# Edison Mucllari

Curriculum Vitae

## Personal Information


Office: 718 Patterson Office Tower, Lexington, Kentucky 40506 USA

| | | | |
|---|---|---|---|
| Email: | edison.mucllari@uky.edu | LinkedIn: | edisonmucllari |
| GitHub: | Edison1994 | Semantic Scholar: | 2181503088 |
| Web-site: | emucllari1.github.io | ORCiD: | 20009-0006-6450-2837 |


## Education

**Doctor of Philosophy in Mathematics** — May 2025
University of Kentucky | Lexington, KY
Dissertation Advisor: Qiang Ye, Ph.D.
Thesis: "*Novel Generative and Language Model Architectures With Applications*"
Research Area: Deep Learning

**Master of Science in Mathematics** — July 2018
University of Tirana | Tirana, Albania
Advisor: Artur Baxhaku, Ph.D.
Master Thesis: "*Factorization of numbers and its applications in asymmetric cryptography*"
Institutional Honors: *Excellent Student*

**Bachelor in Mathematics** — July 2016
University of Tirana | Tirana, Albania
Institutional Honors: *Excellent Student*

## Research Interest

Numerical Analysis, Data Science, Statistics, Algorithmic Development, Natural Language Processing, Computer Vision, Generative AI, LLM, Audio, Speech Enhancement, Continual Learning, Adversarial Machine Learning, Optimal Transport, Causal Inference, HD Computing, Model Predictive Control, Drug Design & Molecular Prediction

## Employment

**Data Scientist Intern** — August 2025 - Present
Infineon Technologies | Lexington, KY

- Developed a hybrid recommender system integrating Graph Neural Networks and Transformer architectures to enhance product recommendation accuracy and customer experience on the Infineon website

**Ph.D. Research Intern** May 2023 - Aug. 2023

SRI International | Princeton, NJ

- Collaborated with leading researchers at SRI International, including Distinguished Computer Scientist, Dr. Michael Isnardi, and worked under the direct supervision of the Senior Technical Manager at Edge Computing Group, Dr. David Zhang
- Enhanced the Transformer architecture by introducing a new and innovative Scalable Transformer that outperformed Transformer XL and Recurrent Memory Transformer, resulting in a more efficient and effective design with a smaller model across various datasets like Word PTB, Wiki-Text 103, enwiki8
- Incorporated a new and novel concept into Hyperdimensional Computing (HD Computing), opening up numerous opportunities in this field
- Collaborated on optimizing Matrix-Matrix and Matrix-Vector Multiplication by focusing on reducing the number of multiplications and improving computation time

**Ph.D. Research Intern** May 2024 - Aug. 2024

SRI International | Princeton, NJ

- Developed novel coreset structures for Continual Learning, demonstrating provable guarantees and significant reduction in catastrophic forgetting, particularly in high-noise environments where our proposed strategies outperformed established methods including Random Replay, EWC, iCaRL, and DER
- Achieved state-of-the-art results on the Toyota Smarthome dataset by implementing CRT, demonstrating its effectiveness for activity recognition in smart home environments

**Research Assistant in Department of Mathematics** May 2020 - June 2025

University of Kentucky (UKY) | Lexington, KY

- Explored the behaviour of language models, specifically the Gated Recurrent Units (GRU) architecture, and analyzed the gradient impact during the training process
- Proposed the usage of orthogonal matrices to prevent exploding gradient problems and enhance long-term memory, where the employed Neumann series-based Scaled Cayley transformation speed up training and achieve significant improvement over many traditional language models
- Introduced Neumann-Cayley Orthogonal Gated Recurrent Units (NC-GRU) as the cells building the AutoEncoder to enable the creation of neural fingerprints, where the proposed architecture results in a more stable training, reliable molecular fingerprints and extensive numerical experiments show that NC-GRU AutoEncoder is a state-of-the-art framework in many molecular prediction benchmarks
- Successfully integrated the combination of Gated Recurrent Units (GRU) and Generative Adversarial Networks (GAN) to model the extremely complex forward process to generate the observed topside welding image from eight consecutive backside image where the presented results show an improved generator architecture due to the GRU mechanism to discriminate bottom images per timing/order
- Implemented a Machine Learning technique that can predict the best time to extubate a preterm baby that need mechanical ventilation in their early days of life considering that extubation failure happens quite often and the babies are required to be reintubated
- Modifying the discriminator loss function in Conformer-Based Metric GAN by incorporating more details from the data distribution to enhance speech signal quality
- Optimizing reasoning capabilities in LLM models through better training techniques utilizing numerical preconditioning

