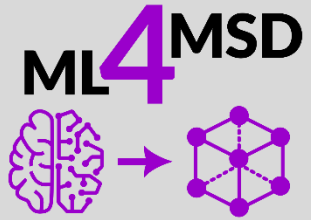


ME 5374-ST



# Machine Learning for Materials Science and Discovery

Fall 2025

Asst. Prof. Peter Schindler

## Lecture 12 – Project Discussions

- Student Paper Presentations
- Individual Final Projects
- Group Projects: Studying the Out-of-distribution Generalizability with *MatFold*

# Student Paper Presentations

- Pick a **recent publication** (or few related ones) in *Materials + AI Domain*
- Focus on topics we have not covered in class or interesting applications
- Sign up for presentation slots (two available dates):  
<https://docs.google.com/spreadsheets/d/1ZbPV3X7NpgXuJB6EMVzSrTUzBinEtFteVY0FivvmCDw/>  
First come first serve (need to select paper to sign up).
- I will approve each paper in the Google Sheet
- **Sign-up and Paper Selection Deadline:** 10/24 at 11:59 pm
- Duration: 15 minutes (12 min presentation + 3 minutes Q&A)

# Timeline of Remainder of Semester

|    |            |   |     |
|----|------------|---|-----|
| 7  | Tue, 10/14 | <i>Project Discussions</i>  | HW3 |
| 7  | Thu, 10/16 | Computational Materials Properties 1                                |     |
| 8  | Tue, 10/21 | Computational Materials Properties 2                                | HW4 |
| 8  | Thu, 10/23 | Deep Learning   |     |
| 9  | Tue, 10/28 | Machine Learning Interatomic Potentials (MLIPs)                     | HW5 |
| 9  | Thu, 10/30 | Large-Language Models (LLMs) and Generative AI in Materials Science |     |
| 10 | Tue, 11/04 | <b>Student Paper Presentations</b>                                  | HW6 |
| 10 | Thu, 11/06 | <b>Student Paper Presentations</b>                                  |     |
| 11 | Tue, 11/11 | Veteran's Day (no class)  |     |
| 11 | Thu, 11/13 | Prof. Peter Traveling (no class)                                    |     |
| 12 | Tue, 11/18 | <i>Guest Speaker: Prof. Robin Walters</i>                           |     |
| 12 | Thu, 11/20 | <i>Guest Speaker</i>  |     |
| 13 | Tue, 11/25 | <i>Guest Speaker</i>  |     |
| 13 | Thu, 11/27 | Thanksgiving recess (no class)                                      |     |
| 14 | Tue, 12/02 | <b>Final project presentations (in-class)</b>                       |     |
| 14 | Thu, 12/04 | <b>Final project presentations (in-class)</b>                       |     |
| 15 | Tue, 12/09 | <b>Final project presentations (in-class)</b>                       |     |

← 10/24: Paper selected

# Individual Final Project – Option 1

- Select **target materials class** (e.g., 2D materials, perovskites, metal oxides) and **target materials property** (can be experimental or computational)
- Conduct **literature review** of available datasets and prior ML efforts
- Collect relevant datasets and perform data cleaning  
Matminer or Materials Project data may **not** be used (all other sources ok)
- Carry out full ML pipeline with **two featurizers**

# Individual Final Project – Option 2

- Identify **atomistic simulation** problem/system
- Conduct **literature review** on appropriate modeling approach
- Simulate this system utilizing **ML Interatomic Potentials (MLIPs)**
- This can either involve *molecular dynamics* workflows  
or DFT-like property predictions involving *total energy predictions*

# Individual Final Project – Deliverables

## Short Project Summary:

2-3 pages summarizing relevant literature and high-level ML pipeline (data cleaning, featurization, ML algorithm, hyperparameters, performance metrics) or MLIP approach (model chosen, summary of model predictions)

## GitHub Repository:

- Contains the entire code to recreate the ML pipeline, plots, and results.
- Contains a summary JSON or CSV file with all main performance metrics.
- Contains PNGs of all relevant plots.

