The Informatics Skunkworks

A Program for Undergraduate Research at the Interface of Data Science and Materials Science and Engineering

Mahmood Mamivand
Department of Mechanical and Biomedical Engineering
July 14, 2021

BOISE STATE UNIVERSITY

ACKNOWLEDGEMENT & TEAM

NSF Cyber Training Implementation (OAC-2017072)

University of Wisconsin

- Dane Morgan
- Wendy Crone
- Anne Lynn Gillian-Daniel
- Benjamin Afflerbach
- Matthew Stilwell

Boise State

- Lan Li
- Eric Jankowski

WHY SHOULD YOU CARE?

How to start a ML research with limited or no ML experience

Faculty not at Boise State

- Email me and we can share our educational materials and experience

Undergraduate student at Boise State

- You can join my Informatics Skunkworks VIP https://www.boisestate.edu/coen-skunkworks/

Faculty or graduate student at Boise State

- You are welcome to join our mentor team

Student out of Boise State

- Stay toned

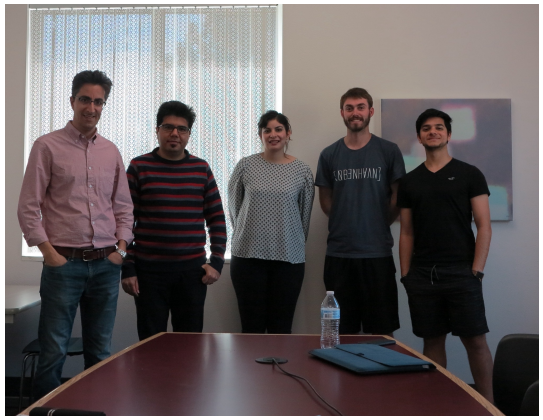
Mahmood Mamivand (mahmoodmamivand@boisestate.edu)

OUTLINE

- What is Informatics Skunkworks
- Motivation
- How does it operate
  - Inputs, Activities, and Outputs
- Invitations for Collaborations

WHAT IS INFORMATICS SKUNKWORKS?

A program that engages undergraduates in research at the intersection of informatics and science and engineering



Founded by Dane Morgan in 2015 at UW-Madison
Adopted at Boise State in 2019

More than 200 student participated at UW-Madison
More than 20 students participated at Boise State

MOTIVATION

Undergraduate research is a high impact teaching practice

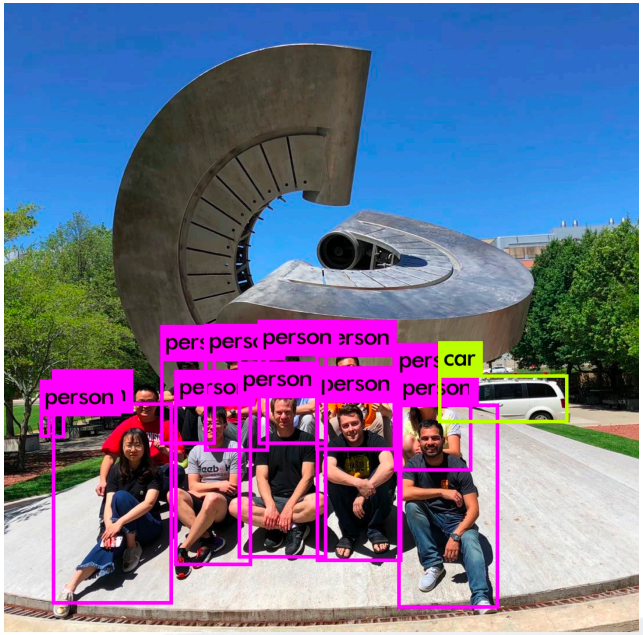
Undergraduate research is:

