Sharon Gannot, Zheng-Hua Tan, Martin Haardt, Nancy F. Chen, Hoi-To Wai,
Ivan Tashev, Walter Kellermann, and Justin Dauwels

Data Science Education: The Signal Processing Perspective

In the last decade, the signal processing (SP) community has witnessed a paradigm shift from model-based to data-driven methods. Machine learning (ML)—more specifically, deep learning—methodologies are nowadays widely used in all SP fields, e.g., audio, speech, image, video, multimedia, and multimodal/multisensor processing, to name a few. Many data-driven methods also incorporate domain knowledge to improve problem modeling, especially when computational burden, training data scarceness, and memory size are important constraints.

Data science (DS), as a research field, emerged from several scientific disciplines, namely, mathematics (mainly statistics and optimization), computer science, electrical engineering (primarily SP), industrial engineering, biomedical engineering, and information technology. Each discipline offers an independent teaching program in its core domain with a segment dedicated to DS studies. In recent years, numerous institutes worldwide have started to provide dedicated and comprehensive DS teaching programs with diverse applications.

Motivation and significance

We believe that there is a unique SP perspective of DS that should be reflected in the education given to our students. Moreover, we think that now is the right time to start defining our needs and inspirations that will reflect the direction the field of SP will take in years to come.

In this article, following a successful panel at IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2022) held in Singapore, we focus on these education aspects and draft a manifesto for an SP-oriented DS curriculum. We hope this article will encourage discussions among SP educators worldwide and promote new teaching programs in the field.

DS, ML, and SP: Interrelations

DS is an interdisciplinary field that can be taught from different perspectives. Indeed, DS-oriented material can be a segment of many existing teaching programs in science, technology, engineering, and mathematics. In this article, we aim at the more ambitious task of defining a complete and comprehensive teaching program in DS that takes the unique SP perspective.

To put things in context, SP is concerned with extracting information and knowledge from signals. Common SP tasks are the analysis, modification, enhancement, prediction, and synthesis of signals [see also https://signalprocessingsociety.org/volunteers/constitution (Article II)]. In parallel to the evolution of the SP methodology, we are witnessing a fast-growing interest in the field of ML. ML is not a new field of knowledge. Perhaps its most widely known definition dates back to 1959 (paraphrased from Arthur Samuel [1]): "Learning algorithms to build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so."

ML is, thus, a method of data analysis that uses algorithms to enable computer systems to identify patterns in data, learn from them, and make predictions or decisions based on that learning.

For the SP community, data come in the form of signals. While the definition of *signals* as the carriers of information remains unchanged, the variety of signal types is rapidly growing. Signals can be either 1D or multidimensional; can be defined over a regular grid (time or pixels) or on an irregular graph; can be packed as vectors, matrices, or higher dimensional tensors; and can represent multimodal data.

As discussed, a significant component of SP is dedicated to extracting, representing, and transforming (raw) data to information that accentuates certain properties beneficial to downstream tasks. While, traditionally, SP focuses on processing raw data that have a physical grounding on planet Earth [e.g., audio, speech, radar, sonar, image, video, electrocardiogram, electroencephalogram, magnetoencephalography, and econometric data], one may not need to be limited to this standard practice.

Digital Object Identifier 10.1109/MSP.2023.3294709
Date of current version: 3 November 2023

A broader and more general definition of *signals* should include semantic data. Semantic information ultimately stems from the cognitive space in the human mind, which originates from neurophysiological activities in our brains. Cognitive neuroscience is currently not advanced enough to pinpoint how to map semantic information represented in a text to brain activation. Still, this limitation does not prevent one from applying the essence of SP approaches to understanding, representing, and modeling text data or, more generally, semantic information. (Text is, ultimately, just a human-made representation for encoding language and knowledge.) Moreover, multimodal signals are jointly analyzed and processed in some modern applications. Audiovisual SP is an excellent example of two physical signals that are jointly processed. Image captioning is a good example that involves both physical signals and semantic information and, hence, should be processed using methodologies adopted from both computer vision and natural language processing (NLP) disciplines.