**Teaching Assistant and Primary Instructor** Aug. 2019 - May 2025

University of Kentucky | Lexington, KY

- Lead interactive discussions about course materials for 60 undergraduate students to encourage collaboration, creating a more inclusive learning environment
- Instructed 15 undergraduate students in an intensive group studying approach to the standard calculus sequence and spent extra time each week with three problem-solving workshops emphasizing collaborative learnings and working closely with a motivated group of students with assistance from an undergraduate assistant
- Supervised the teaching methods of other graduate teaching assistants and provided feedbacks cultivating more effective pedagogy practices
- Designed the curriculum, prepared the lectures for the daily intensive summer class, helped to develop content such as writing or proofreading weekly quizzes, and actively checking in to make sure the students progress is going well
- Provided assistance to undergraduate students in person and/or online based on student need to answer questions or provide time to make up assignments resulting in students showing better understanding of the course material as demonstrated by improved quiz scores each week
- Courses include: Calculus I, II and III; Elementary Calculus and its Applications; Finite Math and Its Applications; Calculus I and II with Life Science Applications; Ordinary Differential Equations

## Doctoral Degree Progress

**Qualifying Exam** March 2021

Title: Cayley Orthogonal Gated Recurrent Units and its application in Bio-molecular data

Abstract: Exploding/vanishing gradients have always been an issue in Gated Recurrent Units (GRU). In our proposed model, NC-GRU, we implement orthogonality using Cayley orthogonal matrices into the GRU gates. Based on the properties of the orthogonal matrices, NC-GRU is able to remember more information from the past. Combining this property with the ability of GRU to forget unnecessary information results in an improved model. Our experiments show that NC-GRU outperforms GRU, LSTM, RNN, scoRNN in different tasks. Even after the introduction of Transformers, GRU is still heavily used in Drug Design. We also conducted experiments in Bio-Molecular Data and our numerical results show a state-of-the-art architecture in many benchmarks, including IGC50, LC50, logP and FreeSolv.

Passed Preliminary Exam in Numerical Analysis Jan. 2021
Passed Preliminary Exam in Modern Algebra June 2020
Passed Preliminary Exam in Analysis June 2020

## Summer School and Workshop Experience

Oxford Machine Learning Summer School,ML x Health | Virtual — Aug. 7 - 10, 2022 / June 27 - 29, 2022

- Organised by AI for Global Goals and in partnership with CIFAR and the University of Oxford's Deep Medicine Program
- Covered 48 hours of lectures of advanced topics in ML theory and its applications in healthcare and medicine
- Accepted to and attended the OxML Health Track ONLINE
- Received the Certificate of Attendance

Workshop on Software and Applications of Numerical Nonlinear Algebra — May 31 - June 2, 2022

- Organised by Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
- Topic of this workshop is solving systems of polynomial equations using numerical algorithms. The participants learned to work with the software HomotopyContinuation.jl, so that they can use it for their own research problems. In addition, the workshop featured a series of lectures on the theory of numerical homotopy continuation, as well as contributed talks by Timothy Duff, Alex Heaton, Julia Lindberg and Maggie Regan
- Accepted to and attended the Virtual event - Videobroadcast

## Service and Outreach

| | |
|---|---|
| Graduate Student Congress Representative of the Mathematics Department at UKY | Sept. 2023 - May 2024 |
| Treasurer of the UKY chapter of Society for Industrial and Applied Mathematics (SIAM) | Sept. 2022 - May 2023 |
| Secretary of the SIAM UKY Chapter | Sept. 2021 - May 2022 |
| Graduate Student Colloquium Coordinator at the University of Kentucky (UKY) | Spring 2021 |
| Co-Coordinator of the UKY Mathematics Department Graduate Student Colloquium | Fall 2021 |
| Peer Mentor of a Graduate Student at the University of Kentucky (UKY) | Fall 2020 |
| University of Kenctucky Julia Robinson Math Festival | Nov. 2019 |
| Proctor at the Math Kangaroo Olympiad | March 2018 |
| Contributed to the coordination and execution of the Conference of Natural Sciences and Technology at University of Tirana | Nov. 2017 |

## Awards and Recognitions

| | |
|---|---|
| Fall 2024 Research Assistantship, NSF | Jan. 2024 |
| Spring 2023 Research Assistantship, NSF | Oct. 2023 |
| Research Assistantship, Math Department, UKY | 2021 - Present |
| Graduate Assistantship, Math Department, UKY | 2019 - Present |
| SIAM Certification for outstanding efforts and accomplishments of the UKY SIAM Chapter | Aug. 2023 |
| Best Student, Master of Science of Mathematics, University of Tirana | 2018 |
| Outstanding Undergraduate Student, University-wide prize and scholarship, University of Tirana | 2014-2016 |
| Excellent Student, Selected by Prime Minister of Albania | 2014 |

## Research

**Continual Learning Project** | Princeton, NJ — May. 2024 - April 2025

Advisors: Zachary Daniels Ph.D.

