

Jacek M. Bajor

CONTACT INFORMATION

Jacek M. Bajor

E-mail: jacek@bajor.me

SUMMARY

Experienced machine learning engineer and software developer specializing in deep learning applications. Proficient in computer vision, medical device development, data mining and natural language processing. With expertise to take a project from the proof of concept all the way through to production, ensuring high-quality and effective solutions.

PROFESSIONAL EXPERIENCE

CARIAD (Volkswagen Group), Berlin, Germany

Data Analyst (contract)

March - December 2022

Central Risk Management Group

- Performed data analysis of the company-wide risk-related data and automated report generation.
- Created data pipelines to process and centralize data from multiple sources.
- Building visual tools to enhance insight and facilitate early detection of threats (PowerBI) and live notification tools.

Data Analysis Business Intelligence Python Microsoft Azure Microsoft PowerBI Docker

Machine Learning Engineer (contract)

June 2021 - February 2022

Offline Perception Group/Cloud Assisted Reference System Group

- Developed a semantic segmentation computer vision model for automatic evaluation of advanced driver-assistance systems (ADAS).
- Researched and evaluated the state of the art models (based on convolutional neural networks and vision transformers) and techniques to optimize semantic segmentation maps.
- Ported a Tensorflow based machine learning pipeline to OpenMMLab (PyTorch based) for the use on the local machines and on the cloud (Azure ML).
- Developed high-performance algorithms (Cython, C++) for post-processing of semantic segmentation maps.
- Assisted the acquisition of the driving dataset (images and LIDAR data) and built algorithmic and visual tools for optimizing the data selection.

Computer Vision Deep Learning Python PyTorch Tensorflow Microsoft Azure ML Docker Airflow

OpenDress GmbH, Konstanz, Germany

Machine Learning Engineer (contract)

April - May 2021

- Consulted on architecture improvements to the convolutional neural network-based deep learning model (PyTorch) and the pipeline for sustainable development and deployment in the production environment.
- Designed a REST API for client access to the model.
- Deployed the model as a service accessible through a REST API.

Computer Vision Deep Learning Python PyTorch FastAPI

Berlin Institute of Health and German Heart Center, Berlin, Germany

Machine Learning Engineer/Senior Data Scientist

May 2019 - January 2021

- Performed analysis and processing of large quantities of continuous patient data and conducted research on predicting severe outcomes in the intensive care setting.
- Developed a recurrent neural network based model for predicting post-operative complications (PyTorch).
- Designed the software architecture incorporating the machine learning model.
- Led and conducted the development in accordance with processes for the certification under Medical Device Regulation.
- Integrated the product with the hospital's infrastructure and deployed it (Docker).

Deep Learning Time Series Medical Data Python PyTorch SQL Docker

Merantix/Vara Healthcare, Berlin, Germany

Data Engineer/Machine Learning Engineer

June 2018 - March 2019

- Led the effort to build a vast dataset of medical mammography images in collaboration with multiple healthcare sites in Europe.
- Anonymized and processed hundreds of thousands of images for use in a machine learning project.
- Developed the backend for an online platform for X-ray image annotation.
- Participated in development of a convolutional neural network based deep learning model for detecting and classifying abnormalities in mammography images.

Computer Vision Deep Learning Medical Data Python Tensorflow DICOM Google Cloud Platform Docker

Vanderbilt University Medical Center, Nashville, TN, USA

Machine Learning Engineer/Research Programmer

Nov. 2013 - May 2018

Department of Biomedical Informatics, Lasko Lab

- Conducted research on computational representation learning and its application to electronic medical records [1].
- Developed supervised models using recurrent neural networks, noisy medical data to predict multi-label targets [3,9].
- Built natural language processing (NLP) based tools for extracting features from medical notes.
- Designed and developed a web application for visualization of medical history employing semantic embedding for medical concepts.
- Compared efficacy of different data formats and common predictive models for medical outcome prediction [2].
- Implemented algorithms for statistical modeling in Python and C optimizing for speed and API simplicity.
- Configured and maintained a GNU/Linux based computation server.