We, therefore, claim that the ICASSP 2020 motto, "From Sensors to Information, at the Heart of Data Science," can be further extended to all types of data: physical, which is indeed captured by sensors, as well as cognitive and semantic. The essence of the processing tasks and the underlining methods remain similar.

The principles of DS education from SP and ML perspectives

This section is dedicated to our view of the essential principles of DS education. Among other topics, we highlight the importance of SP and ML in the DS discipline.

SP and ML methods

Traditionally, we may think of two complementary lists of DS methods stemming from the SP and ML disciplines. A noncomprehensive list can include the following:

- *SP*: convolution, time-frequency analysis (Fourier transform and wavelets), linear systems, state-space representations, the fusion of modalities and sensors, Wiener and Kalman filters, and graph SP.
- *ML*: (variational) expectation maximization, deep learning, reinforcement learning, end-to-end processing, attention, transformers, graphical neural networks, generative models, dimensionality reduction, kernel methods, subspace, and manifold learning.

This dichotomy between the lists is rather artificial. Recent trends have shown that these two paradigms are converging and are now strongly interrelated by routinely borrowing ideas and practices from each other.

We believe that modern teaching programs should, therefore, emphasize the SP and ML aspects of DS without sacrificing other essential and fundamental elements, namely, optimization, statistics, linear algebra, multilinear (tensor) algebra, artificial intelligence (AI), algorithms, data handling, transmission and storage, and programming skills.

From the SP perspective, rigorous training in DS should bring students to think more fundamentally about where the data at hand come from; what the data points and distributions represent; and how to model, sample, represent, and visualize such information robustly so that it is insensitive to various sources and types of noise for different applications and tasks.

Just as important as the technical skills, students must become aware of ethical issues related to DS, e.g., the privacy of the data, biases in collecting the data, and the implications their future techniques and developments might impose on society and humanity.

Teaching methodologies

All modern teaching programs—and, perhaps, specifically DS teaching programs—should give special attention to teaching methodologies that can be more relevant and attractive to the younger generation of students. While we certainly do not claim that "traditional" teaching methods—namely, a teacher lecturing in front of a class—should be abandoned, we encourage educators to incorporate diverse teaching techniques in their curricula. A nonexhaustive list of teaching methodologies may include online courses, labs with interactive programming exercises, flipped classrooms, and hands-on experience that may involve projects and teamwork. As the DS discipline is vast and cannot usually be fully covered by one institute, we encourage educators to consider student exchange programs and joint programs between universities (especially with other countries) and to include internships in the industry. Needless to say, science has no borders, and students will greatly benefit from learning in different schools and listening to many points of view from world-leading experts in their respective disciplines.

Learning outcomes

The graduates of the program are expected to master the theory and practice (including programming skills) of modern and classical SP and ML tools for handling various types of data, most notably, data that originate from signals. They are also expected to thoroughly understand the field's underlying mathematical and statistical foundations as well as related fields, e.g., data handling, storage (databases and clouds), and transmission (over the network), including reliability and privacy preservation. With rigorous training in SP and ML, graduates will be able to identify and apply the correct tools for DS problems. Graduates should specialize in several advanced topics in the general field of DS and become acquainted with several domain-specific applications. Graduates will, thus, be able to address complex DS problems considering ethical aspects and the sustainability of our global environment.

DS undergraduate curriculum: A proposal from the SP perspective

In this section, we draft a proposed curriculum for DS studies from the SP perspective. We are, of course, aware of the different education systems around the world. Nevertheless, we hope that such a list can serve as a source of inspiration to educators and policy leaders in academic institutes.

In the following, we propose a four-year program (in Europe, it is common to have three years of undergraduate studies plus two years of graduate studies) comprising three layers:

- *mandatory*: a strong background in math and statistics, hands-on programming skills, basic data handling and AI, SP and ML, and ethics
- *elective tracks*: data sharing and communication over networks; advanced algorithms and optimization; security, reliability, and privacy preservation; and ML and DL hardware and software tools
- *DS applications*: in diverse domains.

We next discuss each layer in detail and give a list of relevant courses. Naturally, each institute will pave its own way toward the most suitable curriculum.

