I had the great opportunity to get a part-time job in the study coordination of the AIM program, which has been a great support during my studies and has allowed me to acquire new knowledge and build new relationships. I have also been able to contribute to the improvement of our program's website and the way information is communicated, using different media such as social networks and creating engaging videos that attract more people to join the program.

Alicja, MSc AIM student (second year)

> "... imagine studying Computer Vision theory in one class, then learning about its applications in Computer-Assisted Surgery, and later being invited to observe neurosurgeons in the operating room, utilizing these technologies during brain surgery!"

I completed my bachelor's degree in Computer Science in Poland. After working for two years as a Data Scientist across various projects, I realized that the medical applications of data science were the most fascinating to me. That's why I decided to focus on this field by studying AI in Medicine.

The program is a brilliant blend of computer science and biomedical engineering, enriched with specialized courses tailored to our discipline.

The computer science courses provide a strong foundation in AI, with topics like Machine Learning, Computer Vision, and Deep Learning. These courses not only focus on theory but also dive into state-of-the-art solutions, ensuring we are prepared for the technical demands of the field. I've been truly impressed with the quality of teaching and how well the material has been delivered.

The biomedical engineering courses, on the other hand, offer insights into human biology and cutting-edge medical technologies used for diagnosis, treatment, and therapy. This knowledge is crucial for understanding the medical context where AI solutions are applied.

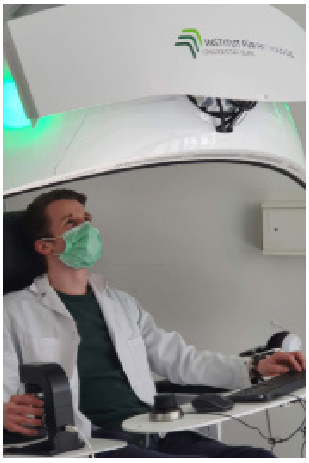
One of the program's standout features is the set of courses created specifically for AI in Medicine. These classes offer an incredible opportunity to interact with clinicians and researchers - people who will likely be our future colleagues or clients. As students, we're also invited to Inselspital, where we can observe the daily workflows of doctors, gaining first-hand experience with the challenges and limitations that will shape the AI solutions we will develop in the future.

For instance, imagine studying Computer Vision theory in one class, then learning about its applications in Computer-Assisted Surgery, and later being invited to observe neurosurgeons in the operating room, utilizing these technologies during brain surgery! This hands-on experience bridges the gap between theory and practice in a way that no other program could.

I truly believe this program offers exceptional preparation for a career in medical AI, and I highly recommend it to anyone passionate about applying AI to solve real-world problems in healthcare.

Clinical Visit Testimonials

Vinzenz, MSc AIM student (second year)



During my first semester in the master's program Artificial Intelligence in Medicine, I attended the course "Introduction to Clinics". Each week, we were introduced to another department of the Bern University Hospital (Inselspital). Pathology was my favorite department! So I had kind of high expectations when registering for the clinical visits that are offered in the course "Clinical Implementations of AI".

These expectations were not only completely fulfilled, but even exceeded. I was surprised how much time all the different people at the institute took to show me their work and answer my long list of questions. I particularly liked the structure of the daily schedule: processes of clinical pathology, how slides get digitalized, all the different deep learning projects with digital data. I also got to produce a digital data sample of a liver tissue. Actually, I can very much picture carrying out a master's thesis project in the AI in pathology research group. My clinical visit to the Institute of Pathology was a perfect day! Thanks a lot to Professor Inti Zlobec and her team!

Graduation ceremony

On March 9th, the University of Bern held its annual graduation ceremony. We are proud to announce that our first four students from the master's program in Artificial Intelligence in Medicine (MSc AIM) have graduated.

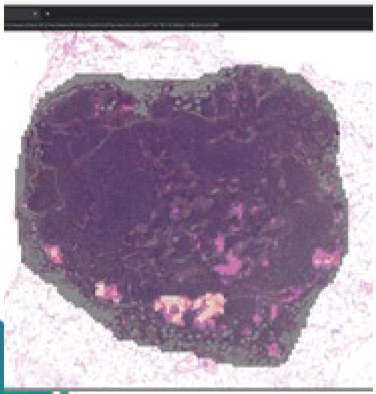
They showed enthusiasm and commitment since the beginning of their studies, and this is reflected in the projects they have undertaken as part of their master's theses. They have applied machine learning and deep learning techniques to various applications. These include the development of tools for viewing Whole Slide Images in Pathology, the detection of Diabetic Retinopathy in Infrared Fundus Images in Ophthalmology, the denoising of Monte Carlo dose distributions for Radiation Therapy, and the analysis of Fluorescence Flow Cytometry data for fast assessment of urine samples.

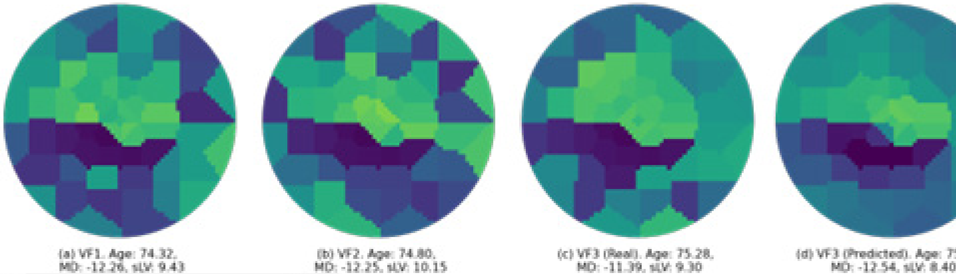
Exchange programs

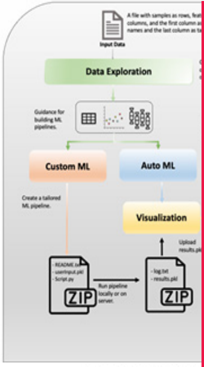
We're very pleased to announce we've signed our first exchange agreement with Harokopio University of Athens.

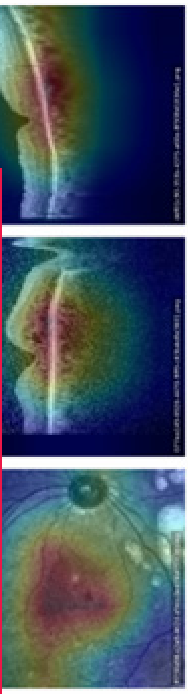
Students will be able to go on exchange to Athens, starting in September 2025 to one of two departments: the department of nutrition and dietetics or the department of informatics and telematics.

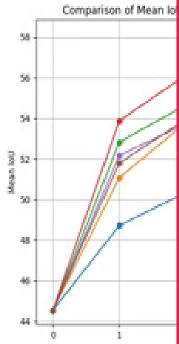
There are 4 spots available in the Department of Informatics and Telematics and 2 available in the Department of Nutrition and Dietetics.





