Nuriye Özlem Özcan Şimşek

Bağlıca Mah.
Etimesgut/ANKARA/TURKEY

Phone: +90 535 414 9841
E-mail: nozlemozcan@gmail.com
Url: nozlemozcan.github.io

RESEARCH INTERESTS

Bioinformatics, Natural Language Processing, Machine Learning, Data Mining

EDUCATION

Boğaziçi University, İstanbul, Turkey   DoG: June 2024
   Ph.D., Computer Engineering
   Advisor: Prof. Fikret Gürgen, Prof. Arzucan Özgür
   Thesis: "Genomic Data Analysis using Machine Learning Methods For Disease and Disease-Gene Prediction"

                                                  DoG: Feb. 2010
Boğaziçi University, İstanbul, Turkey
   M.S., Systems and Control Engineering
   Advisor: Prof. Fikret Gürgen
   Thesis: "A Fuzzy Support Vector Machine Approach For ECG Analysis"

Middle East Technical University,   DoG: June 2005
Ankara, Turkey
   B.S., Computer Engineering

RESEARCH EXPERIENCE

Boğaziçi University, İstanbul, Turkey   DoG: June 2024
Ph.D., Computer Engineering, Advisor: Prof. Fikret Gürgen, Prof. Arzucan Özgür
In this thesis, genomic diseases and disease-gene associations were investigated based on DNA mutations. We used DNA mutations of cancer patients from The Cancer Genome Atlas

(https://www.cancer.gov/ccg/research/genome-sequencing/tcga) repository (ERA Commons authorized access with PID:15042) and UK Biobank (https://www.ukbiobank.ac.uk/) repository (authorized access with PID:44337). The summary of our contributions is as follows:

- We transferred our knowledge from the field of information retrieval to mutations domain. We modelled VCF files as documents consisting of mutations as words and calculated term weighting methods on the mutation words. We wanted to reflect the different effects of each mutation on disease development. We performed a comparison of these data representation schemes for the task of cancer type classification. We proposed a neural network architecture for disease classification and showed that the proposed method achieves significantly better performance in discriminating cancer types.
- We proposed a novel and disease specific gene selection method for gene expression data. We extracted the most effective genes for disease development based on DNA mutations and transferred this information to the gene expression domain as a novel gene selection method. Our test results showed that the new feature selection method leads to equivalent or better performance compared to other feature selection methods and manually curated cancer related gene sets.
- We proposed a novel heterogeneous graph structure for the representation of patients and proteins for disease classification. We modelled patients and proteins as separate nodes in the graph. The edges between proteins were defined according to the PPI network. The edges between patients and proteins were defined based on DNA mutation information. The proposed graph structure was able to represent the protein interactions and gene mutations together. We processed this heterogeneous graph with GAT, which can distinguish the effects of neighbours on a target node. The proposed framework led to a leap in classification performance compared to other machine learning methods, both graph-based and non graph-based.
- We applied explainers to our classification networks. For the neural network architecture, we proposed the explainer method. For the graph based neural network architecture, we used Captum explainer. With these explainers, we identified the most effective genes in cancer type classification. We searched the literature and found that nearly half of the identified genes have been reported as driver or target genes for the corresponding cancer types. This showed us that the proposed methods make predictions based on true causal genes and the rest of the identified genes can be investigated as candidate genes in further cancer study.

Boğaziçi University, İstanbul, Turkey   DoG: 2010
M.S., Systems and Control Engineering, Advisor: Prof. Fikret Gürgen
In this thesis, a fuzzy approach was used for ECG analysis. The ECG dataset in the UCI (https://archive.ics.uci.edu/) repository was used. Five different fuzzy membership functions were proposed and compared. The best performing fuzzy support vector approach was compared to other classification methods for the task of ECG arrhythmia detection. The test results showed that the fuzzy support vector machine outperforms other methods. In order to interpret the classification model, rule base extraction methods were applied. The resulting membership functions identified with ANFIS were found to be consistent with medical knowledge.

PUBLICATIONS

Refereed Journal Papers

[1] Özcan Şimşek, N.Ö., A. Özgür, F. Gürgen, "GNNMutation: a Heterogeneous Graph-Based Framework for Cancer Detection Based on DNA Mutations". Accepted to be published in BMC Bioinformatics. 2025 [Q1, SCIE]

[2] Özcan Şimşek, N.Ö., A. Özgür, F. Gürgen, "A Novel Gene Selection Method for Gene Expression Data for the Task of Cancer Type Classification", Biology Direct, Vol. 16, Article ID 7, 2021 [Q1, SCIE]

[3] Özcan Şimşek, N.Ö., A. Özgür, F. Gürgen, "Statistical Representation Models for Mutation Information within Genomic Data", BMC Bioinformatics, Vol. 20, Article ID 324, 2019 [Q1, SCIE]

Refereed Conference and Workshop Papers

[1] Özcan, N.Ö., F. Gürgen, "Fuzzy Support Vector Machines for ECG Arrhythmia Detection", IEEE International Conference on Pattern Recognition, pp. 2973-2976, İstanbul, Turkey, 2010

Thesis

[1] Nuriye Özlem Özcan Şimşek, "Genomic Data Analysis using Machine Learning Methods for Disease and Disease-Gene Prediction", Ph.D. Thesis, Computer Engineering, Boğaziçi University, İstanbul, Turkey, June 2024. Thesis Advisor: Prof. Fikret Gürgen, Prof. Arzucan Özgür

