

DIMITRIS SPATHIS

Nokia Bell Labs Cambridge
Broers Building, 21 JJ Thomson Ave
Cambridge, CB3 0FA, UK
dimitrios.spathis@nokia.com
ds806@cl.cam.ac.uk
www.cl.cam.ac.uk/~ds806

EDUCATION

- 2017-21 **PhD in Computer Science/Machine Learning**
University of Cambridge, UK (Advisor: Cecilia Mascolo)
Thesis: Deep learning for mobile health sensing
- 2015-17 **MSc in Computational Intelligence/Machine Learning**
Aristotle University of Thessaloniki, Greece (Advisor: Anastasios Tefas)
Thesis: Interactive dimensionality reduction and manifold learning
- 2011-15 **BSc in Computer Science**
Ionian University, Greece (Advisor: Panagiotis Vlamos)
Thesis: Diagnosing respiratory diseases with machine learning

EXPERIENCE

- Nokia Bell Labs**, Cambridge, UK
- 2023- *Senior Research Scientist, Device Intelligence team*
Leading efforts on data-efficient and robust AI with an eye on the future.
- 2022 *Research Scientist, Device Intelligence team*
Worked on multimodal self-supervised learning, continual learning, and fairness/robustness. Submitted multiple papers and filed 2 patents.
- University of Cambridge**, UK
- 2022- *Visiting Researcher, Department of Computer Science and Technology*
Industry partner of the Centre for Mobile/Wearable Systems & Aug. Intelligence.
- 2022 *Postdoctoral Research Associate, Department of Computer Science and Technology*
Led machine learning projects on self-supervised learning and health sensing.
- 2018- *Teaching Assistant/Supervisor, Various Cambridge Colleges and CST Department*
Supervised 3 BSc and 2 PhD theses, and tutored students for the courses *Scientific Computing, Machine Learning & Real-World Data*, and *Mobile & Sensor Systems*.
- 2017-21 *PhD Researcher, Department of Computer Science and Technology*
Developed new AI models for multimodal behavioral and physiological data building on the paradigms of self-supervision, multi-tasking, and transfer learning.
- 2021 **Microsoft Research**, Cambridge, UK (Sep-Nov)
Research Intern in Health Intelligence, hosted by Dr. Stephanie Hyland
Worked on transfer learning & domain generalization for physiological timeseries. Presented the results in a paper at ML4H'22 (co-located with NeurIPS'22).
- 2019 **Ocado**, Barcelona, Spain (Jun-Sep)
Data Science Intern in Predictive Maintenance, hosted by Dr. Laurent Candillier

Developed deep learning anomaly detection models for warehouse robot collision prevention, internal systems monitoring, and industrial conveyor belts.

2017 **Qustudio**, Barcelona, Spain (Jun-Sep)

Data Scientist in Product, Intern

Built data pipelines to automate C-level KPI reporting, conducted A/B tests, custom cohort analyses and clustering to model customer retention and churn.

2016 **Telefonica Research**, Barcelona, Spain (Sep-Dec)

Research Intern, hosted by Dr. Ilias Leontiadis

Worked on next-character prediction language models for hate speech detection and published results at a workshop of ACL 2017 in Vancouver.

2015 **Aristotle University of Thessaloniki**, Greece (Oct-Feb)

Teaching Assistant, supervised by Prof. Anastasios Tefas

Grader for the core undergraduate course in Numerical Analysis.

Center for Research and Technology Hellas (CERTH), Greece (Jul-Sep)

Research Engineer Intern

Designed and developed the prototype mobile app for the Horizon-funded asthma modelling research project *myAircoach*.

2013-15 **Social Informatics Lab**, Ionian University, Greece

Undergraduate Researcher

Gained experience in 3 university labs and co-authored papers in computational linguistics and machine learning for health.

