**The Faculty of Medicine of Harvard University**

**Curriculum Vitae**

|  |  |
| --- | --- |
| **Date Prepared:** | **05/14/2021** |
| **Name:** | **William La Cava** |
| **Office Address:** | **Richards Building D207**  **3700 Hamilton Walk**  **Philadelphia, PA 19104** |
| **Home Address:** | **4406 Pine St Apt 4**  **Philadelphia, PA 19104** |
| **Work Phone:** | **413-320-0544** |
| **Work Email:** | [**lacava@upenn.edu**](mailto:lacava@upenn.edu) |
| **Place of Birth:** | Euclid, OH, USA |

Education:

|  |  |  |  |
| --- | --- | --- | --- |
| Month/Year(s) | Degrees (Honors) | Fields of Study  (Thesis advisor for doctoral research degrees) | Institution |
| 08/05 – 05/09 | B.S. | Mechanical Engineering | Cornell University |
| 08/09 – 05/10 | M.Eng. | Mechanical Engineering | Cornell University |
| 08/12 – 08/16 | Ph.D. | Mechanical Engineering (Kourosh Danai) | Cornell University |

Postdoctoral Training:

|  |  |  |  |
| --- | --- | --- | --- |
| Month/Year(s) | Title | Specialty/Discipline  (Primary mentor/PI, if relevant) | Institution |
| 08/16 – 12/19 | Postdoctoral Fellow | Biomedical Informatics (Jason H. Moore) | University of Pennsylvania |
| 01/20 – | Research Associate | Biomedical Informatics (Jason H. Moore) | University of Pennsylvania |

Other Professional Positions:

|  |  |  |  |
| --- | --- | --- | --- |
| Year(s) | Position Title | Institution | Level of effort  (current roles only) |
| 2015 – 2015 | Visiting Research Fellow | University of Lisbon |  |
| 2010 – 2012 | Research Scientist | National Renewable Energy Laboratory |  |
| 2019 – 2020 | Consultant | National Institute on Aging |  |

Professional Societies:

|  |  |  |
| --- | --- | --- |
| Year(s) | Society Name |  |
|  | Dates of Role(s) | Title of Role(s) |
| 2011 - 2012 | Gearbox Reliability Collaborative, National Renewable Energy Laboratory | Lead organizer |
| 2012 - 2014 | American Institute of Aeronautics and Astronautics (AAAI) | Member |
| 2010 – 2016 | American Society of Mechanical Engineers (ASME) | Member |
| 2014 - | Association of Computing Machinery (ACM) | Member |
| 2014 - | Special Interest Group on Genetic and Evolutionary Computation (SIGEVO) | Member |
| 2016 - | International Society for Computational Biology (ISCB) | Member |

Editorial Activities:

* Ad hoc Reviewer

|  |
| --- |
| List journals for which you serve as a reviewer |

|  |
| --- |
| *AAAI*  *AIAA Wind Energy Symposium (2014)*  *AMIA Annual Symposium*  *BioData Mining*  *ASME Dynamic Systems and Controls Conference (2015)*  *Pacific Symposium on Biocomputing*  *European Conference on Genetic Programming*  *Artificial Life Journal*  *Genetic and Evolutionary Computation Conference*  *Genetic Programming and Evolvable Machines*  *Genetic Programming Theory and Practice*  *EMNLP*  *IEEE Congress on Evolutionary Computation*  *IEEE Control Systems Letter*  *IEEE Transactions on Neural Networks and Learning Systems*  *Information Journal*  *JAMIA*  *Neurips*  *npj Digital Medicine*  *PLOS ONE*  *Renewable Energy Journal*  *Swarm and Evolutionary Computation*  *Wind Energy Journal* |

* Other Editorial Roles

|  |  |  |
| --- | --- | --- |
| Year(s) | Role | Journal/Publication Title |

|  |  |  |
| --- | --- | --- |
| 2017-2019 | Co-editor and organizer | New Standards for Benchmarking in Evolutionary Computation Research (GECCO Workshop) |
| 2020 | Co-editor and organizer | Good Benchmarking Practices for Evolutionary Computation (GECCO Workshop) |

Honors and Prizes:

|  |  |  |  |
| --- | --- | --- | --- |
| Year | Name of Honor/Prize | Awarding Organization | Achievement for which awarded  (if unclear from award title) |

|  |  |  |  |
| --- | --- | --- | --- |
| 2011 | First Place | Green Grand Prix Competition | Design and built a 100+ MPG, four-person sedan as member of the Cornell 100 MPG team |
| 2015 | Best Paper Nomination | Genetic and Evolutionary Computation Conference |  |
| 2017 | Best Paper Nomination | European Conference on Genetic Programming |  |
| 2018 | Best Paper Nomination | Genetic and Evolutionary Computation Conference |  |
| 2019 | Best Informatics Abstract | UPenn Department of Biostatistics, Epidemiology and Informatics |  |
| 2019 | Best Paper Nomination | Genetic and Evolutionary Computation Conference |  |
| 2020 | Best Paper Award | Genetic and Evolutionary Computation Conference |  |