- Inexpensive
- Scalable

MOTIVATION

Machine Learning is diffusing in all fields

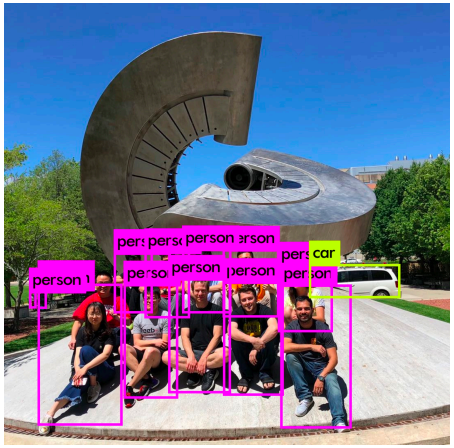
- Machine learning is an excellent research tool
- Machine learning is pervasive in our society and influences our lives
- Students are curious and interested in learning about machine learning
- Data literacy is a need for future jobs



MOTIVATION

Informatics Skunkworks engages undergraduates in machine learning research

MACHINE LEARNING



UNDERGRADUATE RESEARCH

Informatics Skunkworks

HOW DOES INFORMATICS SKUNKWORKS WORK?

Informatics Skunkworks

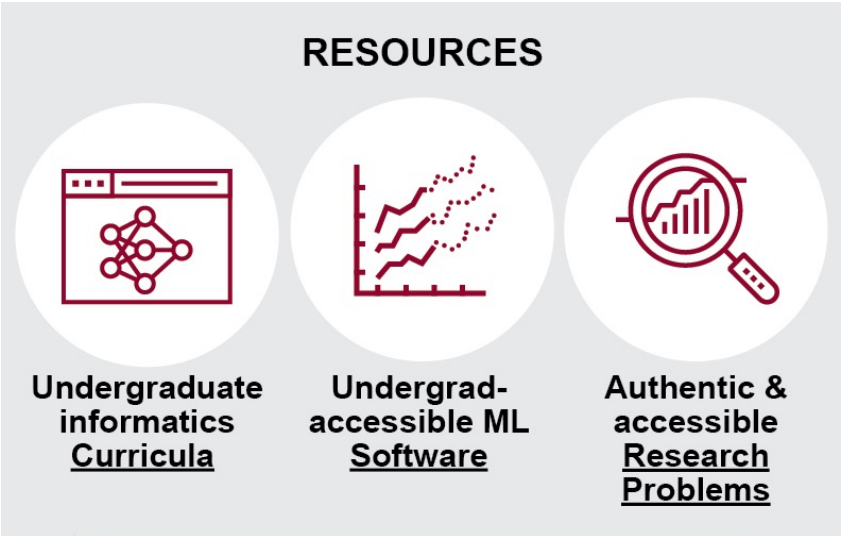
1. 10-week onboarding course
2. Following onboarding, students join mentored research teams

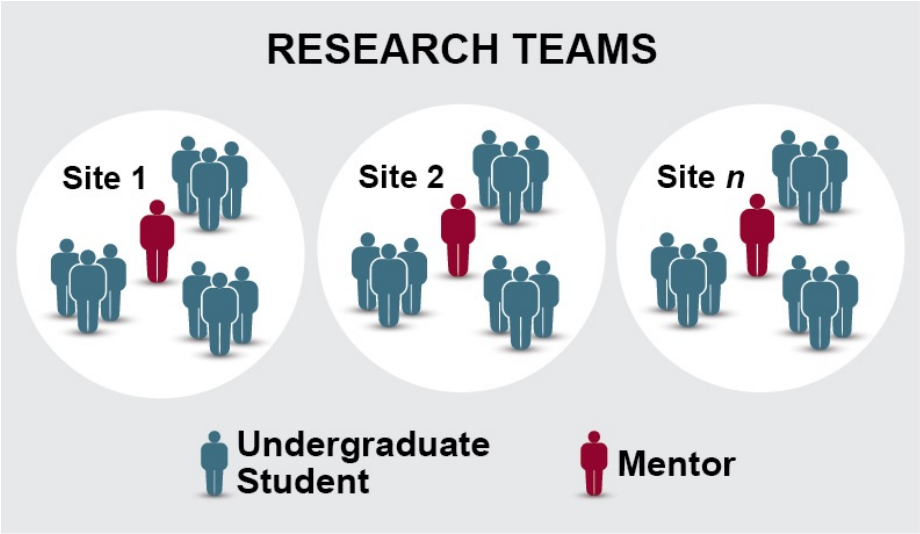
Research Team

Groups of 4-6 undergraduate students who

- Learn relevant background
- Develop data-centric models
- Disseminate results (talks, papers, codes)