DISTINCTIONS

2023 Selected as a 'Rising Star in AI' by Jürgen Schmidhuber's KAUST AI initiative
Invited talk at the honorary symposium (Saudi Arabia)

2022 Hall of Fame Better Future Award, Cambridge Ring
Dept. of Computer Science and Technology, University of Cambridge

MobiCom Travel Award, ACM SIGMOBILE

\$1000 grant to attend *MobiCom'21 (USA)*

2021 Graduate Research Travel Grant, Jesus College, University of Cambridge
£500 grant to attend *MobiCom'21 (USA)*

2019 KDD Student Travel Award, ACM SIGKDD
\$1000 grant to attend *KDD'19 (USA)*

Graduate Research Travel Grant, Jesus College, University of Cambridge
£500 grant to attend *PervasiveHealth'19 (Italy)*

2018 PhD Open Day, Facebook office, London
Selected to attend internal research presentations along with 60 PhD students and presented one of the 10 shortlisted posters

AI Summer School, Microsoft Research, Cambridge

Selected to attend an intensive one-week program for top AI PhD students from EMEA

- 2017 Embiricos Trust Scholarship, Jesus College, University of Cambridge & EPSRC Doctoral Training Partnership (DTP) Grant, University of Cambridge Full PhD scholarship
- 2016 Marie Skłodowska-Curie Research and Innovation Staff Exchange Grant (RISE), Full scholarship for a 4-month secondment at Telefonica Research, Spain
- 2015 Admission Scholarship to the MSc program, Aristotle University of Thessaloniki Full tuition fees scholarship for the first year of the program
- Investigative Data Journalism Contest, *Innovathens*
Glocal News web-app, 3rd prize
- 2014 Erasmus+ multinational project management training in Iceland Full financial support by EU *YouthInAction*, *Youthnet Hellas*
- 2013 Educational trip to Silicon Valley. Selected among 250+ students nationwide. Attended courses at *Stanford University* and visited *UC Berkeley*, *Google*, *Facebook*, *Twitter*, and *Coursera*. Read more: <http://medium.com/p/37e8d2ca0421>

PUBLICATIONS

- Peer-reviewed journals, conferences, workshops, or book chapters
Google Scholar (1153 citations, 16 H-index, as of June 2023)
<https://scholar.google.co.uk/citations?user=rzKY1UAAAAJ>
1. Dang, T., Han, J., Xia, T., Bondareva, E., Brown, C., Chauhan, J., Grammenos, A., Spathis, D., Cicuta, P., & Mascolo, C. (2023). Conditional Neural ODE Processes for Individual Disease Progression Forecasting: A Case Study on COVID-19. *ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, Long Beach, USA (to appear)
 2. Han, J., Montagna, M., Grammenos, A., Xia, T., Bondareva, E., Siegele-Brown C., Chauhan, J., Dang, T., Spathis, D., Floto, A., Cicuta, P., & Mascolo, C. (2023). Evaluating Listening Performance for COVID-19 Detection by Clinicians and Machine Learning: A Comparative Study. *Journal of Medical Internet Research (JMIR.)*, 25. <https://doi.org/kd4b>
 3. Coppock, H., ..., Spathis, D., ..., Schuller, B. (2023). A Summary of the ComParE COVID-19 Challenges. *Frontiers in Digital Health*, 5. <https://doi.org/kd4f>
 4. Spathis*, D., Pozuelo*, I., Gonzales, T., Wu, Y., Brage, S., Wareham, N., & Mascolo, C. (2022). Longitudinal cardio-respiratory fitness prediction through wearables in free-living environments. *Nature Digital Medicine (npj Digit. Med.)*, 5(176). <https://doi.org/jpcc>
☆ PRESS (U CAMBRIDGE, ACM COMMUNICATIONS, DAILY MIRROR, BICYCLING MAG)
☆ ALTMETRIC TOP 5% OF ALL RESEARCH OUTPUTS
 5. Han*, J., Xia*, T., Spathis, D., Bondareva, E., Brown, C., Chauhan, J., Dang, T., Grammenos, A., Hasthanasombat, A., Floto, A., Cicuta, P., & Mascolo, C. (2022). Sounds of COVID-19: exploring realistic performance of audio-based digital testing. *Nature Digital Medicine (npj Digit. Med.)*, 5(16). *equal contribution
<https://doi.org/hf cz>