Mandatory areas (with lists of proposed courses)

We believe that each student should be extensively exposed to the field's theoretical foundations and develop basic hands-on and programming skills:

- *mathematics*: calculus, linear algebra, combinatorics, set theory and logic, harmonic analysis, differential equations (regular and partial), numerical analysis, numerical algebra, multilinear algebra, algebraic structures, optimization, and complex functions
- *statistics*: probability theory, statistics, random processes, information theory, parameter estimation, and statistical theory
- *computer skills and algorithms*: programming basics, data structures and algorithms, Python (including libraries and packages—PyTorch, NumPy, SciPy, and more), object-oriented programming, computer architecture, computability, and cloud computing
- *hands-on*: labs and tools as well as annual projects with real data
- *data handling and AI*: introduction to DS (including the data processing cycle), meetings with industry (R&D in DS, ethics, practical and real-world problems, and needs), data analysis and visualization, data mining, data representations, and introduction to AI
- *SP and ML*: representations and types of signals and systems, SP in the time-frequency domain (Fourier and wavelet transform, filter banks), ML and pattern recognition, statistical algorithms in SP, statistical and model-based algorithms in ML, adaptive SP, generative models, supervised and unsupervised learning, deep learning, time series and sequences analysis and processing, graphical models, and ML operations
- *ethics*: ethical and legal aspects of DS, explainability, General Data Protection Regulation, bias, privacy, and approval processes.

Elective specialization tracks (with lists of proposed courses)

Students should elect courses from two or three specialization tracks to advance their knowledge in the field. Specialization tracks may include advanced SP, ML, and optimization algorithms (we split the SP and ML courses into two lists: basic materials and an elective specialization track); dedicated DS-related hardware and software tools; and data sharing and storing methodologies considering security and privacy preservation:

- *data sharing and communication over networks*: detection theory, communication, wireless communication, ML for communications, computer networks, mathematical analysis of networks, social networks, cloud data handling, and federated learning
- *advanced algorithms and optimization*: online algorithms, advanced algorithms, streaming algorithms, big data, quantum learning, graph theory, advanced databases, game theory, deterministic and stochastic methods in operations research, analysis and mining of processes, distributed computation, and cloud computing
- *advances in SP and ML*: array SP, blind source separation and independent component analysis, data fusion (multiple sensors/modalities), reinforcement learning, distributed processing over networks (federated learning), graph SP, and graph neural networks
- *Security, reliability, and privacy preservation*: coding, cryptography, privacy-preserving computing and communications, safe computing, and anomaly detection
- *ML and DL hardware and software tools*: digital signal processors; field programmable gate arrays; CPUs; GPUs; neuromorphic processing systems; parallel computing architectures; parallel computing platform and application programming interface (CUDA); and Python, C, and C++ computer languages.

Domain-specific DS applications

This track offers a noncomprehensive list of courses in knowledge domains that extensively apply DS tools. Students are encouraged to learn several courses from this list to become acquainted with real-life applications: econometrics, business intelligence, smart cities, blockchain and cryptocurrency, electro-optics, materials, bioinformatics, AI in health care and medical data mining, biomedical SP, DS in brain imaging, audio/speech analysis and processing, music SP and music information retrieval, NLP, image processing and computer vision, computer graphics, wireless communications, and autonomous vehicles. [Students are required to choose only a small number of courses (e.g., three or four) from the list to become acquainted with several domains that apply DS methodologies.]

Summary and further reading

In this article, we proposed a DS curriculum focusing on SP and ML. We believe such a program can be relevant to many educators and researchers in the IEEE Signal Processing Society. This article follows a panel held at ICASSP'22 [2].

There have been several attempts to define the DS discipline and the required curriculum for a major in DS. Interested readers may refer to recent reports by the U.S. National Academies of Sciences, Engineering, and Medicine [3]; Park City Math Institute [4]; and Israeli Academy of Sciences and Humanities [5]. An overview of the history of DS, its prospective future, and some guidelines for educating in the discipline can be found in [6]. All these references address the DS discipline in general. In our article, we attempt to focus on the SP and ML perspectives. Readers are also referred to an interesting discussion between Prof. Alfred Hero and Prof. Anders Lindquist about the impact of ML on SP and control systems, which can be found online [7].