Deep Learning Medical Data Natural Language Processing Python Keras

Software Developer

Sept. 2012 - Nov. 2013

Center for Human Genetics Research

- Developed and maintained several web applications for human genetic research data management and collaboration.
- Designed and developed mobile applications for iOS and Android including a computer vision program for genetic sample management.
- Configured and maintained a GNU/Linux based web and database servers.

Web Development Mobile Development Python Ruby Java

University of Virginia, Charlottesville, VA, USA

Research Assistant

Sept. 2011 - Sept. 2012

Department of Molecular Physiology and Biological Physics, Minor Lab

- Developed large structural biology oriented databases and web applications for research collaboration and communicating results to the public [4,12].
- Configured and administered a network of highly utilized GNU/Linux based servers.
- Expanded the web-based laboratory information management system, by providing unified data storage, improving data visualization and sharing capabilities in structural genomics laboratories [4,10,11].
- Developed a high-throughput pattern recognition system, which automatically scans images in search of protein crystals [13].

Web Development Computer Vision Python Ruby SQL

EDUCATION

Łódź University of Technology, Poland

M.Sc., Computer Science

Sept. 2006 - Sept. 2011

- Department: International Faculty of Engineering
- Thesis Topic: *Automated Crystal Recognition with Haar like Features and Hough Transform*
- Adviser: Professor Paweł Strumiłło

University of Virginia, Charlottesville, VA, USA

Visiting student, Computer Science

Sept. 2010 - Sept. 2011

Copenhagen University College of Engineering, Denmark

Erasmus International Exchange Student, Computer Science Aug. 2008 - June 2009

DEVELOPMENT AND PROGRAMMING SKILLS

Machine Learning

- Data engineering and data analysis:
 - Developing high-performance tools for processing diverse data including:
 - Continuous and discrete medical data
 - High-resolution images (including medical)
 - Natural language processing (NLP)
 - Building data processing pipelines (Airflow, cloud solutions)
 - Data visualization (PowerBI, Dash, matplotlib, plotly, etc.)
- Deep learning:
 - Computer vision with models based on convolutional neural networks, vision transformers
 - Sequential and time-series data models with recurrent neural networks and transformers
 - Familiarity with state of the art deep architectures used in medical applications and computer vision.
- Classical machine learning:
 - Supervised methods including linear regression, logistic regression, random forests, XGBoost, etc.
 - Unsupervised methods including PCA, ICA, t-SNE, autoencoders, clustering, semantic embeddings, etc.
- Tools and libraries:
 - PyTorch, OpenMMLab, Tensorflow, Keras, scikit-learn, Numpy, Scipy, Pandas, Gensim, etc.

Programming concepts and paradigms

- Object-oriented programming, algorithms, design patterns, debugging, performance optimization.

Programming Languages

- Python, Bash scripting, SQL proficient
- C++, C, JavaScript working knowledge
- Ruby, R, Lua, Java, Clojure, Objective-C basic knowledge

Cloud solutions

- Microsoft Azure, Amazon Web Services (AWS), Google Cloud Platform (GCP), etc.

Operating Systems and Administration

- GNU/Linux, networking, security, container orchestration, working with high-performance clusters (HPC).