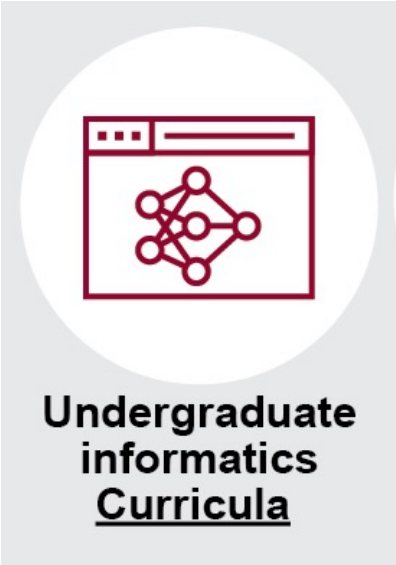
INFORMATICS SKUNKWORKS IS A COMBINATION OF RESOURCES AND A COMMUNITY OF PRACTICE





LEARNING MATERIALS

Accessible curricula enables students to quickly learn machine learning basics



Students work through modular curricula designed around specific learning objectives

Designed as a 10-week course

Students learn key background information from readings and videos

Then students work with example datasets, creating and assessing machine learning models

TOOLS

Software resources enable creation and assessment of machine learning models

Software tools are free and easily accessible

Leveraging previously developed and open-source tools and programs (Citrination, Colab)

Students actively explore data within an hour

EDUCATIONAL MATERIALS

- **Module 0: Machine Learning Predictions Activity**
- Learning Objectives
  - Students will familiarize themselves with a general overview of machine learning and how it can be used to make predictions.
- Activities
  - Machine Learning Predictions video

- **Module 1: Introduction to Citrination**
- Learning Objectives
  - Students will learn how Skunkworks is structured and how it will prepare them for undergraduate research projects
  - Students will access the Citrination client, upload data, and create a random forest model.
  - Students learn about how to establish group expectations, dynamics, communication
  - Students will think critically about a dataset and identify good or bad data.
  - Students discuss ways to address roadblocks and obstacles that frequently arise in research
- Activities
  - Introduction to Citrination
  - Research Compact activity
  - Ethical Data Cleaning

EDUCATIONAL MATERIALS

- **Module 2: Basics of Machine Learning**
- Learning Objectives
  - Students are introduced to model types, key outputs, and metrics for assessing performance
  - Students are introduced to effective practices for maintaining a useful record of research
  - Students continue to learn about model types, key outputs, and metrics for assessing performance
  - Students continue discussing effective practices for maintaining a useful record of research
- Activities
  - Video Introduction to Machine Learning
    - Powerpoint used in Video Introduction (to follow along)
  - Written Introduction to Machine Learning
    - Additional Resources

- **Module 3: Introduction to MAST-ML**
- Learning Objectives
  - Students are introduced to the MAST-ML software and given a few sample workflows to run on their dataset. The models generated here can then be compared against those previously generated with Citrination.
  - Students explore various cross validation methods in MAST-ML
- Activities
  - MASTML Introduction
  - Model Limitations

EDUCATIONAL MATERIALS

- **Module 4: Modifying Workflows with MASTML**
- Learning Objectives
  - Students learn how to modify workflows within MAST-ML to employ the software for various applications.
  - Students understand how MASTML can be used to recreate workflows from other machine learning software (like Citrination).
- Activities
  - Modifying Workflows in MASTML

- **Module 5: Hyperparameter Optimization**
  - Students learn about how model hyperparameters can affect performance and are introduced to some basic ideas on how these hyperparameters can be optimized.