Several institutes worldwide already offer study programs in DS with an SP flavor. A nonexhaustive list of study programs follows. The electrical and computer engineering faculty at the University of Michigan offers an ML curriculum [8]. A new undergraduate program proposed by Bar-Ilan University, Israel, follows the guidelines proposed in this article. This program will be opened in the 2023–2024 academic year. (The full program in Hebrew can be found in [9].) A recent presentation on AI curriculum [10] is exploring several AI and DS teaching programs at both the undergraduate and graduate levels, including at the Technical University of Denmark [11], Carnegie Mellon [12], and the Massachusetts Institute of Technology [13]. Friedrich-Alexander University Erlangen-Nuremberg offers an elite M.Sc. degree program in advanced SP and communications [14]. The corresponding M.Sc. degree program in communications and SP has been taught at Ilmenau University of Technology since 2009 [15], and a similar M.Sc. degree program in signals and systems at Delft University of Technology [16]. While not attempting to be exhaustive, this list demonstrates the broad interest of leading academic institutes in developing study programs in SP-oriented DS for both the undergraduate and graduate levels.

The authors of this article hope that the ideas and guidelines presented here can inspire DS and SP educators to develop new teaching programs in this fascinating field.

Acknowledgment

The authors are grateful to Prof. Mor Peleg from the University of Haifa, Israel, for fruitful discussions and for drawing our attention to some of the references listed in the article as well as to Dr. Ran Gelles for fruitful discussions about the new undergraduate program proposed by Bar-Ilan University, Israel.

Authors

Sharon Gannot (sharon.gannot@biu.ac.il) received his Ph.D. degree in electrical engineering from Tel-Aviv Universty, Israel, in 2000. He is a professor with the Faculty of Engineering, Bar-Ilan University, Ramat-Gan 5290002, Israel, where he heads the data science program; he also serves as the faculty vice dean and served as the deputy director of the Data Science Institute. He served as the chair of the Audio and Acoustic Signal Processing Technical Committee. He will be the general cochair of Interspeech, to be held in Jerusalem in 2024; currently, he is the chair of the IEEE Signal Processing Society (SPS) Data Science Initiative and a member of the SPS Education Center Editorial Board, and EURASIP Signal Processing for Multisensor Systems TAC. He also serves as associate editor and senior area chair for several journals. He is a Fellow of IEEE.

Zheng-Hua Tan (zt@es.aau.dk) received his Ph.D. degree in electronic engineering from Shanghai Jiao Tong University. He is a professor in the Department of Electronic Systems, the Machine Learning Research Group leader, and a cohead of the Centre for Acoustic Signal Processing Research at Aalborg University, 9220 Aalborg, Denmark, as well as a colead of the Pioneer Centre for AI, Denmark. He is an associate editor for the *IEEE Journal of Selected Topics in Signal Processing* inaugural special series on "Artificial Intelligence in Signal and Data Science—Toward Explainable, Reliable, and Sustainable Machine Learning." He is a TPC vice chair for ICASSP 2024 and was the general chair for IEEE MLSP 2018 and a TPC cochair for IEEE SLT 2016. His work has been recognized by the prestigious IEEE Signal Processing Society 2022 Best Paper Award. His research interests include deep representation learning. He is a Senior Member of IEEE.

Martin Haardt (martin.haardt@tu-ilmenau.de) received his Doktor-Ingenieur (Ph.D.) degree from Munich University of Technology in 1996. He has been a full professor in the Department of Electrical Engineering and Information Technology and head of the Communications Research Laboratory at Ilmenau University of Technology, 98684 Ilmenau, Germany, since 2001. He received the 2009 Best Paper Award from the IEEE Signal Processing Society; the Vodafone (formerly Mannesmann Mobilfunk) Innovations Award for outstanding research in mobile communications; the ITG Best Paper Award from the Association of Electrical Engineering, Electronics, and Information Technology; and the Rohde & Schwarz Outstanding Dissertation Award. He has served as a senior editor for *IEEE Journal of Selected Topics in Signal Processing* since 2019. His research interests include wireless communications, array signal processing, high-resolution parameter estimation, and tensor-based signal processing. He is a Fellow of IEEE.

