Suleyman Onur Dogan

onur.dogan@sabanciuniv.edu | sonurdogan.github.io | +90 544 826 7907

EDUCATION

Sabanci University - Master of Science in Computer Science & Engineering *2024-2026*
- *Full Scholarship | %100 English*
  - Courses: Scalable Learning Systems, Machine Learning, NLP, IOT
  - Teaching Assistant: Programming Languages (200+ students)

Antalya Bilim University - Bachelor of Science in Computer Engineering *2019-2024*
- *GPA: 3.71/4 | %100 English*

EXPERIENCE

Graduate Research Assistant Sept 2024 – Present
- *Tastan Lab (Asst. Prof. Oznur Tastan), Sabancı University*
  - Working on NLP applications in Computational Biology, my current research includes Sequential Models, Fine-Tuning Protein Language Models (Language models for protein sequences), the Interpretability in NLP & LLMs.

Junior Data Scientist Sept 2023 – March 2024
- *John Snow Labs, USA (Remote)*
  - Taken parts in Spark-NLP for Healthcare projects that includes building Retrieval-Augmented Generation (RAG) - LLM pipelines, creating NLP pipelines and training NER, Classification models, integrating models on AWS EMR and Glue, and Databricks cloud services.
  - Contributed the development of **open source NLU library** which allows applying latest NLP models from Spark-NLP and HugginFace, directly on any dataframe, with **a single line of code.**

Junior Machine Learning Engineer March – June 2023
- *GrainFox, Canada (Remote)*
  - Improved and **deployed an API which automatically extracts the information from contract** documents using Computer Vision techniques, reducing %90 of manual effort through automation.
  - Developed a **Large Language Model (LLM) agent** using LangChain and OpenAI API as an innovative approach to extract information from poorly processed or unstructured documents.
  - Integrated service APIs with job queue mechanisms into the production pipeline and contributed to **MLOps** implementations using FastAPI, Django, Docker, RabbitMQ, Celery, and Traefik, leading to a 35% reduction in latency.

Undergraduate Research Assistant March 2022 – Nov 2023
- *Computational Biology Group (Professor Hilal Kazan), Antalya Bilim University*
  - Worked on Tubitak (Scientific and Technological Research Council of Turkey) Funded Project – "Neuroimmune guidance cues, MicroRNAs & Inflammatory responses: Sex differences in cardiovascular diseases" under supervision of Professor Hilal Kazan.
  - Focused on the differential gene co-expression network with pairwise and cluster-wise analysis, via ML based clustering algorithms and statistical methods in bulk RNA-Seq datasets.
  - Contributed to the development of SciTuna, scRNA-seq **dataset Integration Tool** which uses Network Alignment algorithm to integrate multiple scRNA-seq dataset into single dataset.

Research Intern July – Sept 2022
- *Tastan Lab (Asst. Prof. Oznur Tastan), Sabancı University*
  - Developed an open-source project called TLMSA for detection of possible SUMOylating sites that emerges through mutations in protein sequence data via fetching data from National Cancer Institute GDC database and UniProt, manipulating data and using deep sequential neural network model.

Undergraduate Research Assistant Sept 2020 – July 2022
- *Statistical Signal Processing Lab, Antalya Bilim University*
  - Conditional information transfer between neurons and accurate neuronal network inference on time series action-potential data investigated with Conditional Transfer Entropy from Information Theory via implementation of IDTxl library.

TECHNICAL SKILLS

- **Programming Languages:** Python, Java, C#, C++, C, R, Matlab, JS, HTML, CSS, SQL
- **ML Frameworks**: PyTorch, Tensorflow, Keras, HuggingFace, Scikit-Learn, LangChain, OpenCV, Spark-NLP, IDTxl
- **Frameworks**: Django, FastAPI, Docker, GitHub, Streamlit, Java Spring, .NET, RabbitMQ, Celery, Traefik.
- **Cloud**: AWS EMR, AWS Glue, AWS S3, Databricks

Publications & Conference Presentations

- Aissa Houdjedj, Yacine Marouf, Mekan Myradov, **Onur Dogan**, Burak Onur Erten, Oznur Tastan, Cesim Erten, Hilal Kazan. SciTuna: Single Cell RNA-seq data integration tool using network alignment. **Under Review**

- Mehmet Uğur KAHRAMAN, Süleyman **Onur DOĞAN**, Sevgi ŞENGÜL AYAN, Yaren ŞEKERCİ, Ferhat KOYUNCU, Hakan Bal. Harmony in Diversity: Exploring Eclectic Chair Designs Through Deep Neural Networks Analysis in Interior Design Education. **Under Review**

- Berke Dilekoglu, **Onur Dogan**, Oznur Tastan. Deep sequential prediction of sumoylation sites with Sumonet and PLMs **Ongoing Publication Process**

- Yacine Marouf, **Onur Dogan**, Ernest Diez Benavente, Gerard Pasterkamp, Hester M. den Ruijter, Janine van Gils, Katey Rayner, Hilal Kazan. Sex-biased Expression of Neuroimmune Guidance Cues in Cardiovascular Diseases. **Presented at Conferences** (ERA-CVD'22, Latvia, 2022) & (HIBIT'22, Turkey, 2022)

- **O.Dogan,** S.Ş. Ayan, D.Gencaga. Analysis of Neuronal Interaction Using Information-Theoretical Models. **Presented at** The 8'th International Congress on Fundamental an Applied Science (ICFAS'21, Oct 19-21, 2021)

External PROJECTS

- **Graph Neural Networks for Directed Multigraph**   Fall 2024 (Active)
    - The motivation of the project is to improve the performance of Multi-GNN with methods such as network centrality. Also, integrating Federated Learning since Multi-GNN works with Financial Dataset. Project is currently active in the scope of Scalable Learning System Course and supervised by **Associate Professor Kubilay Atasu** who is also one of the authors of the paper.

- **Domain Generalization for Surface Anomaly Detection**   Fall 2024 (Active)
    - The motivation of the project is to perform domain generalization for detecting anomalies in different material types used in the industry. Project is currently active in the scope of Machine Learning Course.

- **Healthcare Tourism App – Thesis Project**   Spring 2023
    - Healthcare tourism app is developed as thesis project where Patient can be able to plan the trip and make reservation on hospital and hotel in one app.
    - I developed a **healthcare tourism chatbot** that works based on our SQL database, utilizing RAG with LLM using LangChain, FAISS, Gemini API, FastAPI.

- **The use of Deep Learning in Biomedical Imaging**   Fall 2022
    - The application of CNN and transfer learning-based models to biomedical images is investigated using examples from brain tumor classification and Covid-19 diagnosis as part of Biomedical Imaging course of Prof. Umit Demirbas.

- **Matrix Factorization for Movie Recommendation System with Parallel Computing**   Fall 2021
    - In sophomore year, I focused on application of Matrix Factorization technique with ALS on Collobrative Filtering-based recommendation system using PySparkML.

- **SLAM Simulation of Indoor Robot**   Spring 2021
    - In Freshman year, Simulations performed with a BreezySLAM algorithm by using existing Odometry and Lidar Data for SLAM indoor robot. Throughout the project, Kalman Filtering and Particle Filtering in SLAM algorithms were the main emphasis.

Volunteer EXPERIENCE

Google Developers Group Antalya   Sept 2020 – Nov 2021
- *Co-Organizer*
    - Aim is to gather people interested in technology, to organize activities and to follow the technological agenda.
    - Designed and leaded developers.june robotics event where we hosted five experts in different fields of robotics.

Certificates

- Machine Learning, *Stanford University Online*
- Spark NLP for Data Scientist, *John Snow Labs*