[2] Nuriye Özlem Özcan, "A Fuzzy Support Vector Machine Approach For ECG Analysis", M.S. Thesis, Systems and Control Engineering, Boğaziçi University, İstanbul, Turkey, Feb. 2010. Thesis Advisor: Prof. Fikret Gürgen

Book Chapters

[1] Nuriye Özlem Özcan Şimşek, A. Özgür, F. Gürgen, "Pathogenicity Scoring Methods and Applications", Book Title: "Next Generation Sequencing and Clinical Applications", Editor: Prof. Ahmet Okay Çağlayan, Güneş Tıp Kitabevleri, Ankara, 2023

WORK EXPERIENCE

Boğaziçi University, İstanbul, Turkey   Oct. 2024 – now
Post-Doctoral Researcher, *Department of Computer Engineering*

Project Title: Understanding the Language of Life: Identifying and Characterizing the Language Units in Protein Sequences, LifeLU (ERC-2022-CoG, ID: 101089287)
Principal Investigator: Arzucan ÖZGÜR
Project Details: https://ufukavrupa.org.tr/en/success-stories/arzucan-ozgur-i-erc-2022-consolidator-grant-i-lifelu-i-understanding-language-life

Boğaziçi University, İstanbul, Turkey   Feb. 2022 – Mar. 2024
Research Assistant, Department of Computer Engineering

Boğaziçi University, İstanbul, Turkey   Summer 2017 – Fall 2021
Teaching Assistant, Department of Computer Engineering

TEACHING EXPERIENCE

Boğaziçi University, İstanbul, Turkey   Summer 2017 – Fall 2021
Teaching Assistant, Department of Computer Engineering
- CMPE 250 - Data Structures and Algorithms
  Fall 2017, Fall 2018 (Prof. Ali Taylan Cemgil), Fall 2021 (Assist. Prof. Birkan Yılmaz)
  Responsibilities included leading problem sessions, preparing and grading projects and helping in preparing and grading examinations.
- CMPE 260 - Principles of Programming Languages
  Summer 2017 (Prof. Albert Ali Salah), Summer 2018 (Prof. Tunga Güngör), Summer 2019 (Assist. Prof. Fatma Başak Aydemir)
  Responsibilities included leading problem sessions and preparing and grading projects.
- CMPE 462 – Machine Learning
  Summer 2021 (Assist. Prof. İnci Meliha Baytaş)
  Responsibilities included preparing and grading homeworks and projects.
- CMPE 300 - Analysis of Algorithms
  Fall 2020 (Prof. Tunga Güngör)
  Responsibilities included preparing and grading weekly quizes and projects.
- CMPE 230 - Systems Programming
  Summer 2019 (Prof. Can Özturan)
  Responsibilities included leading problem sessions and helping in grading projects.
- CMPE 321 - Introduction to Database Systems
  Summer 2018 (Prof. Taflan Gündem)
  Responsibilities included preparing and grading projects and helping in grading examinations.

- CMPE 352 - Fundamentals of Software Engineering
  Summer 2017 (Assist. Prof. Suzan Üsküdarlı)
  Responsibilities included supervising student team projects, helping in grading examinations, helping project setup and grading.

PROFESSIONAL ACTIVITIES

- Reviewer for SIU 2024 - 32nd IEEE Conference on Signal Processing and Communications Applications, 2024
- External reviewer for IEEE Journal of Biomedical and Health Informatics, 2021

RESEARCH GRANTS, AWARDS AND SCHOLARSHIP

- Boğazici University Center for Researcher and Academician Development in Information and Telecommunications Technologies, PhD. Scholarship, Feb. 2022 – Mar. 2024
- Boğazici University BAP, "Genomic Data Analysis using Machine Learning Techniques for Disease Prediction", Principal Investigator: Prof. Fikret Gürgen, My Role: Researcher, July 2017 – July 2019
- ECCB Conference Travel Fellowship, 2022
- CAMDA Conference Travel Fellowship, 2019

TALKS/POSTERS AND PRESENTATIONS

- Poster Presentation, "Breast Cancer Detection in UKBiobank Data using Mutation Based Gene Weighting and Deep Learning", ECCB 2022 - 21th European Conference on Computational Biology, Barcelona, Spain, 2022
- Conference Talk, "Statistical Representation Models for Mutation Information within Genomic Data", CIBCB 2019 - 16th IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology, Siena, Italy, 2019
- Conference Talk, "A Novel Gene Selection Method for Gene Expression Data for the Task of Cancer Type Classification", CAMDA 2019 – 18th Annual International Conference on Critical Assessment of Massive Data Analysis (as a part of ISMB 2019), Basel, Switzerland, 2019
- Poster Presentation, "Statistical Representation Models for Mutation Information within Genomic Data", ISMB 2019 - 27th Conference on Intelligent Systems for Molecular Biology, Basel, Switzerland, 2019
- Poster Presentation, "Analysis of Data Representation Techniques and Machine Learning Methods on Variant Calls for Classification of Cancer Types", ECCB 2018 - 17th European Conference on Computational Biology, Athens, Greece, 2018
- Poster Presentation, "A Novel Proof for the Existence of Unknown Causal Genes for Cancer Types", HIBIT 2018 - International Symposium on Health Informatics and Bioinformatics, Antalya, Turkey, 2018

LANGUAGES

Turkish (native), English (advanced)

REFERENCES

Prof. Fikret Gürgen
Department of Computer Engineering
Boğaziçi University
gurgen@bogazici.edu.tr

Prof. Arzucan Özgür
Department of Computer Engineering
Boğaziçi University
arzucan.ozgur@bogazici.edu.tr