6. Dang, T., Han, J., Xia, T., Spathis, D., Bondareva, E., Brown, C., Chauhan, J., Grammenos, A., Hasthanasombat, A., Floto, A., Cicuta, P., & Mascolo, C. (2022). Exploring Longitudinal Cough, Breath, and Voice Data for COVID-19 Progression Prediction via Sequential Deep Learning: Model Development and Validation. *Journal of Medical Internet Research (JMIR)*, 24(6) <https://doi.org/h2p7>
7. Spathis, D., Pozuelo, I., Marques-Fernandez, L., & Mascolo, C. (2022). Breaking away from labels: the promise of self-supervised machine learning in intelligent health. *Cell Patterns*, 3(2). <https://doi.org/hjtt>
8. Greenberg, D., Wride, S., Snowden, D., Spathis, D., Potter, J., & Rentfrow, J. (2022). Universals and variations in musical preferences: A study of preferential reactions to Western music in 53 countries. *Journal of Personality and Social Psychology*, 122(2), 286–309. <https://doi.org/hgjn>
 ☆ PRESS (U CAMBRIDGE, TIMES, TELEGRAPH, CNN, WASHPO, SKY, ABC, TEDX)
 ☆ ALTMETRIC TOP 5% OF ALL RESEARCH OUTPUTS
9. Spathis, D., Hyland, S. (2022). Looking for Out-of-Distribution Environments in Multi-center Critical Care Data. *Machine Learning for Health (ML4H)*. <https://doi.org/kd4k>
10. Wu, Y., Spathis, D., Jia, H., Pozuelo, I., Gonzales, T., Brage, S., Wareham, N., & Mascolo, C. (2022). Turning Silver into Gold: Domain Adaptation with Noisy Labels for Wearable Cardio-Respiratory Fitness Prediction. *Machine Learning for Health (ML4H)*. <https://doi.org/kd4j>
11. Hasthanasombat, A., Ghosh, A., Spathis, D., & Mascolo, C. (2022). Investigating Domain-agnostic Performance in Activity Recognition using Accelerometer Data. *UbiComp workshop on Human Activity Sensing Corpus & Applications (HASCA @ UbiComp)*, Cambridge, UK. <https://doi.org/kd4h>
12. Pozuelo, I., Posa, M., Spathis, D., Westgate, K., Wareham, N., Mascolo, N., Brage, S., & Palloti, J. (2022). Detecting sleep outside the clinic using wearable heart rate devices. *Scientific Reports*, 12, 7956. <https://doi.org/htx2>
13. Xia*, T., Spathis*, D., Brown, C., Grammenos, A., Han, J., Hasthanasombat, Bondareva, E., Chauhan, J., Dang, T., Floto, A., A., Cicuta, P., & Mascolo, C. (2021). COVID-19 Sounds: A Large-Scale Audio Dataset for Digital Respiratory Screening. *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks track*. <https://openreview.net/pdf?id=9KArJb4r5ZQ>
 ☆ PRESS (FINANCIAL TIMES, STAT NEWS)
14. Spathis, D., Pozuelo, I., Brage, S., Wareham, N., & Mascolo, C. (2021). Self-supervised transfer learning of physiological representations from large scale free-living wearable data. *ACM Conference on Health, Inference, and Learning (CHIL)*. <https://doi.org/f6tt>
15. Shah, K., Spathis, D., Tang, I., & Mascolo, C. (2021). Evaluating Contrastive Learning on Wearable Timeseries for Downstream Clinical Outcomes. *Machine Learning for Health (ML4H)*, short paper. <https://doi.org/hhst>
16. Han, J., Brown*, C., Chauhan*, J., Grammenos*, A., Hasthanasombat*, A., Spathis*, D., Xia*, T., Cicuta, P., & Mascolo, C. (2021). Exploring automatic COVID-19 diagnosis via voice and symptoms from crowdsourced data. *IEEE International*

Conference on Acoustics, Speech, and Signal Processing (**ICASSP**).