## Final Presentation:

25 minutes (20 minutes presentation + 5 minutes Q&A)

# Individual Final Project – Timeline

## **Checkpoint 1: 10/30 at 11:59 pm**

Dataset/property/system chosen and signed up in this Google Sheet:

[https://docs.google.com/spreadsheets/d/1BcFXLQ-K0qbLobCU2slovq\\_rKZaYY1Na4R\\_zt89JC08/](https://docs.google.com/spreadsheets/d/1BcFXLQ-K0qbLobCU2slovq_rKZaYY1Na4R_zt89JC08/)

Once completed, can sign up for a presentation slot (first come, first serve)

## **Checkpoint 2: 11/14 at 11:59 pm**

Share detailed list of relevant literature and dataset source

## **Pre-Submission and Feedback: 12/1 at 11:59 pm**

Submit summary report and Repository

I will give you feedback by **12/3**

## **Revisions Due: 12/10 at 11:59 pm [final, hard deadline]**

# Timeline of Remainder of Semester

|    |            |   |     |                                      |
|----|------------|---|-----|--------------------------------------|
| 7  | Tue, 10/14 | <i>Project Discussions</i>  | HW3 |                                      |
| 7  | Thu, 10/16 | Computational Materials Properties 1                                |     |                                      |
| 8  | Tue, 10/21 | Computational Materials Properties 2                                | HW4 |                                      |
| 8  | Thu, 10/23 | Deep Learning   |     | ← 10/24: Paper selected              |
| 9  | Tue, 10/28 | Machine Learning Interatomic Potentials (MLIPs)                     | HW5 |                                      |
| 9  | Thu, 10/30 | Large-Language Models (LLMs) and Generative AI in Materials Science |     | ← 10/30: Final Project CP 1          |
| 10 | Tue, 11/04 | <b>Student Paper Presentations</b>                                  | HW6 |                                      |
| 10 | Thu, 11/06 | <b>Student Paper Presentations</b>                                  |     |                                      |
| 11 | Tue, 11/11 | Veteran's Day (no class)  |     |                                      |
| 11 | Thu, 11/13 | Prof. Peter Traveling (no class)                                    |     | ← 11/14: Final Project CP 2          |
| 12 | Tue, 11/18 | <i>Guest Speaker: Prof. Robin Walters</i>                           |     |                                      |
| 12 | Thu, 11/20 | <i>Guest Speaker</i>  |     |                                      |
| 13 | Tue, 11/25 | <i>Guest Speaker</i>  |     |                                      |
| 13 | Thu, 11/27 | Thanksgiving recess (no class)                                      |     | ← 12/1: Final Project Submission     |
| 14 | Tue, 12/02 | <b>Final project presentations (in-class)</b>                       |     |                                      |
| 14 | Thu, 12/04 | <b>Final project presentations (in-class)</b>                       |     |                                      |
| 15 | Tue, 12/09 | <b>Final project presentations (in-class)</b>                       |     | ← 12/10: Final Project Revisions Due |



# In Distribution (ID) vs. Out-of-Distribution (OOD) Generalization

## Training data



Images generated with DALL-E

*“Is this a cat?”*

**ID:**



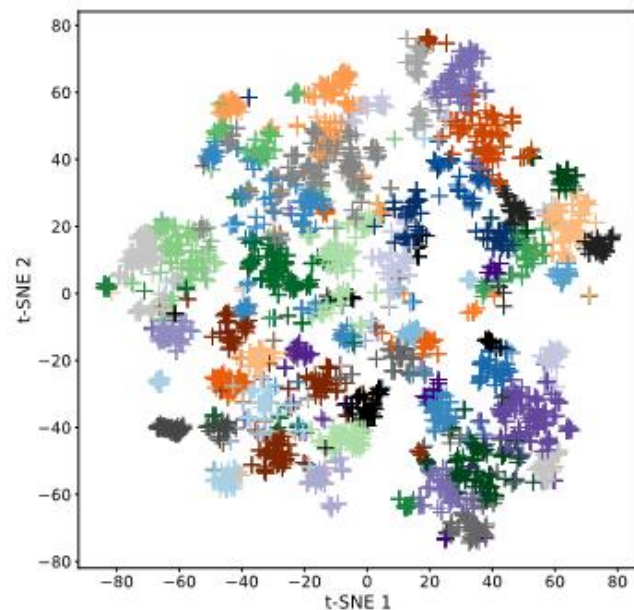
**OOD:**



*“Epistemic”*: Lack of knowledge or sub-optimal model architecture

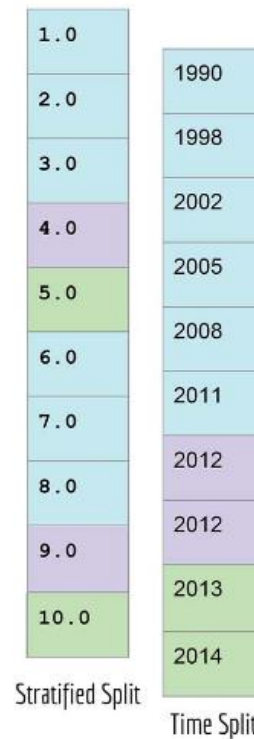
# Assessing OOD Generalizability in Materials Science