  - Students familiarize themselves with a hyperparameter optimization landscape and explore various methods of searching for a global minimum in model errors and avoiding getting trapped in a local minimum.
  - Students familiarize themselves with neural networks, what they are, how to use them, and how to optimize them.
  - Activities
    - Hyperparameter Optimization Presentation Slides
    - Hyperparameter Optimization activity
    - Keeping an Electronic Research Notebook

EVALUATION

- Evaluation of Skunkworks highlights strengths and areas for improvement
- Pre-post surveys measure learning gains in machine learning concepts and research skills
- Working with an evaluator to interview student focus groups on their experience in the program
- Evaluation data will be used to refine curricula and support community of practice

RESEARCH PROBLEMS

Students tackle AUTHENTIC research problems after intro course

Authentic & accessible Research Problems

Research projects from various sources

- Mentor lab
- Literature
- Open datasets
- Industry
- Collaborators

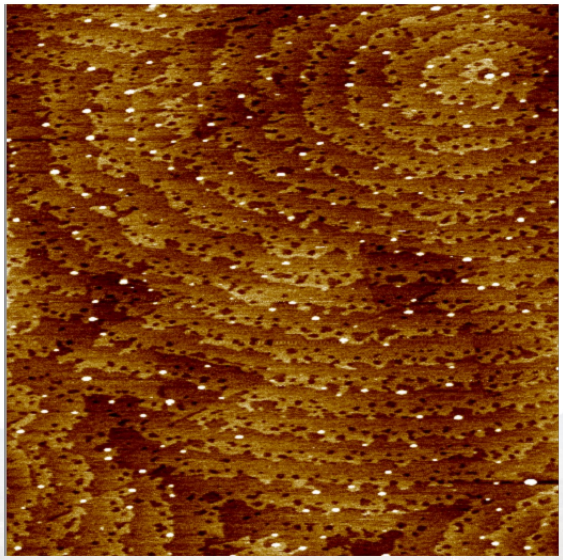
Potential research questions

- Predicting values, properties
- Comparing performance of model types

RESEARCH PROBLEMS

https://www.boisestate.edu/coen-skunkworks/projects/

StuCha	Automated materials structure characterization using machine learning.	The goal of this project is to accelerate and automate the process of materials characterization. Machine learning offers significant opportunities for extracting knowledge from the unlabeled data set. In this project, we will use machine learning to classify images in a popular microstructural dataset. We will use standard pre-trained convolutional neural networks such as VGG16, etc. to extract feature representations for each micrograph. The project will provide opportunities to gain skills in programming in python, using machine learning libraries, and interdisciplinary teamwork. We expect work to yield poster presentations for the Idaho Conference on Undergraduate Research, open-source code packages, and journal publications for motivated students.	**Eric Jankowski** <EricJankowski@boisestate.edu>; **Mahmood Mamivand** <mahmoodmamivand@boisestate.edu>	4	Some for Summer	Yes (Fall and Spring)	Active-Recruiting



MENTORS

- Mentors can be faculty, staff, graduate students, upper-level experienced undergraduates

- Mentors do NOT need to be machine learning experts
- Curriculum brings you up to speed quickly
- Mentors can learn along with students
- Community support

Even if you have NO experience, but are interested in becoming a mentor, we invite you to collaborate with us!

ACTIVITY

Weekly Meetings

Slide decks

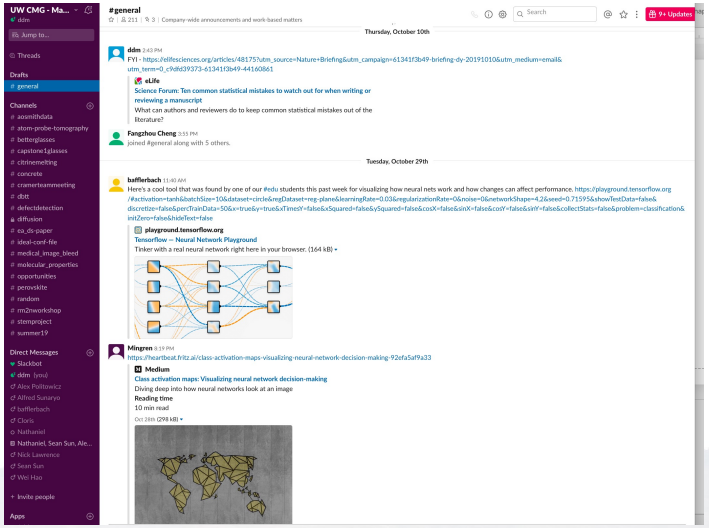
- Hours worked, what you did, found, and plan
- Supports group discussion on next step
- Provides record for grading, references