<https://doi.org/gc25>

17. Tang, C., Pozuelo*, I., Spathis*, D., & Mascolo, C. (2021). SelfHAR: Improving Human Activity Recognition through Self-training with Unlabeled Data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT/Ubicomp)*, 5(1). <https://doi.org/f6tv>
18. Schuller, B., ... Spathis, D., Xia, T., Cicuta, P., Rothkrantz, L., Zwerts, J., Treep, J., & Kaandorp K. (2021). The INTERSPEECH 2021 Computational Paralinguistics Challenge: COVID-19 Cough, COVID-19 Speech, Escalation & Primitives. *Proceedings of the Conference of the International Speech Communication Association (Interspeech)*. <https://doi.org/gzxx>
19. Pozuelo, I., Spathis, D., Gifford-Moore, J., Morley, J., & Cowls, J. (2021). Digital Phenotyping and Sensitive Health Data: Implications for Data Governance. *Journal of the American Medical Informatics Association (JAMIA)*. <https://doi.org/fsxq>
★ PRESS (US NATIONAL COMMITTEE FOR QUALITY ASSURANCE)
20. Searle, B., Spathis, D., Constantinides, M., Quercia, D., & Mascolo, C. (2021). Anticipatory Detection of Body-focused Compulsive Behaviours with Wearables. *ACM Conference on Mobile Human-Computer Interaction (MobileHCI)*. <https://doi.org/gzxx>
21. Pozuelo, I., Spathis, D., Clifton, E., & Mascolo, C. (2021). Wearables, smartphones and artificial intelligence for digital phenotyping and health. **Digital Health**, Elsevier, pp. 33–54. <https://doi.org/fpfv>
22. Spathis, D., Pozuelo, I., Brage, S., Wareham, N., & Mascolo, C. (2020). Learning Generalizable Physiological Representations from Large-scale Wearable Data. *Advances in Neural Information Processing Systems (NeurIPS-W), Machine Learning for Mobile Health workshop, Virtual Event*. <https://doi.org/hjvc>
★ PRESS (VENTUREBEAT, BUSINESS INSIDER, ACM TECHNEWS)
23. Tang, C., Pozuelo, I., Spathis, D., & Mascolo, C. (2020). Exploring Contrastive Learning in Human Activity Recognition for Healthcare. *Advances in Neural Information Processing Systems (NeurIPS-W), Machine Learning for Mobile Health workshop, Virtual Event*. <https://doi.org/hjvb>
24. Brown*, C., Chauhan*, J., Grammenos*, A., Han*, J., Hasthanasombat*, A., Spathis*, D., Xia*, T., Cicuta, P., & Mascolo, C. (2020). Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. *ACM International Conference on Knowledge Discovery and Data Mining (KDD), Virtual Event*. <http://doi.org/d683>
★ ORAL PRESENTATION ★ PRESS (U CAMBRIDGE, NPR, PSYCHOLOGY TODAY, ETC)
★ CAMBRIDGE UNIVERSITY HALL OF FAME BETTER FUTURE AWARD
25. Spathis D., Rodriguez, S., Farrahi, K., Mascolo, C., & Rentfrow, J. (2019). Sequence Multi-task Learning to Forecast Mental Wellbeing from Sparse Self-reported Data. *ACM International Conference on Knowledge Discovery and Data Mining (KDD), Anchorage, USA*. <http://doi.org/gf7nbh>
★ ORAL PRESENTATION, TOP 6% OF SUBMISSIONS

26. Spathis, D., Servia, S., Farrahi, K., Mascolo, C., & Rentfrow, J. (2019). Passive mobile sensing and psychological traits for large scale mood prediction. *International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth)*, Trento, Italy. <http://doi.org/c7hk>
27. Spathis, D., Passalis, N., & Tefas, A. (2019). Interactive dimensionality reduction using similarity projections. *Knowledge-Based Systems*, 165:77-91. <http://doi.org/cxbm>
28. Spathis, D., Passalis, N., & Tefas, A. (2018). Fast, Visual, and Interactive Semi-supervised Dimensionality Reduction. *European Conference on Computer Vision (ECCV-W), Workshop on Compact and Efficient Feature Representation and Learning in Computer Vision*, Munich, Germany. <http://doi.org/cz6d>
29. Spathis, D., Vlamos, P. (2017). Diagnosing Asthma and Chronic Obstructive Pulmonary Disease with Machine Learning. *Health Informatics Journal*, 25(3): 811:827. <http://doi.org/cbzh>
30. Serrà, J., Leontiadis, I., Spathis, D., Stringhini, G., Blackburn, J., Vakali, A. (2017). Class-based Prediction Errors to Detect Hate Speech with Out-of-vocabulary Words. *Annual Meeting of the Association for Computational Linguistics (ACL-W), Workshop on Abusive Language Online*, Vancouver, Canada. <http://doi.org/b94p>
31. Charalampakis, B., Spathis, D., Kouslis, E., & Kermanidis, K. (2016). A comparison between semi-supervised and supervised text mining techniques on detecting irony in greek political tweets. *Engineering Applications of Artificial Intelligence*, 51:50–57. <http://doi.org/bzxq>
32. Charalampakis, B., Spathis, D., Kouslis, E., & Kermanidis, K. (2015). Detecting Irony on Greek Political Tweets: A Text Mining Approach. *International Conference on Engineering Applications of Neural Networks (EANN)*, 17:1–5. <http://doi.org/bzxr>
33. Spathis, D., Mouratidis, T., Sioutas, S., & Tsakalidis, A. (2014). Glocal News: An Attempt to Visualize the Discovery of Localized Top Local News, Globally. *International Conference on Conceptual Modeling (ER), Workshop on Legal and Social Aspects in Web Modeling*, Hong Kong, China. Springer LNCS 8697. <http://doi.org/bzxs>