Collaboration: Joint work with Aswin Raghavan, Ph.D.

Title: Noise-Tolerant Coreset-Based Class Incremental Continual Learning

ABSTRACT: Many applications of computer vision require the ability to adapt to novel data distributions after deployment. Adaptation requires algorithms capable of continual learning (CL). Continual learners must be plastic to adapt to novel tasks while minimizing forgetting of previous tasks. However, CL opens up avenues for noise to enter the training pipeline and disrupt the CL. This work focuses on label noise and instance noise in the context of class-incremental learning (CIL), where new classes are added to a classifier over time, and there is no access to external data from past classes. We aim to understand the sensitivity of CL methods that work by replaying items from a memory constructed using the idea of Coresets. We derive a new bound for the robustness of such a method to uncorrelated instance noise under a general additive noise threat model, revealing several insights. Putting the theory into practice, we create two continual learning algorithms to construct noise-tolerant replay buffers. We empirically compare the effectiveness of prior memory-based continual learners and the proposed algorithms under label and uncorrelated instance noise on five diverse datasets. We show that existing memory-based CL are not robust whereas the proposed methods exhibit significant improvements in maximizing classification accuracy and minimizing forgetting in the noisy CIL setting.
SUPPORT: This research is part of my summer internship and it is internal research and development (IRAD) project at SRI International

**Recurrent Transformer Project** | Princeton, NJ MAY. 2023 - MAY 2025
ADVISORS: Qiang Ye, Ph.D. and David Zhang. Ph.D.
COLLABORATION: Joint work with Zachary Daniels, Ph.D.
TITLE: Compact Recurrent Transformer with Persistent Memory
ABSTRACT: The Transformer architecture has shown significant success in many language processing and visual tasks. However, the method faces challenges in efficiently scaling to long sequences because the self-attention computation is quadratic with respect to the input length. To overcome this limitation, several approaches scale to longer sequences by breaking long sequences into a series of segments, restricting self-attention to local dependencies between tokens within each segment and using a memory mechanism to manage information flow between segments. However, these approached generally introduce additional compute overhead that restricts them from being used for applications where limited compute memory and power are of great concern (such as edge computing). We propose a novel and efficient Compact Recurrent Transformer (CRT), which combines shallow Transformer models that process short local segments with recurrent neural networks to compress and manage a single persistent memory vector that summarizes long-range global information between segments. We evaluate CRT on WordPTB and WikiText-103 for nexttoken-prediction tasks, as well as on the Toyota Smarthome video dataset for classification. CRT achieves comparable or superior prediction results to full-length Transformers in the language datasets while using significantly shorter segments (half or quarter size) and substantially reduced FLOPs. Our approach also demonstrates state-of-the-art performance on the Toyota Smarthome video dataset.
SUPPORT: This research is based upon work supported in part by NSF under the grant IIS-2327113 and DMS-2208314 and the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via Contract No: 2022-21100600001. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

**Improved Techniques for Welding GRU - GAN Training** | Redmond, WA APR. 2023 - MARCH 2024
ADVISORS: YuMing Zhang, Ph.D. and Qiang Ye, Ph.D.
COLLABORATION: Joint work with Yue Cao, M.S.
TITLE: Modeling the Dynamics of Weld Pool Behavior using Generative Adversarial Network (GAN)

Abstract: This paper aims to model the dynamic behaviors of welding process that can be observed by humans and monitored by imaging sensors. This is fundamental in deciding if such dynamic behaviors can be used as raw data, that contains sufficient relevant information, to monitor what occurs underneath the workpiece determining the weld integrity. Challenges arise as featurization of these observations proves difficult, and the governing laws for the underlying process remain largely unknown. Leveraging the ability of Generative Adversarial Networks (GANs) to generate intricate phenomena from simple inputs by automatically approximating unknown underlying laws, we propose using a GAN to predict the dynamic behaviors of welding processes. In this framework, dynamic behaviors serve as the output, while the trained GAN functions as the model for the underlying process, generating the observed behaviors from the inputs. The success of predictions depends not only on the GAN's capability to approximate the process but also on the sufficiency of the inputs provided. With this consideration, we model an exceptionally complex process that correlates weld joint penetration with observed phenomena to determine if the underlying process involves critical inputs beyond penetration alone. Our findings reveal the necessity of welding currents for accurate predictions by the GAN model. This implies that the combined information from human-observed/imaged phenomena and welding currents provides a more comprehensive basis for monitoring weld penetration, contributing to the advancement of robotizing welding processes with the necessary intelligence.
Support: Supported in part by NSF under grants No. 2024614 and DMS-2208314, the Institute for Sustainable Manufacturing, Department of Electrical and Computer Engineering, and Department of Mathematics at the University of Kentucky

**Novel Welding GRU - GAN Project** | University of Kentucky — Dec. 2022 - April 2023

Advisors: YuMing Zhang, Ph.D. and Qiang Ye, Ph.D.