**Report of Funded and Unfunded Projects**

Past

|  |  |
| --- | --- |
| Year(s) | Grant title |
|  | Funding Agency, Grant type and Grant number |
|  | Official role on project |
|  | Description of the major goals |

|  |  |
| --- | --- |
| 2012 – 2016 | IGERT: Offshore Wind Energy Engineering, Environmental Science, and Policy |
|  | NSF |
|  | Role: Graduate Fellow $285,000 |
|  | Interdisciplinary program that trains graduate students in the technology, environmental implications, and social/economic/regulatory challenges of offshore wind farms. |
| 2014 – 2016 | Extreme Science and Engineering Discovery Environment (XSEDE) Startup Allocation Award |
|  | Role: PI (50,000 compute hours) |
|  | Using machine learning to identify dynamic models of complex systems |
| 2016 – 2017 | Postdoctoral Fellowship |
|  | Warren Center for Network and Data Sciences, University of Pennsylvania |
|  | Role: Fellow $30,000 |
|  | Using graph database methods to understand problems in computational genetics |

Current

|  |  |
| --- | --- |
| Year(s) | Grant title |
|  | Funding Agency, Grant type and Grant number |
|  | Official role on Project (if PI or Site PI, report total direct costs for all years) |
|  | Description of the major goals |

|  |  |
| --- | --- |
| 2019 - | Multi-objective representation learning methods for interpretable predictions of patient outcomes using electronic health records |
|  | National Library of Medicine / NIH K99 LM012926 |
|  | Role: PI $179,000 |

**Projects Submitted for Funding**

|  |  |
| --- | --- |
| Status | Grant title |
| Grant type and number |
|  | Role on Project (if in PI role, list requested direct costs |
|  | Description of the major goals (include preliminary scores or review date if known) |

|  |  |
| --- | --- |
| Awarded; | Development of methods to improve identification of patients with rare or complex diseases |
| starting | PCORI Methods to Improve Patient-Centered Outcomes Research |
| 07/21 | Role: Co-Investigator (Other PI: Herman) $45,000 |
|  |  |
|  |  |
| In process; start date | Multi-objective representation learning methods for interpretable predictions of patient outcomes using electronic health records |
| 09/21 | National Library of Medicine / NIH R00 (Independent phase of K99/R00) |
|  | Role: PI $747,000 |

Unfunded Current Projects

|  |  |
| --- | --- |
| Year(s) | Title of Project |
|  | Role |
|  | One sentence description of the purpose of the project |

|  |  |
| --- | --- |
| 2020 - | Predictive modeling of CAR T-Cell immunotherapy outcomes |
|  | Co-Investigator |
|  | Using machine learning to understand the role of the product manufacturing process on clinical outcomes of CAR T-cell immunotherapy |
| 2021 - | Improving fetal heart rate monitoring using deep learning |
|  | Co-Investigator |
|  | Developing machine learning approaches to improve the success of interventions that utilize fetal heart rate signatures |

**Report of Local Teaching and Training**

Teaching of Students in Courses:

|  |  |  |
| --- | --- | --- |
| Year(s) | Course Title | Location |
|  | Type of Student/Audience | Level of Effort |

|  |  |  |
| --- | --- | --- |
| 2014 – 2015 | Control Systems Laboratory | University of Massachusetts Amherst |
|  | Senior undergraduate engineering students | Two hour sessions, 4 times per semester |
|  |  |  |
| 2015 | System Dynamics | University of Massachusetts Amherst |
|  | Senior undergraduate engineering students | Two 1.5-hour lectures |
|  |  |  |
| 2016 | Offshore Wind Energy Design | University of Massachusetts Amherst |
|  | Graduate engineering students | 1.5-hr lecture |
|  |  |  |
| 2017 – 2020 | Data Science for Biomedical Informatics | University of Pennsylvania |
|  | Masters and CE students in informatics | Intro to Machine Learning module; two 1.5-hr lectures per semester, plus final project advising |
|  |  |  |
| 2020 | AI III: Advanced Methods and Health Applications in Machine Learning | University of Pennsylvania |
|  |  | 1.5-hr lecture |
|  |  |  |
| 2018 – 2021 | Special Topics in Biomedical and Health Informatics | University of Pennsylvania |
|  | Masters and CE students in informatics | Module on Bio-inspired Computing (2018-2020); Module on Fairness and Ethics in Medicine (2021). 2 3-hour sessions per semester |
|  |  |  |
| 2020 – 2021 | Foundations of Artificial Intelligence | University of Pennsylvania |
|  | Masters and CE students in informatics | 1.5-hr lecture per semester |

Formal Teaching of Residents, Clinical Fellows and Research Fellows (post-docs):

|  |  |  |
| --- | --- | --- |
| Year(s) | Title | Location |
|  | Type of Trainee/Audience | Level of Effort |

|  |  |  |
| --- | --- | --- |
| 2021 | Predicting COVID-19 mortality with electronic medical records | COVID-19 Population Sciences Journal Club |
|  | Research Fellows | Led 30-minute discussion portion following presentation by trainee |
| 2021 | CHIP Fellows Didactic Session | Boston Children’s Hospital |
|  | Research fellows | One hour lecture and discussion |