ACADEMIC SERVICE & LEADERSHIP

— Organizer

Co-chair of **FairComp** and **WellComp** workshops at **UbiComp 2023**, Mexico.
 Co-organizer and track chair of **CHIL 2023**, USA.
 Senior panel/roundtable chair at **ML4H 2022**, New Orleans, USA.
 Chair of **WellComp** workshop at **UbiComp 2022**, Cambridge, UK.
 Co-organizer of the *Federated Sensing* tutorial at **MobiCom 2021**, USA.

— Program Committee Member

AAAI 2021-2023, **IJCAI 2020**, **KDD 2020-2023** (PC & Session Chair), **FACCT 2023**, **SIAM SDM 2022**, **Sensiblend @ UbiComp 2021**, **Mobiquitous 2022**.

— Reviewer

NeurIPS, ICLR, ICML, AAAI, IJCAI, KDD, CHI, Ubicomp/IMWUT, CHIL, Nature Digital Medicine, Nature Scientific Reports, ICASSP, Expert Systems with Applications, Neurocomputing, WWW/The Web Conference, MobileHCI etc.

PROJECTS

— Open-source code and contributions on GitHub <https://github.com/sdimi>

1. *Covid-19 Sounds* app. Looking for COVID-19 biomarkers in respiratory sounds. 2020. (180,000+ visits) <http://covid-19-sounds.org>
☆ PRESS (BBC, FORBES, GUARDIAN, FINANCIAL TIMES, SLATE, NPR, AND MORE)
2. Anonymize kids' faces before posting online. 2018.
<https://devpost.com/software/patronus-k61iv4>
Mobile app with face recognition, age estimation, & emotion recognition to blur kids' faces or replace with emotion-based emoji (*HackZurich* project).
☆ PRESS (CNN MONEY SWITZERLAND)
3. Map out your musical taste on *Spotify*. Published in *Cuepoint Magazine*. 2016.
(25,000+ visits) <https://medium.com/p/fe50c94b8af3>
"Nifty music-oriented dataviz" – (tweet) Paul Lamere, Director Dev. Platf., *Spotify*.
4. How do popular book authors use language differently? 2015.
Text mining *Game of Thrones*, *Harry Potter*, *Hunger Games* and *Lord of the Rings* books. (15,000+ visits) <http://medium.com/p/100290c94242>
Tweeted by @Medium to 2 million followers, featured in *Editor Picks*.
5. Visualizing pop-culture TV references. 2015.
Every pop-culture reference on the TV series *Community* broken down by episode and character. (15,000+ visits) <http://communitypoprefs.com>
"This is very cool" – (tweet) Josef Adalian, Editor, *New York Magazine*.
6. Google News mashup on top of Google Maps. 2013.
<https://github.com/sdimi/glocalnews.js/> (now defunct due to API closure)
The U.S. Dept. of Health expressed interest in using the app for policy research.

SKILLS

— Computing

Python: Numpy, Pandas, Scipy, Jupyter, Tensorflow, PyTorch Sklearn, Keras, Matplotlib, Seaborn, Statsmodels etc. *Systems*: Unix/Bash, Slurm/GPU clusters, SQL. *Collaboration*: Jira, Git, Slack, Overleaf, GDocs. *Previous experience*: C++, Java/Android, PHP, Javascript, R, D3.js, Plotly, Node.js, Google Cloud ML, Spark, NetworkX, Igraph, Gephi, Matlab, NLTK, Gensim, PyMC.

— Language certifications

English (C2), Spanish (C1), German (B2), Greek (native)