Nancy F. Chen (nfychen@i2r.a-star.edu.sg) received her Ph.D. degree in biomedical engineering from the Massachusetts Institute of Technology and Harvard in 2011. She is a Fellow, Senior Principal Scientist, Principal Investigator and Group Leader at the Institute for Infocomm Research and Centre for Frontier AI Research, Agency for Science, Technology, and Research (A*STAR), Singapore, 138632. She leads research efforts in generative artificial intelligence with a focus on speech language technology with applications in education, healthcare, and defense. Speech evaluation technology from her team has been deployed at the Ministry of Education in Singapore to support home-based learning and led to commercial spin-offs. She has received numerous awards, including being named among the Singapore 100 Women in Tech in 2021, the Young Scientist Best Paper Award at MICCAI 2021, the Best Paper Award at SIGDIAL 2021, and the 2020 P&G Connect + Develop Open Innovation Award. She is currently the program chair of the International

Conference on Learning Representations (ICLR), IEEE Distinguished Lecturer, a Board Member of the International Speech Communication Association (ISCA), and a Senior Area Editor of *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. She is a Senior Member of IEEE.

Hoi-To Wai (htwai@cuhk.edu.hk) received his Ph.D. degree from Arizona State University (ASU) in electrical engineering in 2017. He is an assistant professor with the Department of Systems Engineering and Engineering Management at the Chinese University of Hong Kong, Hong Kong, China. His research interests include signal processing, machine learning, and distributed optimization, with a focus on their applications to network science. His dissertation received the 2017 Dean's Dissertation Award from the Ira A. Fulton Schools of Engineering at ASU, and he was a recipient of a Best Student Paper Award at ICASSP 2018. He is a Member of IEEE.

Ivan Tashev (ivantash@microsoft.com) received his Ph.D. degree in computer science from the Technical University of Sofia, Bulgaria, in 1990. He is a partner software architect and leads the Audio and Acoustics Research Group in Microsoft Research, Redmond, WA 98052 USA; is an affiliate professor at the University of Washington in Seattle; and is an honorary professor at the Technical University of Sofia, Bulgaria. He also coordinates the Brain–Computer Interfaces project in Microsoft Research. He has published two books, two book chapters, and more than 100 scientific papers, and he is listed as an inventor for 50 U.S. patents. His research interests include audio signal processing, machine learning, multichannel transducers, and biosignal processing. He is a member of the Audio Engineering Society and the Acoustical Society of America and a Fellow of IEEE.

Walter Kellermann (walter.kellermann@fau.de) received his Dr.-Ing. degree in electrical engineering from Technical University Darmstadt, Germany, in 1988. He is a professor of communications at the University of Erlangen-Nuremberg, 91058 Erlangen, Germany. His service to the IEEE Signal Processing Society includes Distinguished Lecturer (2007–2008), Chair of the Technical Committee for Audio and Acoustic Signal Processing (2008–2010), Member of the IEEE James L. Flanagan Award Committee (2011–2014), Member at Large SPS Board of Governors (2013–2015), Vice President Technical Directions (2016-2018), Member SPS Nominations Appointments Committee (2019–2022), and Member of the SPS Fellow Evaluation Committee (2023–). He has served as the general chair of eight mostly IEEE-sponsored workshops and conferences. He is a corecipient of 10 best paper awards, was awarded the Julius von Haast Fellowship by the Royal Society of New Zealand in 2012, and received the Group Technical Achievement Award of the European Association for Signal Processing (EURASIP) in 2015. His research interests include speech signal processing, array signal processing, and machine learning, especially for acoustic signal processing. He is a fellow of EURASIP and a Life Fellow of IEEE.

Justin Dauwels (j.h.g.dauwels@tudelft.nl) received his Ph.D. degree in electrical engineering from the Swiss Polytechnical Institute of Technology in Zurich in 2005. He is an associate professor of signal processing systems, Department of Microelectronics, Delft University of Technology, 2628 CD Delft, The Netherlands. He is an associate editor of *IEEE Transactions on Signal Processing*, an associate editor of the Elsevier journal *Signal Processing*, a member of the editorial advisory board of *International Journal of Neural Systems*, and an organizer of IEEE conferences and special sessions. His research team has won several best paper awards at international conferences and from journals. His research interests include data analytics with applications to intelligent transportation systems, autonomous systems, and analysis of human behavior and physiology.