Research Supervisory and Training Responsibilities:

|  |  |  |
| --- | --- | --- |
| Year(s) | Type of Responsibility | Location |
|  | Type of Trainee/Audience | Level of Effort |

|  |  |  |
| --- | --- | --- |
| 2015 | NSF Research Experiences for Undergraduates (REU) Mentor | University of Massachusetts Amherst  2 hours per week during the summer |

**Mentored Trainees and Faculty:**

|  |  |
| --- | --- |
| Year(s) | Names and degrees / Current position, Institution |
|  | Note the mentee’s career stage during the mentorship period and your mentoring role. Please describe the accomplishments of your mentee as a direct result of your mentorship (maximum one sentence) |

*Note: these are students I mentored for research projects as a postdoc and/or research associate at UPenn6.*

|  |  |
| --- | --- |
| 2017 | Rishabh Gupta, M.S. EMBS / Bloomberg |
| 2018 | Sophia Moses, B.S. SEAS / current UPenn student |
| 2018-2019 | Tilak Raj Singh, M.S. SEAS / Microsoft |
| 2019 | James Taggart, B.S. CIS / Google |
| 2018 | Srinivas Suri, M.S. SEAS / Microsoft |
| 2018-2019 | Max Roling, B.S. Wharton / Morgan Stanley |
| 2018 | Efe Ayhan, B.S. CSE / current UPenn student |
| 2019-2020 | Isabel Lee, B.S. SEAS / current UPenn student |
| 2019 | Saurav Bose, M.S. SEAS / Children’s Hospital of Philadelphia |
| 2019-2021 | Nupur Baghel, M.S. SEAS / Facebook |

Local Invited Presentations:

|  |
| --- |
| Check the **single** most appropriate statement below (double click the chosen box and change the default value to “checked”) |

*No presentations below were sponsored by 3rd parties/outside entities*

*Those presentations below sponsored by outside entities are so noted and the sponsor(s) is (are) identified.*

|  |  |
| --- | --- |
| Year(s) | Title of presentation / Type of presentation |
|  | Department and Institution where presented (Sponsor, if any) |

|  |  |
| --- | --- |
| 2015 | Genetic programming with epigenetic local search / invited talk |
|  | Laboratory of Agent Modeling, University of Lisbon |
| 2017 | Multidimensional Feature Learning for Biomedical Classification / invited talk |
|  | Penn Genetics and Computational Biology Retreat, College of Physicians |
| 2019 | Towards Automated Feature Engineering and Automated Data Science / invited talk |
|  | Wharton Undergraduate Data Analytics Club |
| 2019 | Learning Features from Longitudinal Data / invited talk |
|  | Research Day: Penn Department of Biostatistics, Epidemiology and Informatics |
| 2021 | Ethics of AI in Medicine / invited panelist |
|  | Penn Medicine |
| 2021 | K Award Panel Discussion / invited panelist |
|  | Biomedical Postdoctoral Program, University of Pennsylvania |

**Report of Regional, National and International Invited Teaching and Presentations**

|  |
| --- |
| Check the **single** most appropriate statement below (double click the chosen box and change the default value to “checked”) |

*No presentations below were sponsored by 3rd parties/outside entities*

*Those presentations below sponsored by outside entities are so noted and the sponsor(s) is (are) identified.*

Regional

|  |  |
| --- | --- |
| Year(s) | Title of presentation or name of course / Type of presentation/role(s)  (note if presentation was the result of a selected abstract) |
|  | Location (Sponsor, if any) |

|  |  |
| --- | --- |
| 2015 | Developing compact nonlinear dynamic models with biologically inspired algorithms / invited talk |
|  | CSAIL, Massachusetts Institute of Technology |
|  | Cambridge, MA |
| 2019 | Automating Data Science for Biomedical Informatics / invited talk |
|  | Bioinformatics Studio, Temple University |
| 2019 | Automated data integration approaches for multi-omics analysis / invited talk |
|  | National Institute on Aging Winter Meeting |
| 2020 | Transformative AI-based strategies to identify determinants of exceptional health and life span / invited talk |
|  | National Institute on Aging Summer Retreat |

National

|  |  |
| --- | --- |
| Year(s) | Title of presentation or name of course / Type of presentation/role(s)  (note if presentation was the result of a selected abstract) |
|  | Location (Sponsor, if any) |

