

# Jacek M. Bajor

---

## CONTACT INFORMATION

Jacek M. Bajor

*E-mail:* jacek@bajor.me

## SUMMARY

Skilled data scientist and software developer proficient in machine learning with the background in computational medicine and medical imaging. Experienced in deep learning applications. Well versed in object-oriented programming languages, algorithms, and advanced data structures. Skilled in GNU/Linux administration. Eagerly explores new technologies, techniques, and trends in machine learning and software development.

## PROFESSIONAL EXPERIENCE

**Berlin Institute of Health**, Berlin, Germany

*Senior Data Scientist/Machine Learning Engineer* **May 2019 - December 2020**

- 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.
- Designed the software architecture incorporating the machine learning model.
- Led and conducted the development in accordance with processes for the product to be certified under Medical Device Regulation.
- Integrated the product with the hospital's infrastructure and deployed it.
- Project was a part of the [Digital Health Accelerator](#) at Berlin Institute of Health in collaboration with [Deutsches Herzzentrum Berlin](#).

**Merantix/MX Healthcare**, Berlin, Germany

*Data 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.

**Vanderbilt University Medical Center**, Nashville, TN, USA

*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 [3,8].
- 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.

*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.

**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].

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:
  - Processing of vast amounts of continuous (vital signs) and discrete (lab results, medications) medical data
  - Data visualization
  - Anonymization of electronic medical records
  - Processing of medical images (DICOM)
- Neural networks and deep learning:
  - Convolutional neural networks, recurrent neural networks, autoencoders, etc.
  - State of the art deep architectures used in medical applications
- Supervised learning:
  - Linear regression, logistic regression, random forests, XGBoost, etc.
- Unsupervised learning:
  - PCA, ICA, t-SNE, autoencoders, clustering, semantic embeddings, etc.
- Natural Language Processing
- Tools and libraries:
  - PyTorch, TensorFlow, Keras, Scikit-learn, NumPy, SciPy, Pandas, Gensim, etc.

**Programming concepts and paradigms**

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

**Programming Languages**

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

**Operating Systems and Administration**

- GNU/Linux, networking, security, container orchestration.

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.