Collaboration: Joint work with Rui Yu, M.S. and Yue Cao, M.S.

Title: Do We Need a New Foundation to Use Deep Learning to Monitor Weld Penetration?

Abstract: Deep learning has been successfully used to automate the modeling process that trains a network/model from a given experimental dataset to calculate the output directly using highdimensional complex raw data. However, the trained network is an inverse of the welding process (forward process) that produces the welding phenomena/measured raw data as the output with the penetration as the input of the forward process. Now the question is in addition to the current state of the weld penetration to be estimated if the forward process also has other inputs to determine its output. If it has, then the inverse model has to be constructed accordingly. This will call for a new foundation for deep learning-based monitoring of penetration. This letter proposed a novel innovative generative adversarial network (GAN) with GRU (Gated Recurrent Unit) in the generator, i.e., GRU-GAN, to model the extremely complex forward process to generate the observed topside welding image (output of the forward process) from the backside images (as comprehensive quantification of weld penetration). It is found that the produced topside welding image is not only determined by the current backside image but also by its history. A new foundation thus must be established to guide deep learning-based monitoring of weld penetration. The prediction model/network as an inverse model must be in compliance with the forward process that includes the history of the state of the weld penetration as its input.
Support: Supported in part by NSF under grants No. 2024614 and DMS-2208314, the Institute for Sustainable Manufacturing, Department of Electrical and Computer Engineering, and Department of Mathematics at the University of Kentucky

**Deep Learning in the Drug Discovery Project** | University of Kentucky — Sept. 2021 - Sept. 2022

Advisors: Duc Nguyen, Ph.D. and Qiang Ye, Ph.D.

Collaboration: Joint work with Vasily Zadorozhnyy, Ph.D.

Title: Novel Molecular Representations Using Neumann-Cayley Orthogonal Gated Recurrent Unit

Abstract: Advances in Deep Neural Networks (DNNs) have made a very powerful machine learning method available to researchers across many fields of study, including the bio-medical and cheminformatics communities, where DNNs help to improve tasks such as protein performance, molecular design, drug discovery, etc. Many of those tasks rely on molecular descriptors for representing molecular characteristics in cheminformatics. Despite significant efforts and the introduction of numerous methods that derive molecular descriptors, quantitative prediction of molecular properties remains challenging. One widely used method of encoding molecule features into bit strings is molecular fingerprint. In this work, we propose using new Neumann-Cayley Gated Recurrent Units (NC-GRU) inside the Neural Nets encoder (AutoEncoder) to create neural molecular fingerprints, NC-GRU fingerprints. NC-GRU AutoEncoder introduces orthogonal weights into widely used GRU architecture, resulting in faster, more stable training, and more reliable molecular fingerprints. Integrating novel NC-GRU fingerprints and Multi-Task DNN schematics improves the performance of various molecular-related tasks such as toxicity, partition coefficient, lipophilicity, and solvation-free energy, producing state-of-the-art results on several benchmarks.
Support: Supported in part by NSF under grants DMS-1821144, DMS-2053284, and DMS-2151802

**Orthogonality Preserving Recurrent Nets** | University of Kentucky — Aug. 2021 - July 2022
Advisors: Duc Nguyen, Ph.D. and Qiang Ye, Ph.D.
Collaboration: Joint work with Edison Mucllari, M.S. and Cole Pospisil, M.S.
Title: Orthogonal Gated Recurrent Unit with Neumann-Cayley Transformation
Abstract: In recent years, using orthogonal matrices has been shown to be a promising approach in improving Recurrent Neural Networks (RNNs) with training, stability, and convergence, particularly to control gradients. While Gated Recurrent Unit (GRU) and Long Short Term Memory (LSTM) architectures address the vanishing gradient problem by using a variety of gates and memory cells, they are still prone to the exploding gradient problem. In this work, we analyze the gradients in GRU and propose the usage of orthogonal matrices to prevent exploding gradient problems and enhance long-term memory. We study where to use orthogonal matrices and we propose a Neumann series-based Scaled Cayley transformation for training orthogonal matrices in GRU, which we call Neumann-Cayley Orthogonal GRU, or simply NC-GRU. We present detailed experiments of our model on several synthetic and real-world tasks, which show that NC-GRU significantly outperforms GRU as well as several other RNNs.
Support: Supported in part by NSF under grants DMS-1821144, DMS-2053284 and DMS-21518