|  |  |
| --- | --- |
| 2012 | Gearbox reliability collaborative: test and model investigation of sun orbit and planet load share in a wind turbine gearbox / conference presentation |
|  | AIAA 53rd Structures, Structural Dynamics, and Materials and Colocated Conferences |
|  | Honolulu, HI |
| 2015 | Model Structure Adaptation: A Gradient-based Approach / conference presentation |
|  | ASME 2015 Dynamic Systems and Control Conference |
|  | Columbus, OH |
| 2015 | Automated Identification of Closed-Loop Wind Turbine Dynamics via Genetic Programming / conference presentation |
|  | ASME 2015 Dynamic Systems and Control Conference |
|  | Columbus, OH |
| 2015 | Intelligible system modeling with applications to wind energy (and genomics!) / invited talk |
|  | Institute for Biomedical Informatics, University of Pennsylvania |
|  | Philadelphia, PA |
| 2016 | Epsilon-Lexicase Selection for Regression / conference presentation |
|  | Genetic and Evolutionary Computation Conference |
|  | Denver, CO |
| 2017 | Symbolic Representation Learning / invited talk |
|  | EDGE Workshop |
|  | Key West, FL |
| 2017 | Data-driven Advice for Applying Machine Learning to Bioinformatics Problems / conference presentation |
|  | Pacific Symposium on Biocomputing (PSB) |
|  | Big Island, HI |
| 2020 | Randomized search heuristics for learning fair classifiers / invited lunch talk |
|  | Department of Biomedical Informatics, Harvard |
|  | Boston, MA |
| 2020 | Learning interpretable models from biomedical data / invited talk |
|  | Northeastern Department of Bioengineering |
|  | Boston, MA |
| 2020 | Learning interpretable models from biomedical data / invited talk |
|  | Mt. Sinai |
|  | New York, NY |
| 2020 | Automated data integration approaches for multi-omics analysis / invited lecture |
|  | Computational and Functional Genomics Group, National Institute on Aging |
|  | Washington, D.C. (online) |
| 2021 | Learning interpretable and fair models for predictive health / invited talk |
|  | University of Massachusetts Boston |
|  | Boston, MA (online) |
| 2021 | Learning interpretable and fair models for predictive health / invited talk |
|  | Boston Children’s Hospital |
|  | Boston, MA (online) |

International

|  |  |
| --- | --- |
| Year(s) | Title of presentation or name of course / Type of presentation/role(s)  (note if presentation was the result of a selected abstract) |
|  | Location (Sponsor, if any) |

|  |  |
| --- | --- |
| 2012 | Determining wind turbine gearbox model complexity using measurement validation and cost comparison / conference presentation |
|  | European Wind Energy Association annual event |
|  | Copenhagen, DE |
| 2013 | Gearbox Reliability Collaborative: Findings from Phase 1 and 2 |
|  | Norwegian University of Science and Technology |
|  | Trondheim, Norway |
| 2015 | Genetic Programming with Epigenetic Local Search / conference presentation, best paper session |
|  | Genetic and Evolutionary Computation Conference |
|  | Madrid, Spain |
| 2017 | A general feature engineering wrapper for machine learning using ϵ-lexicase survival / conference presentation, best paper session |
|  | European Conference on Genetic Programming |
|  | Amsterdam, Netherlands |
| 2017 | Genetic Programming Representations for Multi-dimensional Feature Learning in Biomedical Classification / conference presentation |
|  | European Conference on Genetic Programming |
|  | Amsterdam, Netherlands |
| 2017 | Ensemble representation learning: an analysis of fitness and survival for wrapper-based genetic programming methods / conference presentation |
|  | Genetic and Evolutionary Computation Conference |
|  | Berlin, Germany |
| 2018 | New Standards for Benchmarking in Evolutionary Computation Research / workshop presentation |
|  | Genetic and Evolutionary Computation Conference |
|  | Kyoto, Japan |
| 2018 | Behavioral search drivers and the role of elitism in soft robotics / conference presentation |
|  | Artificial Life |
|  | Tokyo, Japan |
| 2019 | Semantic variation operators for multidimensional genetic programming / conference presentation, best paper session |
|  | Genetic and Evolutionary Computation Conference |
|  | Prague, Czech Republic |
| 2020 | Genetic programming approaches to learning fair classifiers / conference presentation |
|  | Genetic and Evolutionary Computation Conference |
|  | Presented online due to COVID-19, scheduled for Cancun, Mexico |
| 2020 | What are good benchmarking practices? / workshop presentation |
|  | Genetic and Evolutionary Computation Conference |
|  | *Presented online due to COVID-19, scheduled for Cancun, Mexico* |

**Report of Technological and Other Scientific Innovations**

|  |  |
| --- | --- |
| Innovation  (dates, if applicable) | Patent, if any, pending or awarded; if described in print or on the web, provide citation |
|  | Describe the influence or potential influence of the innovation on research or clinical care, including how the material is used locally (at HMS), regionally, nationally or internationally; if developed as a member of a team, describe your contribution. (1-3 sentences) |
| Tuned Mass Damper Module for FAST v8 (2015) | <https://nwtc.nrel.gov/tmd>  A software module that made it possible for wind researchers to simulate and develop tuned mass damper systems when designing next generation wind turbines. I was the main developer, supervised by Matthew Lackner. |
| ellynGP: a Linear Genetic Programming System for Python (2016) | <http://github.com/EpistasisLab/ellyn>  The first Sklearn-compatible genetic programming system for symbolic regression. Allows for interpretable machine learning. Sole developer. |
| FEW: a feature engineering wrapper for Scikit-learn (2016) | <http://github.com/lacava/few>  DOI: 10.5281/zenodo.205105  Allows researchers to learn symbolic features for any machine learning method. Sole developer. |
| PMLB: Penn Machine Learning Benchmarks (2017) | <http://github.com/EpistasisLab/pmlb>  Standardized, easy-to-use interface for hundreds of open source datasets. I serve as a main contributor alongside Randal Olson, Weixuan Fu, Trang Le, Joseph Romano, and other open source contributors. |
| FEAT: Feature Engineering Automation Tool (2018) | <http://github.com/lacava/feat>  The current state-of-the-art method for learning symbolic features for regression. Originally published in ICLR 2019. Sole developer. |

