SURYABHAN SINGH HADA

shada@ucmerced.edu | http://graduatestudents.ucmerced.edu/shada
+1 (209) 777-2068 | 2916 Oleander Ave, Merced, CA 95340

SHORT BIO

I am an AI and ML researcher with an extensive track record of peer-reviewed publications. My area of expertise is **interpretable machine learning**. At the same time, I have experience with a wide range of projects involving **natural language processing (NLP), computer vision (CV), financial analytics**, and others.

EDUCATION

University of California, Merced, Merced, California, USA 08/2016 - 05/2022 (expected)

- **Ph.D. Candidate** in Electrical Engineering and Computer Science (Machine Learning).
- *Advisor:* Miguel Á. Carreira-Perpiñán.

Indian Institute of Technology (BHU), Varanasi, India 06/2009 - 05/2014

- **Integrated Masters of Technology** (B. Tech & M. Tech) in Mathematics and Computing.
- First-class honours.
- *Advisor:* Dr. Santwana Mukhopadhyay.

PROFESSIONAL EXPERIENCE

University of California, Merced (Graduate Research Assistant) 08/2016 - present

- **Interpretable word-embeddings:** Created an effective algorithm to determine what concepts are represented by the dimensions of a word-embedding vector, which allows us to detect human biases associated with different words in a language.
- **Class-specific neurons in a deep net:** Created an algorithm to find a small subset of neurons in a deep neural network associated with a specific class and visualization of the distribution of classes in the latent space.
  - By controlling the activation of these neurons, we can make the net only to predict or never predict a given class.
  - These neurons can find the input features essential to a given class.
  - Our approach also allows us to extract if-else type rules from a deep net.
- **Inverse-set of a neuron in a deep net:** Created an algorithm to characterize the behavior of neurons in a deep neural network in terms of what concept in a given class changes the activation of a neuron.
- **Exact counterfactual explanations:** Created exact and efficient algorithms for the non-differentiable problem of counterfactual explanations to interpret large decision trees for datasets with both continuous and categorical variables.
  - Our approach allows us to answer various questions, like finding the closest class to a given input or finding the critical feature to flip the model decision.
  - The algorithms are fast enough for real-time use.

Cvent, Inc. (Software Engineer) 07/2014 - 06/2016

- Developed a new platform for event management and customized data integration for third parties using Microsoft .net as web tier for UI, Drop Wizard for scalable rest layer, RabbitMq as message broker, Couchbase as NoSql datastore.
- Created an android app that helps the employees to navigate inside the office and broadcast messages at team and department levels.

PUBLICATIONS

- **S. S. Hada**, M. Á. Carreira-Perpiñán and A. Zharmagambetov: "*Understanding and Manipulating Neural Net Features Using Sparse Oblique Classification Trees*" in proceedings of IEEE International Conference on Image Processing (ICIP) 2021.
- **S. S. Hada** and M. Á. Carreira-Perpiñán: "*Sampling the "Inverse Set" of a Neuron: An Approach to Understanding Neural Nets*" in proceedings of IEEE International Conference on Image Processing (ICIP) 2021.

- M. Á. Carreira-Perpiñán and **S. S. Hada**: "*Counterfactual Explanations for Oblique Decision Trees: Exact, Efficient Algorithms*" in AAAI conference on Artificial Intelligence (AAAI) 2021.
- **S. S. Hada** and M. Á. Carreira-Perpiñán: "*Style Transfer by Rigid Alignment in Neural Net Feature Space*" in proceedings of IEEE conference on Winter Conference of Applications on Computer Vision (WACV) 2021.
- A. Zharmagambetov, **S. S. Hada** and M. Á. Carreira-Perpiñán, M. Gabidolla: "*Non-Greedy Algorithms for Decision Tree Optimization: An Experimental Comparison*" in proceedings of IEEE International Joint Conference on Neural Networks (IJCNN) 2021.
- M. Á. Carreira-Perpiñán and **S. S. Hada**: "*Inverse classification with logistic and softmax classifiers: efficient optimization*" under review in AAAI Conference on Artificial Intelligence (AAAI) 2022.
- M. Á. Carreira-Perpiñán and **S. S. Hada**: "*More Interpretable Decision Trees: Pruning via Node Descent and the Delta Penalty*" under review in International Conference on Artificial Intelligence and Statistics (AISTATS) 2022.
- **S. S. Hada**, M. Á. Carreira-Perpiñán and A. Zharmagambetov: "*Sparse Oblique Decision Trees: A Tool to Understand and Manipulate Neural Net Features*." arXiv:2104.02922.

WORKSHOPS

- **S. S. Hada** and M. Á. Carreira-Perpiñán: "*Exploring counterfactual explanations for classification and regression trees*." ECML 2021.

TECHNICAL SKILLS

- **Programming Languages:** Python, MATLAB, Java, C++, C#
- **Deep learning tools:** Pytorch, Tensorflow, MatConvNet
- **Databases:** SQL, Couchbase, Cassandra

TEACHING EXPERIENCE

Teaching Assistant at University of California, Merced

- CSE120: Software Engineering (Spring & Fall 2021)
- Introduction to Computing (Spring 2020, Spring 2017 & Fall 2016)
- Algorithm Design and Analysis (Spring & Fall 2019)
- Discrete Mathematics (Summer 2019, Summer 2018 & Fall 2017)

OTHER

- Reviewer for AISTATS 2021.
- Reviewer for Explainable Artificial Intelligence Approaches for Debugging and Diagnosis Workshop, NEURIPS 2021.

AWARDS

- UC Merced Bobcat Fellowship, 2018 and 2019
- UC Merced Travel Fellowship, 2019 and 2020