## Invited/Conference Talks

- International Conference on Automation Science and Engineering — Aug. 26 - Aug. 30, 2023
  **Title:** Do we need a new foundation to use Deep Learning to monitor weld penetration?
- SRI International Intern Presentation Day — Aug. 2023
  **Title:** Towards Trainable Deep Hyperdimensional Computing
- SRI International (Online) — April 2023
  **Title:** Novel Welding GRU-GAN architecture and Neumann-Cayley Orthogonal Gated Recurrent Units with Some Applications in Bio-molecular Data
- SIAM Conference on Computational Science and Engineering (CSE23) — Feb. 26 - March 3, 2023
  **Title:** Neumann-Cayley Orthogonal Gated Recurrent Units with Some Applications in Bio-molecular Data
- University of Kentucky Applied Math Seminar — March 2022
  **Title:** Cayley Orthogonal Gated Recurrent Units and its application in Bio-molecular data
- University of Kentucky Department of Mathematics Graduate Student Colloquium — Oct. 2021
  **Title:** Optimal Transport, Monge Problem and Kantarovich Formulation
- University of Kentucky Department of Mathematics Graduate Student Colloquium — May 2021
  **Title:** Convolutional Neural Networks with Applications
- Faculty of Natural Sciences in University of Tirana — July 2018
  **Title:** Factorization of numbers and its applications in asymmetric cryptography

## Publications and Works-in-Progress

1. **Edison Mucllari**, Aswin Raghavan, Zachary Daniels. *Noise-Tolerant Coreset-Based Class Incremental Continual Learning.* arxiv.org/abs/2504.16763.
2. **Edison Mucllari**, Aswin Raghavan, Zachary Daniels. *Class-incremental SAR ATR in noisy and adversarial environments* Automatic Target Recognition XXXV.
3. **Edison Mucllari**, Zachary Daniels, David Zhang, Qiang Ye. *Compact Recurrent Transformer with Persistent Memory.* arxiv.org/abs/2505.00929.
4. Aswin Raghavan, David Zhang, Saurabh Farkya, Zachary Daniels, Michael Piacentino, Gooitzen Van Der Wal, Philip Miller, Michael Lomnitz, Abrar Rahman, **Edison Mucllari**, Shahir Rahman. *Energy efficient machine learning on the edge with query-based knowledge assistance.* US Patent App. 18/944,744.
5. **Edison Mucllari**, Yue Cao, Qiang Ye, YuMing Zhang. Modeling Imaged Welding Process Dynamic Behaviors Using GAN for a New Foundation to Monitor Weld Penetration Using DL. Journal of Manufacturing Processes3
6. **Edison Mucllari**, Rui Yu, Yue Cao, Qiang Ye, YuMing Zhang. Do we need a new foundation to use Deep Learning to monitor weld penetration. IEEE Robotics and Automation Letters, 2023
7. **Edison Mucllari***, Vasily Zadorozhnyy*, Qiang Ye, Duc Nguyen. Novel Molecular Representations using Neumann-Cayley Orthogonal Gated Recurrent Units. Journal of Chemical Information and Modeling, 2023 (*-equal contribution )
8. Vasily Zadorozhnyy*, **Edison Mucllari***, Cole Pospisil, Duc Nguyen, Qiang Ye. Orthogonal Gated Recurrent Unit with Neumann-Cayley Transformation Unit. Neural Computation 1–26. (*-equal contribution )

## Languages and Computer Skills

| | |
|---|---|
| Languages: | English(fluent), Albanian(fluent), Italian(basic knowledge) |
| Programming languages: | PYTHON, JULIA, BASH, JAVA, C/C++, R, HTML/CSS, REST, SQL |
| Mathematical and statistical softwares: | Matlab, Mathematica, Maple, SAS, LaTeX, Tableau, PowerBI |
| Machine and Deep learning libraries: | NumPy, Pandas, SciPy, CuPy, TensorFlow, PyTorch, Jax |
| Containerization and orchestration: | Docker, Kubernetes |
| Cloud platforms and services: | AWS, Azure, Google Cloud |
| DevOps and CI/CD: | Jenkins, GitHub Actions, GitLab CI/CD |