**Report of Education of Patients and Service to the Community**

|  |
| --- |
| Check the most appropriate statement below (double click the chosen box and change the default value to “checked”) |

*No presentations below were sponsored by 3rd parties/outside entities*

*Those presentations below sponsored by outside entities are so noted and the sponsor(s) is (are) identified.*

Activities

|  |  |
| --- | --- |
| Year(s) | Organization or institution / Role (Sponsor, if any) |
|  | One sentence description (optional) |

|  |  |
| --- | --- |
| 2013 – 2014 | Four Rivers Charter School / Invited Science Teacher |
| 2015 – 2016 | Hampshire Regional High School / Science Fair Judge |
|  | Volunteer visiting high school science teacher |

**Report of Scholarship**

ORCID ID: 0000-0002-1332-2960

Peer-Reviewed Scholarship in print or other media:

|  |
| --- |
| Group peer reviewed publication in three categories under the following headings:   * Research Investigations (full-length manuscripts that contain new data or new meta-analysis) * Other peer-reviewed publications (e.g., reviews, case reports, UpToDate and MedEdPORTAL submissions) * Scholarship without named authorship |

**Research Investigations**

1. **La Cava W**, McNiff B, van Dam J. NREL Gearbox Reliability Collaborative: Comparing In-field Gearbox Response to Different Dynamometer Test Conditions. In Anaheim, California: AWEA; 2011. Available from: <http://www.nrel.gov/docs/fy11osti/51690.pdf>

2. Guo Y, Keller J, **La Cava W**. Combined effects of gravity, bending moment, bearing clearance, and input torque on wind turbine planetary gear load sharing. In: AGMA Fall Technical Meeting. Dearborn, MI: AGMA; 2012.

3. Keller J, Link HF, Guo Y, **La Cava W**, McNiff BP. Gearbox reliability collaborative phase 1 and 2: testing and modelling results. In: Conference proceedings of ISMA2012-USD2012 [Internet]. Leuven, Belgium; 2012. Available from: <http://www.isma-isaac.be/past/conf/isma2012/proceedings/papers/isma2012_0114.pdf>

4. **La Cava W**, Keller J, McNiff B. Gearbox reliability collaborative: test and model investigation of sun orbit and planet load share in a wind turbine gearbox. In: AIAA 53rd Structures, Structural Dynamics, and Materials and Colocated Conferences [Internet]. Honolulu, Hawaii; 2012. Available from: <http://arc.aiaa.org/doi/pdf/10.2514/6.2012-1418>

5. **La Cava W**, Xing Y, Guo Y, Moan T. Determining wind turbine gearbox model complexity using measurement validation and cost comparison. In: European Wind Energy Association annual event [Internet]. Copenhagen, Denmark; 2012. Available from: <http://www.nrel.gov/docs/fy12osti/54545.pdf>

6. **La Cava W**, Guo Y, Marks C, Xing Y, Moan T. Three-dimensional bearing load share behaviour in the planetary stage of a wind turbine gearbox. IET Renewable Power Generation. 2013 Jul;7(4):359–69.

7. Guo Y, Keller J, **La Cava W**. Planetary gear load sharing of wind turbine drivetrains subjected to non-torque loads. Wind Energy. 2014 Mar;18:757–68.

8. **La Cava W**, Spector L, Danai K, Lackner M. Evolving differential equations with developmental linear genetic programming and epigenetic hill climbing. In: Companion proceedings of the 2014 conference on Genetic and Evolutionary Computation [Internet]. Vancouver, B.C.: ACM Press; 2014. p. 141–2. Available from: <http://dl.acm.org/citation.cfm?doid=2598394.2598491>

9. Guo Y, Keller J, **La Cava W**, Austin J, Nejad A, Halse C, et al. Recommendations on Model Fidelity for Wind Turbine Gearbox Simulations. In: Conference for Wind Power Drives (CWD) 2015. AAchen, Germany; 2015.

10. **La Cava W**, Danai K. Gradient-based adaptation of continuous dynamic model structures. International Journal of Systems Science. 2015 Aug;47(1):249–63.

11. **La Cava W**, Danai K. Model Structure Adaptation: A Gradient-based Approach. In: ASME 2015 Dynamic Systems and Control Conference. Columbus, Ohio: ASME; 2015.

12. **La Cava W**, Danai K, Spector L, Fleming P, Wright AD, Lackner M. Automated Identification of Closed-Loop Wind Turbine Dynamics via Genetic Programming. In: ASME 2015 Dynamic Systems and Control Conference. Columbus, Ohio: ASME; 2015.