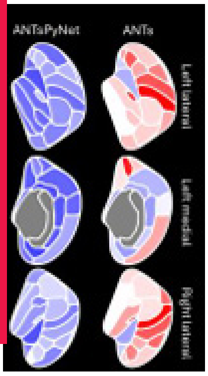


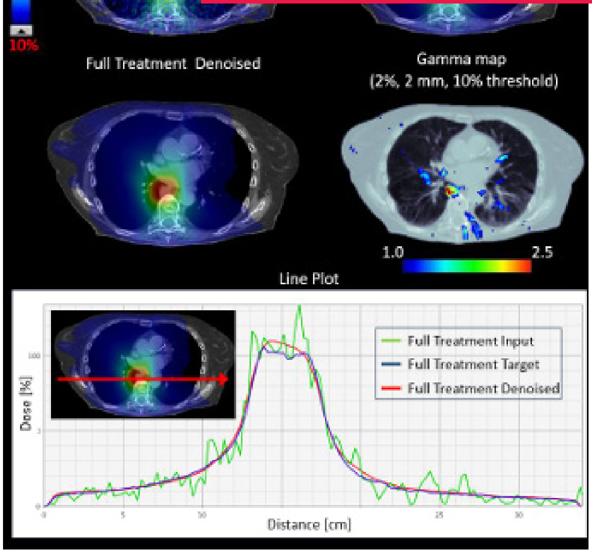


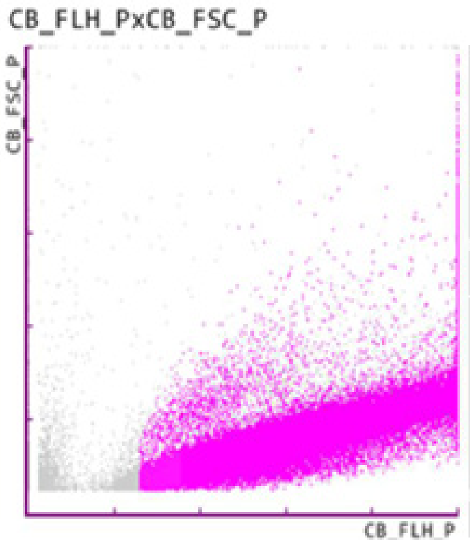


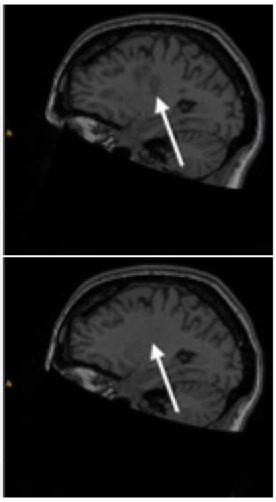
Master Theses

2023
2024









Development of a User Interface for Deep-zoom Whole Slide Images to Overlay Predictions from Deep Learning Models

Chris Rüttimann

Supervisor(s): Dr. Amjad Khan
Institution(s): Institute of Tissue Medicine and Pathology, University of Bern

Introduction

With the surge in Artificial Intelligence (AI) applications in healthcare, there's a pressing need for tools that allow for the effortless integration of AI into clinical settings. One such application is the viewing of Whole Slide Images (WSIs) along the results of deep learning algorithms in pathology settings.

Current tools, primarily designed for research purposes, often present clinicians with a steep learning curve due to their intricate interfaces and the prerequisite of additional software installations.

Additionally, these tools fall short when it comes to integrating with established high-performance computing (HPC) infrastructures, such as Ubelix at the University of Bern. In this study, we aim to develop a comprehensive tool that facilitates the viewing of WSIs by seamlessly integrating advanced AI models through HPC, thereby enabling the streamlined visualization of AI predictions within a single platform for pathology experts.

Materials and Methods

The application was developed using the Vue.js [1] framework for the frontend, which interacts with the backend via protocols such as HTTP and WebSockets. The backend was implemented in Python, handles complex tasks such as data synchronization with a shared drive and connectivity with Ubelix via SSH. The application currently uses an SQLite [2] database, with plans for future replacement with a more sophisticated system. The application was tested and evaluated for performance, usability, and functionality.

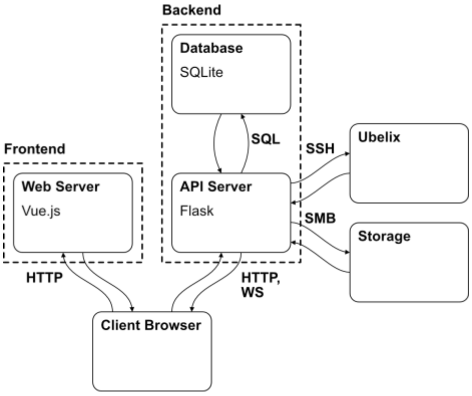


Fig. 1 Architectural Overview of the Application: Featuring Vue.js-based Frontend, Python Backend with SQLite Database, and Connectivity Mechanisms (HTTP, WebSockets, SSH).

Results

The application successfully enables users to view WSIs and deep learning outcomes for lymph node metastasis detection [3]. It exhibits high usability, supported by an intuitive user interface, and showcases commendable performance in terms of its speed and responsiveness. However, several challenges and areas for improvement were identified, including an offset issue in the classification mask, a significant bottleneck in the connection with the storage module, and limitations in user-started classification due to the duration of the classification process.

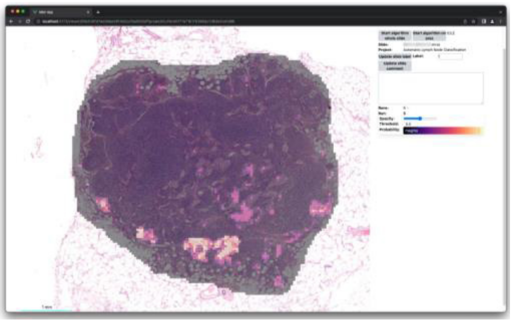


Fig. 2 WSI overlayed with the result of the deep learning algorithm from [3] shown in the final version of the frontend.

Discussion

Despite the identified challenges, the project was deemed successful, achieving its primary objective. Future work includes optimizing the system architecture by deploying dedicated servers for both frontend and backend, replacing the SQLite database with a more sophisticated system, and incorporating a user management system to facilitate concurrent use and project sharing among users.

References

[1] https://vuejs.org/
[2] https://www.sqlite.org/index.html
[3] A. Khan et al., Computer-assisted diagnosis of lymph node metastases in colorectal cancers using transfer learning with an ensemble model, Modern Pathology, vol. 36, no. 5, p. 100 118, 2023.