LANGUAGES

English

proficient (speaking, reading, writing)

German

intermediate, currently learning

Polish

native language

PUBLICATIONS

- [1] **Bajor, J.**, Mesa, D.A., Osterman, T.J., Lasko, T.A., Embedding Complexity In the Data Representation Instead of In the Model: A Case Study Using Heterogeneous Medical Data. *ACM Conference on Knowledge Discovery and Data Mining*, August 2018 (*submission*, <https://arxiv.org/abs/1802.04233>).
- [2] Lasko, T.A., **Bajor, J.**, Using Scalable, Generic Data Representations with Computational Feature Learning for Clinical Prediction Problems. (*in preparation*)
- [3] **Bajor, J.**, Lasko, T.A., Predicting Medications from Diagnostic Codes with Recurrent Neural Networks. *5th International Conference on Learning Representations*, April 2017 (<https://openreview.net/pdf?id=rJEgeXFex>).
- [4] Copper D., Grabowski M., Zimmerman M., Porebski P., Schabalin I., Woinska M., Domagalski M., Zheng H., Sroka P., Cymborowski M., Czub M., Niedzialkowska E., Venkataramany B. S., Osinski T., Fraczak Z., **Bajor, J.**, Gonera J., MacLean E., Wojciechowska K., Konina K., Wajerowicz W., Chruszcz M., W. Minor, State-of-the-Art Data Management: Improving the Reproducibility, Consistency, and Traceability of Structural Biology and in Vitro Biochemical Experiments. *Part of the Methods in Molecular Biology book series*, (MIMB, volume 2199), 2021.
- [5] **Bajor, J.**, Tkaczuk, K.L., Kagan, O, Chruszcz, M, Savchenko, A., Joachimiak, A., Minor, W., The first structure of NMT1/THI5-like domain-containing protein, CAE31940 from *Bordetella bronchiseptica* RB50. *Journal of Structural and Functional Genomics*, 15(2):73–81, June 2014.
- [6] Shabalin, I.G., Bacal, P., **Bajor, J.**, Winsor, J., Grimshaw, S., Anderson, W.F., Minor, W., CSGID, Crystal structure of unliganded anabolic ornithine carbamoyltransferase from *Vibrio vulnificus*. *Protein Data Bank deposit 4KWT*, May 2013
- [7] Shabalin, I.G., Winsor, J., Grimshaw, S., Osinski, T., **Bajor, J.**, Chordia, M.D., Shuvalova, L., Anderson, W.F., Minor, W., CSGID, Crystal structure of anabolic ornithine carbamoyltransferase from *Vibrio vulnificus* in complex with carbamoyl phosphate, *Protein Data Bank deposit 4JFR*, February 2013
- [8] **Bajor, J.**, Kagan, O., Chruszcz, M., Savchenko, A., Joachimiak, A., Minor, W. Structure of CAE31940 from *Bordetella bronchiseptica* RB50, *Protein Data Bank deposit 3QSL*, March 2011

POSTER
PRESENTATIONS

- [9] **Bajor, J.**, Lasko, T.A., Predicting Medications from Diagnostic Codes with Recurrent Neural Networks. *5th International Conference on Learning Representations, Toulon, France*, April 2017.
- [10] **Bajor, J.**, Blus, B.J., Dworzynski, P., Wojciechowska, K., Zimmerman, M.D., Minor, W. Tracking fluorescence-based thermal shift assay experiments: data management of functional experiments. *High-Throughput Structural Biology Symposium, Keystone, Colorado, USA*, January 2012.
- [11] Zimmerman, M.D, **Bajor, J.**, MacLean, E., Cymborowski, M., Fraczak, Z., Wojciechowska, K., Wajerowicz, W., Grabowski, M., Minor, W. The LabDB crystallographic laboratory information management system. *High-Throughput Structural Biology Symposium, Keystone, Colorado, USA*, January 2012.
- [12] Grabowski, M., Zimmerman, M.D, **Bajor, J.**, Domagalski, M., Fraczak, Z., Minor, W. The Architecture of a Data Management System for PSI:Biology Centers. *High-Throughput Structural Biology Symposium, Keystone, Colorado, USA*, January 2012.
- [13] **Bajor, J.**, Dayal, A., Fraczak, Z., Minor, W. Automated Crystal recognition with Haar like features and Hough transform. *41st Mid-Atlantic Protein Crystallography Workshop, Frederick, Maryland, USA*, June 2011.