References

[1] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM J. Res. Develop.*, vol. 3, no. 3, pp. 210–229, Jul. 1959, doi: 10.1147/rd.33.0210.

[2] S. Gannot, Z.-H. Tan, M. Haardt, N. F. Chen, H.-T. Wai, and I. Teshev, "Data science education: The signal processing perspective," *Panel at IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)* May 2022. [Online]. Available: https://rc.signalprocessingsociety.org/conferences/icassp-2022/SPSICASSP22VID1984.html?source=IBP

[3] *Envisioning the Data Science Discipline: The Undergraduate Perspective*. Washington, DC, USA: National Academy Press, 2018.

[4] R. D. De Veaux et al., "Curriculum guidelines for undergraduate programs in data science," *Annu. Rev. Statist. Appl.*, vol. 4, no. 1, pp. 15–30, Mar. 2017, doi: 10.1146/annurev-statistics-060116-053930.

[5] N. Ahituv, J. Ben-Dov, Y. Benjamini, Y. Bronner, Y. Dudai, D. Raban, and R. Sharan, *Teaching Data Science in Universities in All Disciplines*. Jerusalem, Israel: Israel Academy of Sciences and Humanities, 2020. [Online]. Available: https://www.academy.ac.il/SystemFiles2015/2-1-21-English.pdf

[6] D. Donoho, "50 years of data science," *J. Comput. Graphical Statist.*, vol. 26, no. 4, pp. 745–766, Aug. 2017, doi: 10.1080/10618600.2017.1384734.

[7] C. June, "Machine learning and systems: A conversation with 2020 Field Award winners Alfred Hero and Anders Lindquist," Elect. Comput. Eng., Univ. of Michigan, Ann Arbor, MI, USA, Oct. 2019. [Online]. Available: https://ece.engin.umich.edu/stories/machine-learning-and-systems-a-conversation-with-2020-field-award-winners-al-hero-and-anders-lindquist

[8] C. June, "Teaching machine learning in ECE," Elect. Comput. Eng., Univ. of Michigan, Ann Arbor, MI, USA, Mar. 2022. [Online]. Available: https://ece.engin.umich.edu/stories/teaching-machine-learning-in-ece

[9] "Data engineering - Bachelor's degree," Faculty Eng., Bar-Ilan Univ., Ramat Gan, Israel, 2023. [Online]. Available: https://engineering.biu.ac.il/datascience

[10] Z.-H. Tan, "On artificial intelligence curriculum and problem-based learning [Slides]," Aalborg Univ., Aalborg, Denmark, 2021. [Online]. Available: https://people.es.aau.dk/~zt/online/AI-curriculum-Tan.pdf

[11] [Online]. Available: https://www.dtu.dk/english/education/undergraduate/undergraduate-programmes-in-danish/bsc-eng-programmes/artificial-intelligence-and-data

[12] "B.S. in artificial intelligence," School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, 2023. [Online]. Available: https://www.cs.cmu.edu/bs-in-artificial-intelligence/

[13] "Interdisciplinary programs," Massachusetts Inst. Technol., Cambridge, MA, 2022–2023 USA. [Online]. Available: http://catalog.mit.edu/interdisciplinary/undergraduate-programs/degrees/

[14] "Elite master's study programme: Advanced signal processing and communications engineering," Inst. Digit. Commun., Erlangen, Germany, 2023. [Online]. Available: https://www.asc.studium.fau.de

[15] "Master of science in communications and signal processing," TU Ilmenau, Ilmenau, Germany, 2023. [Online]. Available: https://www.tu-ilmenau.de/mscsp

[16] "Track: Signals & systems," TU Delft, Delft, The Netherlands, 2023. [Online]. Available: https://www.tudelft.nl/en/education/programmes/masters/electrical-engineering/msc-electrical-engineering/track-signals-systems

SP