Acknowledgements

I am grateful for everyone who supported me during this project. Calculations were performed on UBELIX (http://www.id.unibe.ch/hpc), the HPC cluster at the University of Bern.

Proliferative Diabetic Retinopathy Detection with Multimodal Deep Learning

Yanis Schärer

Supervisor: Prof. Dr. Raphael Sznitman
Institutions: University of Bern, ARTORG Center for Biomedical Engineering Research
University Hospital Bern (Inselspital), Department of Ophthalmology

Introduction

Diabetic retinopathy (DR) is a prevalent and potentially blinding eye disease that affects individuals with diabetes. Early and accurate detection of DR is crucial for timely intervention to prevent irreversible vision loss. However, manual diagnosis of DR can be tedious and error-prone, underlining the need for automated and reliable detection systems. The first aim of this thesis is to classify infrared (IR) fundus images and optical coherence tomography (OCT) scans into non-proliferative diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). Additionally, the thesis explores the fusion of different imaging modalities to investigate the potential benefits of combining IR fundus images and OCT scans.

Methods

The dataset used in this study was acquired from the University Hospital of Bern. The models are trained and validated using triplets consisting of an IR fundus image, a horizontal OCT scan and a vertical OCT scan from the same eye.

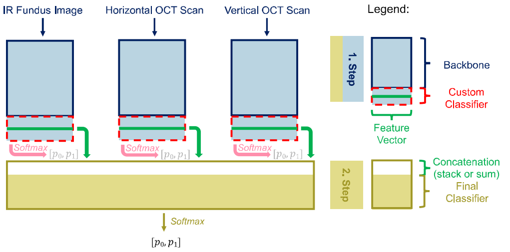


Figure 1: Architecture of the model. p_0 and p_1 are the probabilities of class 0 (NPDR) and class 1 (PDR), respectively.

The approach to fuse modalities involves pre-training unimodal backbone models in combination with a customized classifier, one for each modality. During the fusion process, all three images are passed through their respective unimodal model, and the resulting feature vectors, that is, the last linear layer of each model before classification, are extracted and subsequently fused together using a concatenation operation. Finally, the fused representation is fed into the final classifier, which generates the class prediction as the output.
In order to provide reliable visual interpretability, two established but distinct methods, HiResCAM and occlusion maps, were chosen and implemented.

Results

Training and evaluation was repeated three times to consider statistical variability. The AUC of the final multimodal model is 0.84 ± 0.01 and the AP is 0.68 ± 0.02. The narrow standard deviation indicates minimal variability across training runs. Among the unimodal backbones, the model trained on horizontal OCT scans consistently exhibits the highest AUC and AP (0.83 ± 0.03 and 0.63 ± 0.02, respectively).

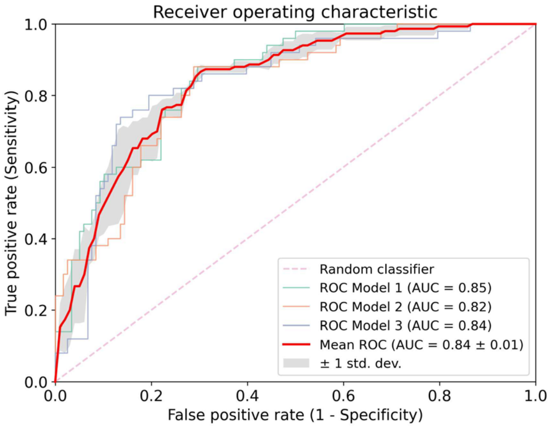


Figure 2: Receiver operating characteristic (ROC) curve for the multimodal model.

Fig. 3 representatively displays the heat maps of a correctly classified PDR (class 1) sample. The focus of the model on the IR fundus images is on areas known to be associated with the presence of neovascularization, a characteristic feature of PDR. In the OCT scans, the region of and around the macula is consistently highlighted, showing its significance in the model's decision-making process.

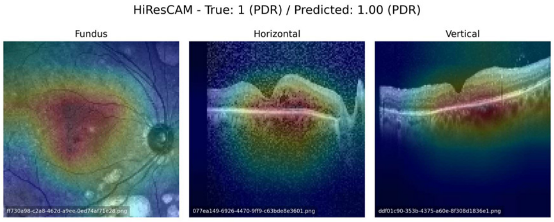


Figure 3: HiResCAM applied to a correctly classified PDR (class 1) triplet.

Discussion

The results are promising, particularly when considering the small size and the low quality of the dataset. The findings of this thesis demonstrate the potential benefits of multimodal fusion, as it enhances the performance of PDR detection compared to unimodal classification. It is important to note that the absence of a benchmark, such as the evaluation by medical specialists, limits the ability to contextualize the performance of the model in a clinical setting.

Deep Learning based Monte Carlo Dose Denoising for Radiation Therapy

Raphael Joost

Supervisor(s): Prof. Dr. Peter Manser, Prof. Dr. Stavroula Mougiakakou, Prof. Dr. Michael K. Fix, Hannes A. Löbner
Institution(s): University Hospital Bern (Inselspital), Division of Medical Radiation Physics
University of Bern, ARTORG Center for Biomedical Engineering Research

Introduction

Monte Carlo (MC) dose calculation is widely considered as the gold-standard for calculating dose distributions of radiotherapy plans, owing but not limited to its ability to accurately simulate dose distributions in inhomogeneous media. However, its clinical applicability is limited due to its long calculation-time for dose distributions with low statistical uncertainty. The purpose of this work is to develop a fast, deep learning based approach to denoise Monte Carlo dose distributions (MCDD) of high statistical uncertainty for 6, 10 and 15 MeV volumetric modulated arc therapy and to integrate it into the Swiss Monte Carlo Plan (SMCP)[1].

Materials and Methods

The proposed approach first calculates a MCDD of low statistical uncertainty, approximately 60%, of each field/arc of a radiotherapy treatment plan using 1.5 million primary particles. The resulting MCDD is subsequently denoised by a 4-layer 3D-U-Net, to predict a denoised MCDD of low statistical uncertainty, <1%. The model is trained on dose distributions of 106 clinically motivated VMAT arcs applied on 29 CTs. The plans are augmented to 3074 different arcs by random changes of machine parameters. MCDDs of low statistical uncertainty are used as ground truths (targets) to train the model. Different model input sizes are evaluated in this work. Patient-geometries that are smaller than the input size are padded with zero voxels, larger geometries are handled using a patch-based approach.

The final model accuracy is evaluated by means of gamma passing rates (3% (global)/ 3 mm, 10% threshold) and root-mean-squared error (RMSE) between target and denoised dose distributions for 307 samples from the test set, as well as a radiotherapy plan for a lung case consisting of two arcs. Dose distributions of individual arcs are summed to yield the full treatment dose distribution of the lung case. The case is additionally evaluated using a dose line plot and gamma map. Model performance is assessed by computation time.