Slides, references talks, etc. all on cloud (gdocs, Mendeley)

ACTIVITY

Slack

- Easy, archived communication for many teams
- Includes flexible communications, links, small and large texts, images, etc.
- People can share results, opportunities, questions, etc.
- Supports a community



CREDIT/PAY

- Pay is great if you have funds but not needed and likely too expensive. However, this topic attracts support. We focus funds on summer and infrastructure support.
- Credits help students meet graduation requirements, take responsibility, and gives record of the work.
- At Boise we use Vertically Integrated Project (VIP)

ALL-HANDS

All-Hands meetings at end of each semester

Students/mentors present on most interesting projects

Sat. 10-2 with food

Great for creating community

OUTPUT: SCIENCE (PAPERS, TALKS)

Papers: ~1 / year (~50% undergraduate first author) – graduate students help get them completed.

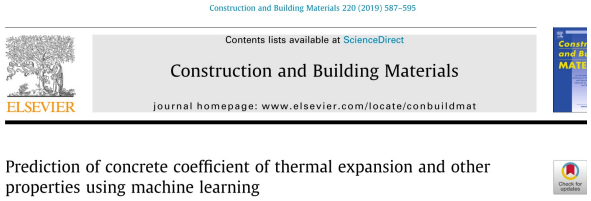
Talks: ~10 / year, many local (e.g., undergraduate symposiums) but some at major conferences (TMS, MRS, ACS).


Computational Materials Science 134 (2017) 160–165

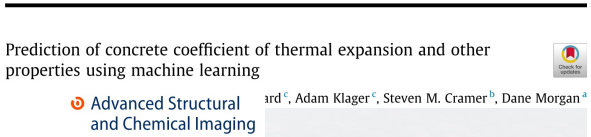
Contents lists available at ScienceDirect

Computational Materials Science

journal homepage: www.elsevier.com/locate/commatsci


Robust FCC solute diffusion predictions from ab-initio machine learning methods

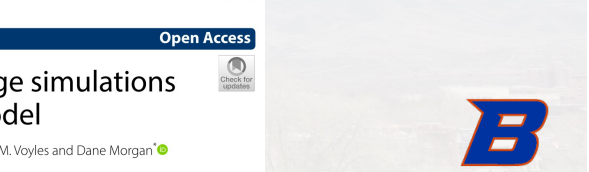
Henry Wu[a], Aren Lorenson[a], Ben Anderson[a], Liam Witteman[a], Haotian Wu[a], Bryce Meredig[b], Dane Morgan[a,*]


Construction and Building Materials 220 (2019) 587–595

Contents lists available at ScienceDirect

Construction and Building Materials

journal homepage: www.elsevier.com/locate/conbuildmat


Prediction of concrete coefficient of thermal expansion and other properties using machine learning

ard[c], Adam Klager[c], Steven M. Cramer[b], Dane Morgan[a]


Combs et al. Adv Struct Chem Imag (2019) 5:2
https://doi.org/10.1186/s40679-019-0064-2

Advanced Structural and Chemical Imaging

RESEARCH Open Access

Fast approximate STEM image simulations from a machine learning model

Aidan H. Combs, Jason J. Maldonis, Jie Feng, Zhongnan Xu, Paul M. Voyles and Dane Morgan*

OUTPUT: AWARDS

Students: ~15% of participants / year win awards. Typically campus awards, summer fellowships.

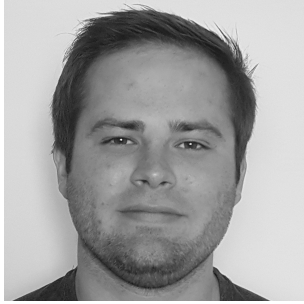
Skunkworks projects provide excellent foundation for research-centric award applications.

Skunkworks highlighted as one of the 32 accomplishments of the first 5 years of the MGI

Excellent for NSF Broader Impact

OUTPUT: JOBS/GRADUATE SCHOOL/VALUED EXPERIENCE

Students get experiences and resources (e.g., essay materials, letters) to support strong nest steps to top schools (MIT, Duke) and companies (Google, Tesla).