13. **La Cava W**, Danai K, Spector L, Fleming P, Wright A, Lackner M. Automatic identification of wind turbine models using evolutionary multiobjective optimization. Renewable Energy [Internet]. 2015 Nov. Available from: <http://www.sciencedirect.com/science/article/pii/S0960148115303475>

14. **La Cava W**, Helmuth T, Spector L, Danai K. Genetic Programming with Epigenetic Local Search. In: Proceedings of the Genetic and Evolutionary Computation Conference [Internet]. Madrid, Spain: ACM Press; 2015. p. 1055–62. Available from: <http://dl.acm.org/citation.cfm?doid=2739480.2754763>

* *Nominated for Best Paper Award*

15. **La Cava W**, Danai K, Spector L. Inference of compact nonlinear dynamic models by epigenetic local search. Engineering Applications of Artificial Intelligence. 2016;55:292–306.

16. **La Cava W**, Spector L, Danai K. Epsilon-Lexicase Selection for Regression. In: Proceedings of the 2016 on Genetic and Evolutionary Computation Conference [Internet]. ACM; 2016. p. 741–8. Available from: <http://dl.acm.org/citation.cfm?id=2908898>

17. Park S, Lackner MA, Cross-Whiter J, Tsouroukdissian AR, **La Cava W**. An Investigation of Passive and Semi-Active Tuned Mass Dampers for a Tension Leg Platform Floating Offshore Wind Turbine in ULS Conditions. In: ASME 2016 35th International Conference on Ocean, Offshore and Arctic Engineering. American Society of Mechanical Engineers; 2016. p. V003T02A061-V003T02A061.

18. Rodriguez Tsouroukdissian A, Lackner M, Cross-Whiter J, Park SM, Pourazarm P, **La Cava W**, et al. Smart Novel Semi-Active Tuned Mass Damper for Fixed-Bottom and Floating Offshore Wind (Paper). In: Offshore Technology Conference. Houston, TX: U.S. DOE Office of Science and Technical Information; 2016.

19. **La Cava W**, Moore JH. A general feature engineering wrapper for machine learning using ϵ-lexicase survival. In: Genetic Programming. Springer; 2017. p. 80–95. (Lecture Notes in Computer Science).

* *Nominated for Best Paper Award*

20. **La Cava W**, Moore JH. Ensemble representation learning: an analysis of fitness and survival for wrapper-based genetic programming methods. In: GECCO ’17: Proceedings of the 2017 Genetic and Evolutionary Computation Conference. ACM; 2017.

21. **La Cava W**, Sahare K, Danai K. Restructuring Controllers to Accommodate Plant Nonlinearities. J Dyn Sys, Meas, Control. 2017;139(8):081004-081004–10.

22. **La Cava W**, Silva S, Vanneschi L, Spector L, Moore J. Genetic Programming Representations for Multi-dimensional Feature Learning in Biomedical Classification. In: Applications of Evolutionary Computation. Springer International Publishing; 2017. p. 158–73. (Lecture Notes in Computer Science; vol. 10199).

23. Olson RS, **La Cava W**, Orzechowski P, Urbanowicz RJ, Moore JH. PMLB: A Large Benchmark Suite for Machine Learning Evaluation and Comparison. BioData Mining [Internet]. 2017; Available from: <https://arxiv.org/abs/1703.00512>

24. Olson\* RS, **La Cava W**\*, Mustahsan Z, Varik A, Moore JH. Data-driven Advice for Applying Machine Learning to Bioinformatics Problems. In: Pacific Symposium on Biocomputing (PSB) [Internet]. 2017. Available from: <http://arxiv.org/abs/1708.05070>

25. **La Cava W**, Moore JH. An Analysis of epsilon-lexicase Selection for Large-scale Many-objective Optimization. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion [Internet]. ACM; 2018. p. 185–6. (GECCO ’18). Available from: <http://doi.acm.org/10.1145/3205651.3205656>

26. **La Cava W**, Moore JH. Behavioral search drivers and the role of elitism in soft robotics. In: Artificial Life [Internet]. MIT Press; 2018. p. 206–13. Available from: <https://www.mitpressjournals.org/doi/abs/10.1162/isal_a_00044>

27. **La Cava W**, Silva S, Danai K, Spector L, Vanneschi L, Moore JH. Multidimensional genetic programming for multiclass classification. Swarm and Evolutionary Computation [Internet]. 2018; Available from: <http://www.sciencedirect.com/science/article/pii/S2210650217309136>

28. Orzechowski P, **La Cava W**, Moore JH. Where are we now? A large benchmark study of recent symbolic regression methods. In: Proceedings of the 2018 Genetic and Evolutionary Computation Conference [Internet]. 2018. (GECCO ’18). Available from: <http://arxiv.org/abs/1804.09331>

* *Nominated for Best Paper Award*

29. Urbanowicz RJ, Meeker M, **La Cava W**, Olson RS, Moore JH. Relief-based feature selection: Introduction and review. Journal of Biomedical Informatics. 2018;85:189–203.

30. **La Cava W**, Bauer CR, Moore JH, Pendergrass SA. Interpretation of machine learning predictions for patient outcomes in electronic health records. In AMIA; 2019. (AMIA 2019 Annual Symposium; vol. abs/1903.12074). Available from: <http://arxiv.org/abs/1903.12074>

31. **La Cava W**, Helmuth T, Spector L, Moore JH. A probabilistic and multi-objective analysis of lexicase selection and epsilon-lexicase selection. Evolutionary Computation. 2019;27(3):377–402.