Results

Model accuracy on the test set shows an average gamma passing rate of 94.0±2.3% (standard deviation) and RMSE of 0.12±0.02 mGy/MU for voxels with dose values greater than 10% of the maximum dose. Model performance on a CPU is on average 32.7±7.5 s for the input MCDD. The subsequent data-loading, preprocessing, and denoising take on average of 2.6±0.3 s on a GeForce RTX 3090 GPU. This compares to a MCDD with low statistical uncertainty taking on average 3.3±0.7 h on a CPU. The clinical lung case achieves an average gamma passing rate per arc of 97.37% and 99.83% for the full treatment (figure 1). The average computation time per arc is 32.5 s (input: 30.0 s, denoising 2.5 s).

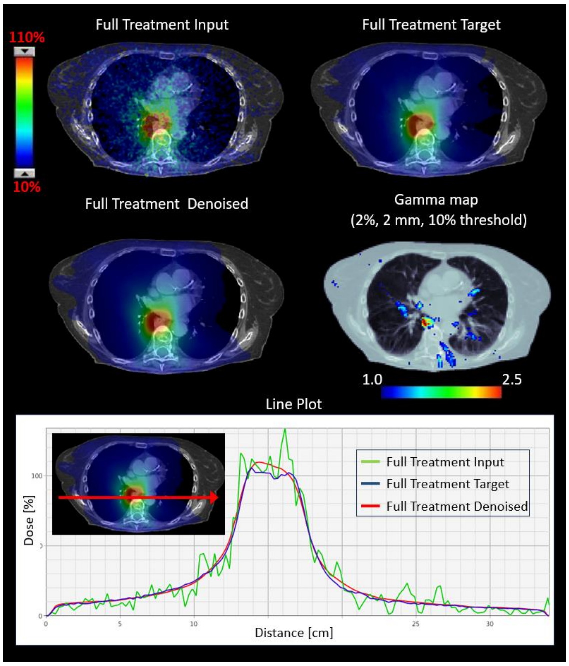


Fig. 1 2D slices of input, target and denoised dose distributions of the full treatment plan. The dose values are represented relative to the prescribed dose. The line plot shows dose values along the red arrow. The gamma map shows voxels that do not pass the gamma criterion (2% (global)/ 2 mm, 10% threshold), resulting in a gamma passing rate of 97.66% for this criterion.

The proposed approach is integrated into the SMCP framework and can be selected on its graphical user interface. All calculations are executed on the high-performance computing cluster UBELIX.

Discussion

A deep learning based approach to denoise MCDD is successfully developed and integrated into the SMCP framework. It offers a substantial reduction in computational time compared to a full MCDD calculation, while achieving reasonable accuracy.

References

[1] Fix et al., An efficient framework for photon Monte Carlo treatment planning, Phys. Med. Biol. 52, 2007.

Machine Learning Analysis of Automated Fluorescence Flow Cytometry Data for Fast and Efficient Microbiological Analysis of Urine Samples

Tim Graf

Supervisor(s): PD Dr. Alban Ramette
Institution(s): University of Bern, Institute for Infectious Diseases (IFIK)

Introduction

Urinary tract infections (UTIs) are among the most common infections, and urine microbiological cultures are the gold standard for diagnosis. As culturing is time-consuming, it cannot guide initial therapy, leading to unnecessary antibiotic treatments. Urine flow cytometry (UFC) can analyze urine samples within minutes and can be used as a screening tool to rule out negative cultures. Recent UFC analyzers have shown some ability to predict the Gram type of the bacteria in urine, which can help guide initial antibiotic treatment, if the prediction is reliable.

Materials and Methods

Retrospective data from microbiological analysis and UFC analysis of 3,148 urine samples (1,775 female, 1,373 male) sent to the microbiology lab at IFIK for routine diagnosis of urinary tract infection by urine culture was used. The primary aim was to develop thresholds for bacteria and leukocyte counts from UFC analysis to rule out negative urine cultures, with examination of different thresholds depending on age and sex. Additionally, extreme gradient boosting (XGBoost) models were used to incorporate more UF-5000 measurements into the decision process in order to better predict negative urine cultures.

In a secondary aim, evaluation of the Sysmex UF-5000 Gram typing performance was conducted, and machine learning algorithms were trained that aim to improve Gram typing performance from the bacteria scattergram produced by the UF-5000.

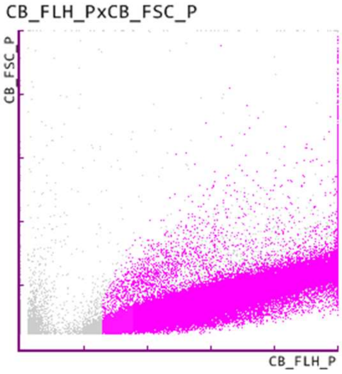


Fig. 1 Example bacterial scattergram produced by the SYSMEX UF-5000

Outcome measures included the area under the receiver operating characteristic curve (AUC), sensitivity (SE), specificity (SP), negative predictive value (NPV), and positive predictive value (PPV).

Results

Regarding the primary aim, thresholding of leukocyte and bacterial counts to rule out negative cultures yielded an overall AUC of 0.755 with SE 91.0% and SP 44.2%. Specific thresholds for women showed AUC of 0.702 with SE 92.5% and SP 31.5%, while the threshold for men yielded AUC of 0.823 with SE 88.1% and SP 55.6%. For patients older than 55 years, AUC of 0.735 with SE 91.9% and SP 37.3% was achieved, while thresholds for patients of age 55 or younger yielded AUC of 0.749 with SE 91.7% and SP 49.3%. Predictions with XGBoost yielded AUC of 0.787 with SE 86.1% and SP 56.8%.

Concerning the secondary aim, the UF-5000 detected Gram negative bacteria, Gram positive bacteria and Mixed Gram respectively with SE 72.7%, 50.0%, 45.5% and SP 84.0%, 85.1%, 94.6%. The machine learning approach using the data from the scattergram yielded SE 81.8%, 80.6%, 45.5% and SP 78.7%, 74.7%, 92.3% for Gram negative, Gram positive and Mixed Gram bacteria, respectively.

Discussion

These results are in line with the findings of similar studies performed with the Sysmex UF-5000. Bacterial counts from the UF-5000 have a higher predictive power on the growth result of urine culture than leukocyte counts, and thresholds for women are generally higher. Gram type prediction from both the UF-5000-interpreted data directly and the raw scattergrams yielded better results for Gram negative bacteria than Gram positive bacteria. Prediction of cultures with mixed Gram type was shown to be most difficult. While the Gram type predictions might help guide initial treatment, they are not accurate enough to replace urine cultures.