Joining the Skunkworks was one of the most beneficial activities I did during my undergraduate career. … It directly contributed to my admission to many top graduate schools including MIT, UC-Berkeley, and Northwestern. - Zach Jensen, 6/4/17

Joining the Informatics Skunkworks group was one of the best decisions I made during my undergraduate career. … I intend to go to graduate school to study computational social science, and my Skunkworks experience helped me get admission to some of the top programs in the field. - Aidan Combs, 4/5/18

OUTPUT: SKILLS

For Students

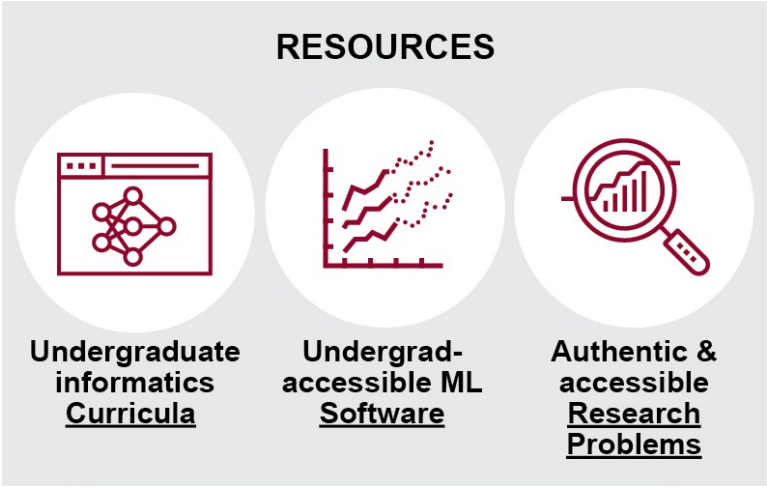
- Personal & Professional Skills
- Problem solving
- Project management
- Applied machine learning
- Teamwork
- Communication
- Data analysis

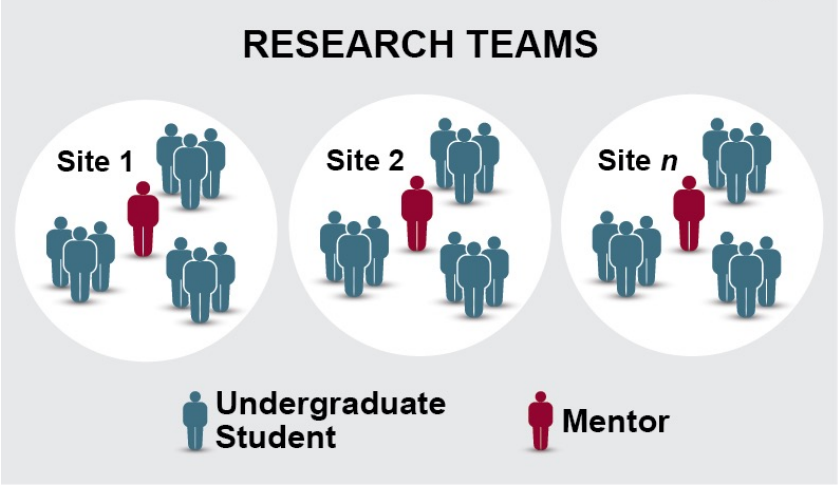
For Faculty

- Start (or expand) a new research direction

JOIN US!

Our community of practice can include you!
Institutions of all types welcome! – community college through research intensive





WHY SHOULD YOU CARE?

How to start ML research with limited or no ML experience

Faculty not at Boise State

- Email me and we can share our educational materials and experience

Undergrad student at Boise State

- You can join my Informatics Skunkworks VIP https://www.boisestate.edu/coen-skunkworks/

Faculty or grad student at Boise State

- You are welcome to join our mentor team

Student out of Boise State

- Stay toned

Mahmood Mamivand (mahmoodmamivand@boisestate.edu)

Thank you!

Mahmood Mamivand, mahmoodmamivand@boisestate.edu

Learn More:
Boise: https://www.boisestate.edu/coen-skunkworks/
UW-Madison: https://skunkworks.engr.wisc.edu/