32. **La Cava W**, Moore JH. Learning concise representations for regression by evolving networks of trees. In: International Conference on Learning Representations. 2019. (ICLR). Available from: <http://arxiv.org/abs/1807.00981>

33. **La Cava W**, Moore JH. Semantic variation operators for multidimensional genetic programming. In: Proceedings of the 2019 Genetic and Evolutionary Computation Conference [Internet]. Prague, Czech Republic: ACM; 2019. (GECCO ’19). Available from: <http://arxiv.org/abs/1904.08577>

* *Nominated for Best Paper Award*

34. **La Cava W**, Moore JH. Genetic programming approaches to learning fair classifiers. In: Proceedings of the 2020 Genetic and Evolutionary Computation Conference [Internet]. 2020. (GECCO ’20). Available from: <https://arxiv.org/abs/2004.13282>

* *Best Paper Award*

35. **1. La Cava W**, Moore JH. Learning feature spaces for regression with genetic programming. Genet Program Evolvable Mach [Internet]. 2020 Mar 11; Available from: <https://doi.org/10.1007/s10710-020-09383-4>

36. **La Cava W**, Williams H, Fu W, Vitale S, Srivatsan D, Moore JH. Evaluating recommender systems for AI-driven biomedical informatics. Bioinformatics [Internet]. 2020 Aug 7; (btaa698). Available from: <https://doi.org/10.1093/bioinformatics/btaa698>

37. Danai K, **La Cava W**. Controller design by symbolic regression. Mechanical Systems and Signal Processing. 2021;151:107348.

**Non-peer reviewed scholarship in print or other media:**

|  |
| --- |
| Group materials into the following categories:   * Proceedings of meetings or other non-peer reviewed research scholarship * Reviews, chapters, monographs and editorials * Books/Textbooks for the medical or scientific community   + Only include books for which you are listed as an author. Books that you edited should be listed under Other Editorial Activities. * Case reports * Letters to the Editor |

**Proceedings of meetings or other non-peer reviewed scholarship**

1. Link H, **La Cava W**, van Dam J, McNiff B, Sheng S, Wallen R, et al. Gearbox reliability collaborative project report: findings from phase 1 and phase 2 testing. National Renewable Energy Laboratory; 2011. Report No.: https://www.nrel.gov/docs/fy11osti/51885.pdfNREL/TP-5000-51885.

2. Sheng S, Link H, **La Cava W**, Van Dam J, McNiff B, Veers P, et al. Wind turbine drivetrain condition monitoring during GRC phase 1 and phase 2 testing [Internet]. National Renewable Energy Laboratory; 2011. Report No.: http://www.nrel.gov/docs/fy12osti/52748.pdfNREL/TP-5000-52748. Available from: <http://www.nrel.gov/docs/fy12osti/52748.pdf>

3. **La Cava W**, Lackner M. Theory manual for the tuned mass damper module in FAST 8 [Internet]. University of Massachusetts Amherst; 2015 Mar. Available from: <https://nwtc.nrel.gov/TMD>

**Reviews, chapters, monographs and editorials**

1. Kannappan K, Spector L, Sipper M, Helmuth T, **La Cava W**, Wisdom J, et al. Analyzing a Decade of Human-Competitive (“HUMIE") Winners: What Can We Learn? In: Genetic Programming Theory and Practice XII [Internet]. Springer; 2015. p. 149–166. Available from: <http://link.springer.com/chapter/10.1007/978-3-319-16030-6_9>

2. **La Cava W**, Spector L. Inheritable Epigenetics in Genetic Programming. In: Riolo R, Worzel WP, Kotanchek M, editors. Genetic Programming Theory and Practice XII [Internet]. Cham: Springer; 2015. p. 37–51. Available from: <http://link.springer.com/10.1007/978-3-319-16030-6_3>

3. Olson RS, Sipper M, **La Cava W**, Tartarone S, Vitale S, Fu W, et al. A System for Accessible Artificial Intelligence. In: Genetic Programming Theory and Practice XIV. Springer; 2017.

4. Spector L, **La Cava W**, Shanabrook S, Helmuth T, Pantridge E. Relaxations of Lexicase Parent Selection. In: Banzhaf W, Olson RS, Tozier W, Riolo R, editors. Genetic Programming Theory and Practice XV. Cham: Springer International Publishing; 2018. p. 105–120.

5. Urbanowicz RJ, Meeker M, **La Cava W**, Olson RS, Moore JH. Relief-based feature selection: Introduction and review. Journal of Biomedical Informatics. 2018;85:189 – 203.