Acknowledgements

The contribution of Dr. Benjamin Schoenbeck, Ulrike Pauliks and Cara-Dorothea Schneider from SYSMEX for helping with the export of the data and answering technical questions regarding the UF-5000 is gratefully appreciated, as is the contribution from Dr. Peter Keller for the fruitful discussions and ideas.

Advancing Glaucoma Progression Prediction with Machine Learning Models

Marta Colmenar Herrera

Supervisor(s): Prof. Dr. Raphael Sznitman, Dr. Pablo Márquez Neila, Dr. Serife Seda Kucur
Institution(s): University of Bern, ARTORG Center for Biomedical Engineering Research
University Hospital Bern (Inselspital), Department of Ophthalmology

Introduction

Glaucoma is one of the leading causes of blindness worldwide. Accurate assessment of disease progression is crucial to intervene early and prevent irreversible vision loss. Traditional linear models often fail to capture the complex dynamics of glaucoma progression. This study uses machine learning (ML) techniques to improve current methods for predicting glaucoma progression.

Materials and Methods

We used three clinical datasets containing visual field (VF) data from glaucoma patients: the Bern dataset from Inselspital Eye Clinic (Switzerland) and the publicly available Rotterdam and Washington datasets. We assessed glaucoma progression in three tasks: stage classification prediction, binary classification on progression, and VF forecasting. For that, we built triplets of VFs, using the first two to forecast a property of the third. First, we compared non-linear models like the AdaBoost classifier with traditional linear models, such as Logistic Regression, to classify glaucoma stages. Next, the study explores various criteria for binary progression classification, highlighting the need for criteria independent of disease severity to reduce bias and improve prediction accuracy. Furthermore, an autoencoder (AE) model is developed to forecast glaucoma progression by reconstructing VFs (Fig. 1).

Results

In glaucoma stage prediction, non-linear models achieved high kappa metric values of 0.82 ± 0.02, 0.86 ± 0.00, and 0.85 ± 0.01 across the datasets. In comparison, Logistic Regression achieved kappa values of 0.66 ± 0.05, 0.86 ± 0.05 and 0.76 ± 0.01 for the same datasets. Binary progression classification yielded Receiver Operating Characteristic Area Under the Curve (ROC AUC) values exceeding 0.80 across all datasets. The AE model achieved Mean Absolute Error (MAE) values of 2.90 ± 3.16, 2.41 ± 2.87, and 2.56 ± 3.35, for the respective datasets.

This AE model demonstrates superior performance in capturing detailed VF changes over time, especially in progressed cases, surpassing the predictive accuracy of linear models, such as linear regression. The linear model achieved MAE values of 3.28 ± 3.47, 2.12 ± 2.49, and 2.76 ± 3.19, respectively. This AE model is also evaluated for its classification capacity, achieving kappa metric values of 0.83 ± 0.08, 0.95 ± 0.05, and 0.85 ± 0.06 for the respective datasets.

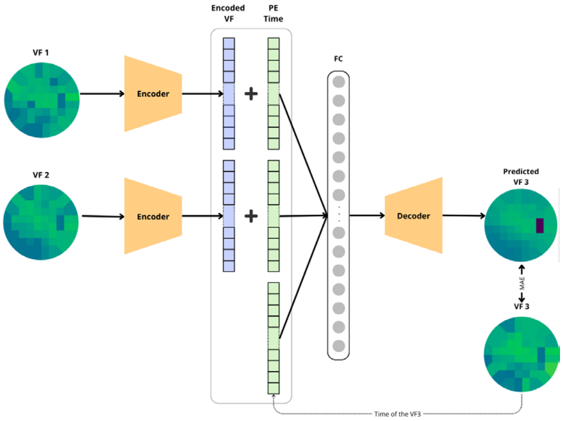


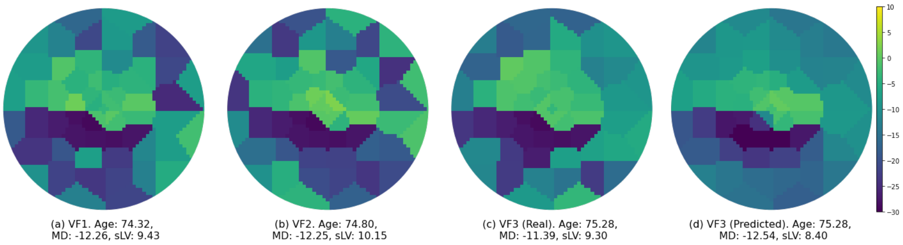
Fig. 1 Architecture of the Forecasting Visual Field Model.

Discussion

Overall, non-linear models consistently perform better than traditional ones in classifying glaucoma stages and forecasting disease progression. These findings underscore the advantages of using ML solutions for predicting glaucoma progression and reconstructing VF data, providing valuable insights that can enhance patient care and pave the way for future advancements in the field.

Acknowledgments

This research was funded by Innosuisse Grant 100.767 IP-LS.



(a) VF1. Age: 74.32, MD: -12.26, sLV: 9.43

(b) VF2. Age: 74.80, MD: -12.25, sLV: 10.15

(c) VF3 (Real). Age: 75.28, MD: -11.39, sLV: 9.30

(d) VF3 (Predicted). Age: 75.28, MD: -12.54, sLV: 8.40

Fig. 2 Example from the <u>Bern Dataset</u> showing Visual Field (VF) forecasting on the test dataset, with real and predicted VF images alongside patient age, Mean Defect (MD), and Loss Variance square root (sLV).

Machine Learning Made Easy (MLme) 2.0: Enhancing Analytical Capability with Regression Analysis

Jiahui Yu

Supervisor(s): Dr. Ali Hashemi Gheinani
Prof. Dr. Katia Monastyrskaya
Institution(s): University of Bern, ARTORG Center for Biomedical Engineering Research
University Hospital Bern (Inselspital), Department for BioMedical Research (DBMR)

Introduction

Machine learning (ML) has become an indispensable tool for the analysis of complex datasets in research, yet its integration often necessitates extensive domain knowledge and programming skills. In this thesis, we expanded, Machine Learning Made Easy (MLme), initially tailored for classification tasks, to its latest version, MLme2, to include regression analysis. This version enhances data handling capabilities and runtime efficiency, thereby increasing its utility for a broad spectrum of research applications.

MLme2 integrates four main components to facilitate its adoption among researchers: Data Exploration, which enables users to visualize dataset characteristics and perform data engineering tasks such as missing value imputation; AutoML, which automates the construction of ML pipelines across various regression algorithms, yielding detailed outputs including residual plots and feature importance diagrams; CustomML, which allows users with domain expertise in ML to design and implement custom pipelines through a user-friendly graphical interface; and Visualization, which provides tools for the effective presentation of both AutoML and CustomML results.

MLme2 has been rigorously tested across 35 benchmark datasets and eight omics-related datasets, demonstrating its robust performance and adaptability.

Materials and Methods

Figure 1 illustrates the graphical abstract of the MLme2. The input for regression analysis consists of a data file with samples as rows and features as columns, where the target variable is selected via a dropdown menu. Users typically begin by exploring their dataset using the Data Exploration module, which helps them gain an initial understanding of their data and prepare it for the next steps. Once well-prepared with the data structure, users can utilize the AutoML feature, which provides results automatically by simply uploading the data and clicking the start button. The AutoML feature includes automatic data checks, cleaning, and imputation, ensuring that users can obtain results even without prior data exploration. For experienced users, they can also use CustomML to create their own ML pipeline and run it on their own computer or server. Both AutoML and CustomML generate a results.pkl file, which users can then visualize and download reports using the Visualization module.

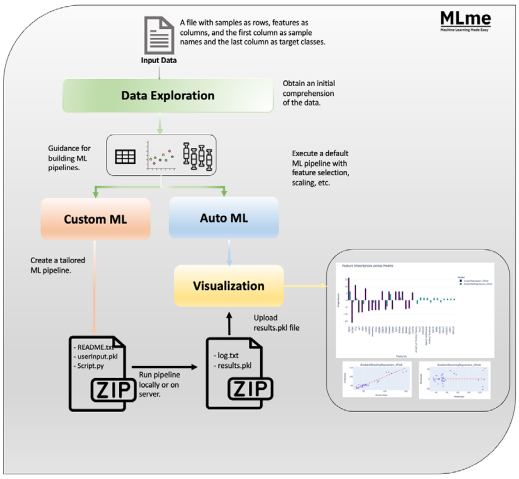
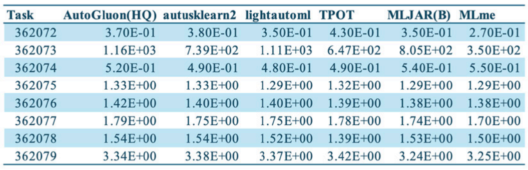


Fig. 1 Graphical abstract of MLme 2.

Results

Figure 2 shows the results of the collected omics datasets, comparing to other 5 AutoML tools.



Task	AutoGluon(HQ)	autusklearn2	lightautoml	TPOT	MLJAR(B)	MLme
362072	3.70E-01	3.80E-01	3.50E-01	4.30E-01	3.50E-01	2.70E-01
362073	1.16E+03	7.39E+02	1.11E+03	6.47E+02	8.05E+02	3.50E+02
362074	5.20E-01	4.90E-01	4.80E-01	4.90E-01	5.40E-01	5.50E-01
362075	1.33E+00	1.33E+00	1.29E+00	1.32E+00	1.29E+00	1.29E+00
362076	1.42E+00	1.40E+00	1.40E+00	1.39E+00	1.38E+00	1.38E+00
362077	1.79E+00	1.75E+00	1.75E+00	1.78E+00	1.74E+00	1.70E+00
362078	1.54E+00	1.54E+00	1.52E+00	1.39E+00	1.53E+00	1.50E+00
362079	3.34E+00	3.38E+00	3.37E+00	3.42E+00	3.24E+00	3.25E+00

Fig. 2 Results for the collected omics datasets, detailing the mean RMSE computed across 10 folds within a four-hour budget.

Discussion

From the results above, MLme2 has shown particular efficacy in medical and biological research, managing high-dimensional data and facilitating the prediction of clinical outcomes and the analysis of gene expression patterns. These results underscore MLme's potential as a transformative tool for advancing regression analysis across various scientific fields.

Acknowledgements
I extend my heartfelt thanks to everyone who have generously contributed their time and insights to this project

Diffusion-Based Filling and Synthesis of Multiple Sclerosis Lesions

Vinzenz Uhr

Supervisor(s): PD Dr. Richard McKinley
Institution(s): University Institute of Diagnostic and Interventional Neuroradiology, University of Bern, Inselspital, Bern University Hospital, Bern, Switzerland

Introduction

White matter (WM) lesions, common in neurological conditions like multiple sclerosis (MS), can significantly distort cortical thickness measurements derived from magnetic resonance imaging (MRI). Traditional methods often rely on lesion filling techniques to mitigate this issue. This thesis explores the potential of deep learning to improve the accuracy and efficiency of cortical thickness measurement in the presence of WM lesions. Additionally, it compares the robustness of various computational methods for cortical thickness measurement in the presence of WM lesions. A major challenge in deep learning projects for medical images is the scarcity of large datasets. To address this limitation, this thesis investigates the generation of synthetic data.

Materials and Methods

We explored different techniques and architectures for filling and synthesizing MS lesions in MRI images using noise diffusion models. We compared 2D and pseudo-3D architectures, as well as conditional and unconditional approaches. Additionally, we investigated the use of a synthetic mask distribution for training. To evaluate the robustness of various processing tools, we calculated the average absolute changes relative to the mean (%) when comparing with and without lesion filling. To assess the quality of the generated synthetic lesions, two trained neurosurgeons were asked to identify the synthetically added lesions among other lesions in 20 patients.

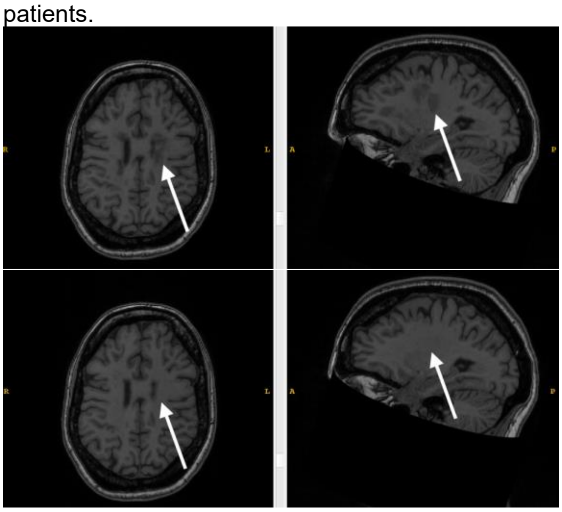


Fig. 1 T1w before (top) and after (bottom) lesion filling.

Results

The most effective model for lesion filling was a diffusion model using a pseudo-3D U-Net architecture conditioned on binary masks. This model was trained with a combination of real lesion masks and synthetic random circle masks. The best model for lesion synthesis was also a conditional model with the same pseudo-3D U-Net architecture. In the neurosurgeon evaluation, only three synthetic lesions were correctly identified out of 20. When comparing the robustness of different computational methods for cortical thickness, newer deep learning-based approaches exhibited superior performance.