Thesis:

|  |
| --- |
| Provide full citation for doctoral thesis |

|  |
| --- |
| 1. **La Cava W**. Automatic Development and Adaptation of Concise Nonlinear Models for System Identification. In: Doctoral Dissertations May 2014 - current [Internet]. 2016. Available from: <http://scholarworks.umass.edu/dissertations_2/731/> |

Abstracts, Poster Presentations and Exhibits Presented at Professional Meetings:

|  |
| --- |
| List abstracts published and exhibits presented at meeting during the last 3 years which have not already been published as full-length manuscripts. May also list all abstracts or exhibits, regardless of date or publication as full-length manuscript, which received special recognition at a meeting (e.g., juried poster presentation, meeting commendation). |

|  |
| --- |
| 1. **La Cava W**. Testing and Modeling of a 3-MW Wind Turbine Using Fully Coupled Simulation Codes. AWEA Wind Power. 2012.  2. **La Cava W**. Automatic Development of Dynamic Models for Complex Systems. High Performance Computing Day, UMass Darthmouth. 2014.  3. **La Cava W**. Identifying Epistatic Interactions in Biomedical Data by Learning Multidimensional Feature Transformations. Pacific Symposium on Biocomputing. 2017.  4. Wojcieszynski Jr AP, **La Cava W**, Baumann BC, Lukens JN, Fotouhi Ghiam A, Urbanowicz RJ, et al. Machine Learning to Predict Toxicity in Head and Neck Cancer Patients Treated with Definitive Chemoradiation. International Journal of Radiation Oncology • Biology • Physics. 2019.  5. **La Cava W**, Lee PC, Ajmal I, Ding X, Cohen JB, Moore JH, et al. Application of flexible machine learning to construct accurate and interpretable EHR computable phenotypes. In 2020. (Symposium on Artificial Intelligence for Learning Health Systems (SAIL)).   * *Symposium cancelled due to COVID-19* |

Narrative Report

|  |
| --- |
| The narrative should describe your major contributions and achievements and should be clear to those outside your field. Readers should understand your career trajectory and impact. You need not reiterate your history of training, but the chronology of your accomplishments should be clear. You should make clear the extent of your reputation and the key achievements for which you are known. If your accomplishments bridge difference arenas (e.g., research and clinical care), you may wish to clarify any connections that are not readily apparent. Additionally, feel free to include any meaningful educational experiences (e.g., leadership or specialized course training) that you feel have impacted your career trajectory.  The narrative should be written in the first person and should be concise. In general, length should be commensurate with rank. Most narratives for Instructor and Assistant Professor candidates should ≤ 1 page; no narrative should exceed two pages.  Please include:   * Your contributions in your Area of Excellence (Investigation, Teaching and Educational Leadership, or Clinical Expertise and Innovation). You may want to include a description of work in progress that may otherwise not be reflected in your CV. * Your contributions in teaching (if not already described under your Area of Excellence). * Your contributions in any Significant Supporting Activities (Investigation, Clinical Expertise, Special Merit in Education, Administration & Institutional Service, Education of Patients and Service to the Community). |

|  |
| --- |
| My research focuses on developing multi-objective learning methods and using them to explain the principles underlying complex, biomedical processes. I use these methods to learn predictive models from electronic health records (EHR) that are both interpretable to clinicians and fair to the population on which they are deployed. My long-term goals are to positively impact human health by developing methods that are flexible enough to automate entire computational workflows underlying scientific discovery and medicine.  My dissertation focused on the automatic derivation of concise, nonlinear dynamic models from measurements. I contributed methods for learning the structural *form* of these models, rather than their parameters. I proposed local and global search methods for tackling two different scenarios: one in which we start with an expert model of the process, and one in which we start from scratch. In the first, I showed that one could use gradients to adapt model structures to observations. In the second scenario, I contributed improvements to Pareto optimization, a multi-objective search method for characterizing trade-offs (in our case, between model accuracy and simplicity). These improvements allowed us to identify exact solutions to several nonlinear differential equation systems from various fields of engineering, and to identify simple and accurate dynamic models of industrial wind turbines from operational data.  My postdoctoral work has focused on learning intelligible representations that couple with standard machine learning (ML) approaches and produce transparent predictive models from electronic health records. This work is motivated by the need for representations that can 1) represent complex interactions in the data, 2) produce independent features that capture relevant factors of variation, and 3) satisfy multiple performance criteria. I have extended these methods to handle the many constraints required for rich definitions of fairness. I have contributed methods that are currently state-of-the-art for solving multi-objective problems, including models that balance accuracy and simplicity in regression, and methods for satisfying intersectional definitions of fairness in classification.  In addition to learning intelligible and fair models, my work has improved 1) accessibility of ML methods to non-experts and 2) the benchmarking methodologies used to evaluate ML methods. I helped develop an AI data science assistant known as PennAI that chooses and runs analyses for users with the click of a button. These methods have been used to produce state-of-the-art predictions for septic shock in critical care patients and to produce highly accurate and interpretable models of resistant hypertension for patient screening.  For the past four years, I have helped organize conference workshops on benchmarking. My work has significantly impacted benchmarking methodologies used in different areas of machine learning, including regression, feature selection, and in applications to biomedical informatics.  I have advised several trainees at the undergraduate and graduate level through research projects leading to impactful publications. These projects have been at the intersection of biomedical informatics and machine learning. In addition, I have experience teaching informatics, medical, and engineering students at the undergraduate and graduate level in data science, machine learning, and ethical artificial intelligence. |