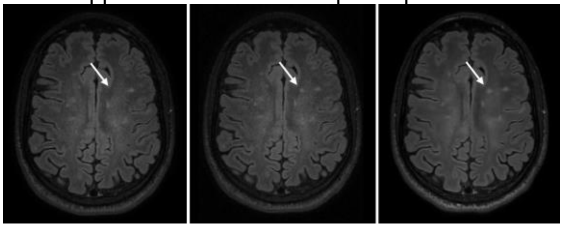


Fig. 2 Comparison of a patient's brain scans: (1) before adding a synthetic lesion to healthy tissue, (2) after adding the synthetic lesion, and (3) at a later timepoint with a natural new lesion at the same location.

Discussion

The trained models exhibited high-quality performance in lesion filling and lesion synthesis. A qualitative evaluation by two neuroradiologists confirmed the authenticity of the synthetic lesions. The superior performance of newer deep learning-based models for cortical thickness measurements suggests that lesion filling techniques may become less essential as these tools evolve.

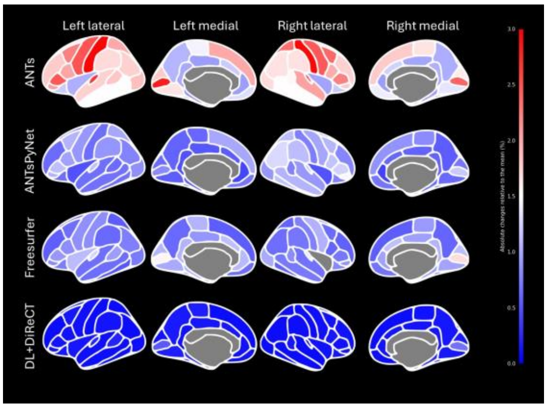


Fig. 3 Reproducibility errors of the ROI-wise average cortical thicknesses between different processing tools.

Applying active learning methods on state-of-the art pretrained models

Shunyu Wu

Supervisor(s): Fei Wu, Prof. Dr. Raphael Sznitman
Institution(s): University of Bern, ARTORG Center for Biomedical Engineering Research

Introduction

Active learning is designed to reduce the amount of labeled data required for training models by selecting the most informative samples for labeling. This research focuses on integrating state-of-the-art pretraining models such as SimSiam and DenseCL into an active learning framework, aiming to improve performance in semantic segmentation tasks. The study investigates whether the pretraining provided by self-supervised methods can further enhance the efficiency of active learning by reducing labeling costs and improving model accuracy on datasets such as Auris, Intuitive(Endovis), Pascal VOC, and Cityscapes.

Materials and Methods

Self-supervised learning (SSL) allows models to learn meaningful representations from unlabeled data through pretext tasks, which is crucial when labeled data is scarce. In contrastive learning (CL), a popular SSL method, models learn by contrasting positive pairs (similar images or views) with negative pairs (dissimilar images or views), effectively capturing image features. In this study, SimSiam, which focuses on image-level features without negative samples, and DenseCL, designed for dense pixel-level predictions, were pretrained on large datasets. These models were then fine-tuned for semantic segmentation on various target datasets, including medical (Auris, Intuitive) and urban scenes (Pascal VOC, Cityscapes). During active learning, several strategies were employed—BvSB, CBAL, entropy, pixelBal, random, and revisiting—and performance was evaluated using mean Intersection over Union (IoU) over multiple rounds.

Results

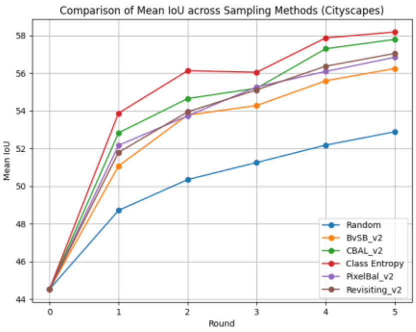


Fig. 1 Comparison of Mean IoU across 6 sampling methods over 6 rounds for baseline active learning training on Cityscapes datase

Fig.1 shows that class entropy is the best-performing method on cityscapes dataset, and also on other three datasets which is not shown here. Fig.2 and Table.1 shows that DenseCL pretraining improves performance in the active learning framework, especially for complex large datasets such as Pascal VOC and Cityscapes. DenseCL demonstrated approximately 1% improvement in each mean IoU compared to baseline (deeplabV3 with Resnet50 pretrained) models, whereas SimSiam's impact was limited especially in dense segmentation tasks. DenseCL outperformed the baseline in cumulative IoU area under curve (cAUC) across rounds, proving its effectiveness for pixel-level prediction tasks.

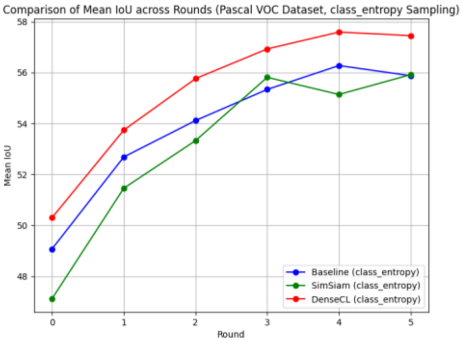
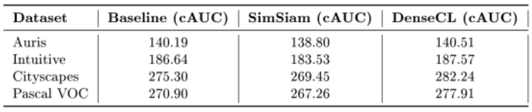


Fig. 2 Comparison of cAUC across Rounds for Class Entropy Sampling with SimSiam and DenseCL pretraining on Pascal VOC dataset



Dataset	Baseline (cAUC)	SimSiam (cAUC)	DenseCL (cAUC)
Auris	140.19	138.80	140.51
Intuitive	186.64	183.53	187.57
Cityscapes	275.30	269.45	282.24
Pascal VOC	270.90	267.26	277.91

Table. 1 Comparison of cAUC values across datasets with class entropy sampling for Baseline, SimSiam, and DenseCL pretraining.

Discussion

DenseCL's pixel-level feature extraction aligns well with the requirements of semantic segmentation, making it particularly effective in reducing labeling effort in active learning. In contrast, SimSiam's focus on object detection made it less suitable for dense tasks like segmentation. This study highlights the importance of selecting the appropriate pretraining model for the task at hand. Future research could explore the combination of active learning with other self-supervised models or domain-specific tasks, especially for medical imaging.

AI-based Analysis of Abdominal Ultrasound Images to Support Medical Diagnosis in Emergency Departments

Zahra Hamedi

Supervisor(s): Prof. Dr. Stavroula Mougiakakou, Dr. Lorenzo Brigato
Institution(s): University of Bern, ARTORG Center for Biomedical Engineering Research
University Hospital Bern (Inselspital)

Introduction

Ultrasound imaging is a non-invasive diagnostic tool that offers real-time visualization of internal organs and pathological conditions. With unclear boundaries, noise, and low contrast, ultrasound images present challenges for manual segmentation, which is highly reliant on the clinician's experience. However, Artificial Intelligence (AI)-driven methods can alleviate this difficulty by reducing the time and effort needed while making large-scale dataset analysis more feasible and faster. Our research seeks to locate and classify abdominal organs within ultrasound (US) images using AI-based techniques to aid medical diagnoses. Furthermore, we also concentrate on detecting free fluid, which can be a sign of malignancy.

Materials and Methods

We developed a tool utilizing MedSAM to simplify the creation of segmentation masks for ultrasound images, playing a crucial role in the data collection process. Using this tool, we gathered a dataset of ultrasound images, along with retrospective segmentation masks and organ-level labels, from Inselspital. Then, we used SAMed, a foundational segmentation model, to detect and classify abdominal organs in ultrasound images. Due to the small size of our dataset, we pre-trained SAMed on a larger public ultrasound dataset to fine-tune it for ultrasound imaging. We then applied the trained model to our dataset to identify and classify the organs visible in each image.

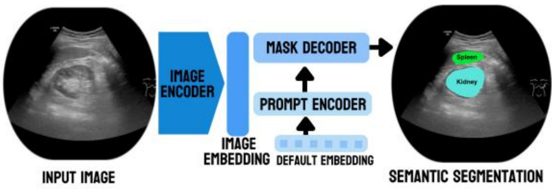


Fig. 1 Overview of the SAMed framework: The input image is processed through an image encoder, and a default embedding is passed to the prompt encoder. The resulting outputs are integrated in the decoder to generate the segmentation mask.

Results

To evaluate the model's performance, we calculated metrics such as Dice, HD95, Sensitivity, Precision, IoU, and Accuracy. Comparing results before and after fine-tuning SAMed on US images, revealed noticeable improvements across all metrics, notably enhancing Dice (+2.53), Sensitivity (+2.02), Precision (+2.33), IoU (+2.29), and Accuracy (+1.74), while HD95 saw a slight decline (-0.87).

It emphasizes the effectiveness of transfer learning in improving performance, especially when dealing with small datasets.

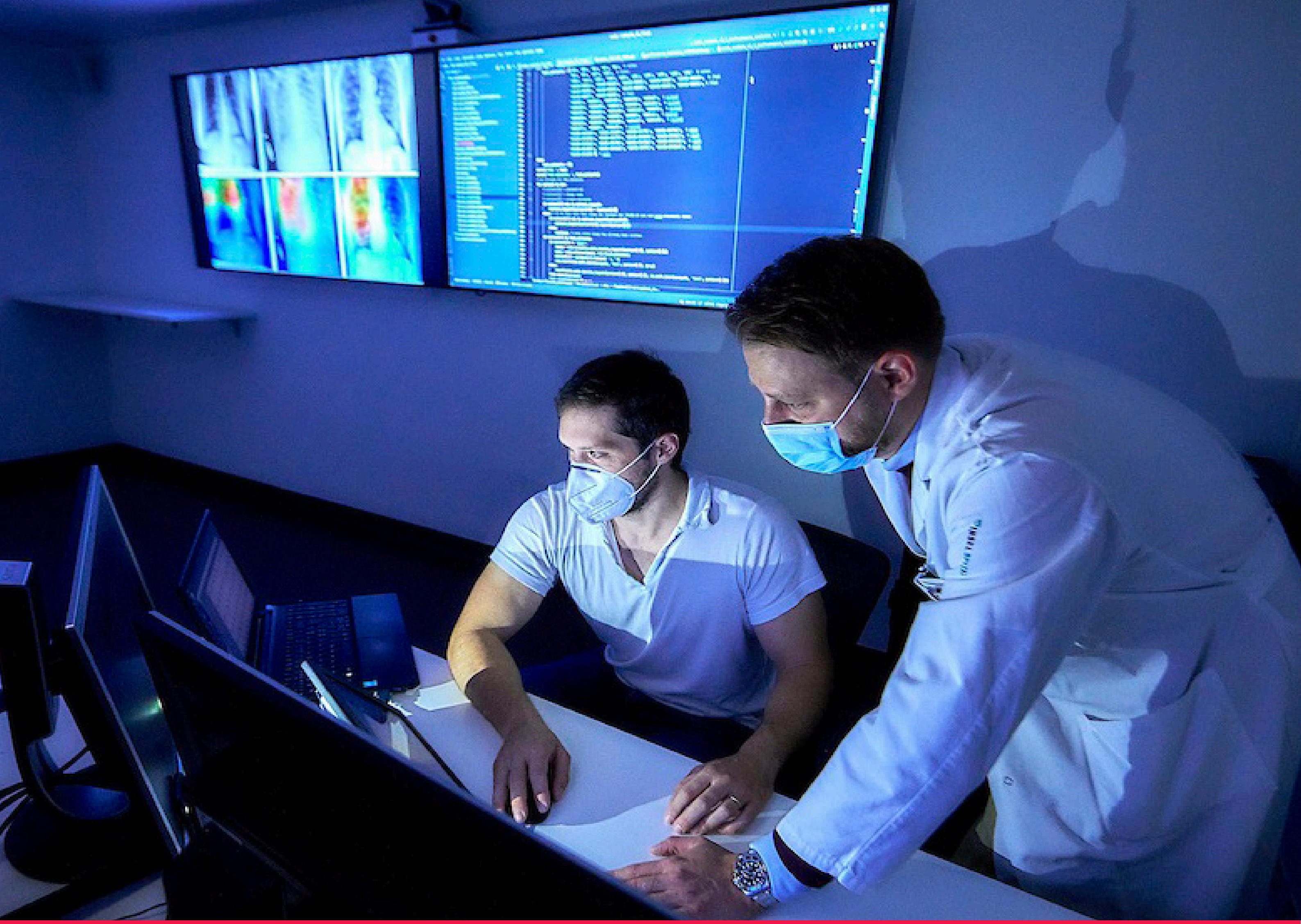
Fig. 2 Visualization of SAMed's predicted segmentation masks compared to the ground truth

Discussion

In this work, we demonstrated the practical application of foundation models on ultrasound images, focusing on the segmentation and classification of organs and free fluids to support medical diagnosis. While SAM shows great promise in general imaging tasks, its performance in medical imaging is often limited by severe degradation and poor generalization without proper adaptation to medical images. Consequently, we fine-tuned this model on ultrasound images to improve its performance in segmenting US data. Additionally, we handled the challenge of limited labeled data by employing transfer learning.

Acknowledgments

I extend my sincere thanks to Prof. Dr. Stavroula Mougiakakou and Dr. Lorenzo Brigato for their mentorship and for providing me with valuable feedback and guidance during this thesis.



Universität Bern
CAIM Center for Artificial Intelligence in Medicine
Master in Artificial Intelligence in Medicine
Freiburgstrasse 3
3010 Bern
Switzerland

https://www.caim.unibe.ch/education/msc__aim