**Feature Space:  
LOCO-CV, KDE Distance, Range**



Omee, et al. npj Comput. Mater. 2024, 10, 144.

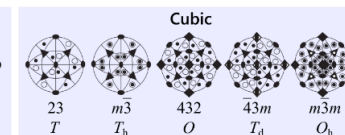
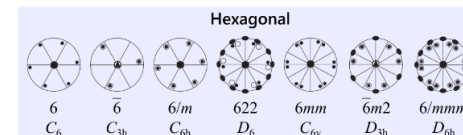
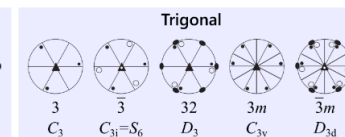
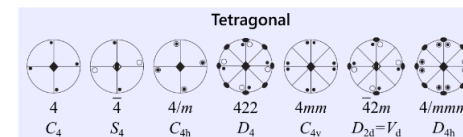
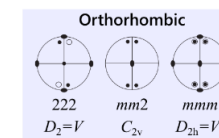
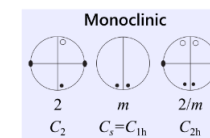
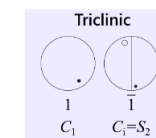
**Target Property Range  
or Time-based**



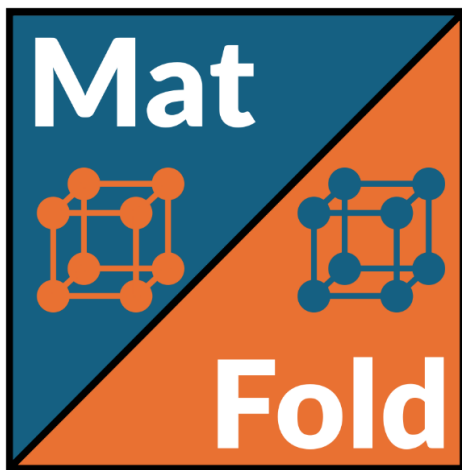
Wu et al. Chem. Sci. 2018, 9, 513–530

**Structure and Chemistry**

| Group  | 1        | 2     | 3     | 4     | 5     | 6     | 7     | 8     | 9     | 10    | 11    | 12    | 13    | 14     | 15     | 16     | 17     | 18     |
|--------|----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|--------|--------|
| Period | 1<br>1 H |       |       |       |       |       |       |       |       |       |       |       |       |        |        |        |        | 2 He   |
| 2      | 3 Li     | 4 Be  |       |       |       |       |       |       |       |       |       |       | 5 B   | 6 C    | 7 N    | 8 O    | 9 F    | 10 Ne  |
| 3      | 11 Na    | 12 Mg |       |       |       |       |       |       |       |       |       |       | 13 Al | 14 Si  | 15 P   | 16 S   | 17 Cl  | 18 Ar  |
| 4      | 19 K     | 20 Ca | 21 Sc | 22 Ti | 23 V  | 24 Cr | 25 Mn | 26 Fe | 27 Co | 28 Ni | 29 Cu | 30 Zn | 31 Ga | 32 Ge  | 33 As  | 34 Se  | 35 Br  | 36 Kr  |
| 5      | 37 Rb    | 38 Sr | 39 Y  | 40 Zr | 41 Nb | 42 Mo | 43 Tc | 44 Ru | 45 Rh | 46 Pd | 47 Ag | 48 Cd | 49 In | 50 Sn  | 51 Sb  | 52 Te  | 53 I   | 54 Xe  |
| 6      | 55 Cs    | 56 Ba | 57 La | 58 Ce | 59 Pr | 60 Nd | 61 Pm | 62 Sm | 63 Eu | 64 Gd | 65 Tb | 66 Dy | 67 Ho | 68 Er  | 69 Tm  | 70 Yb  | 71 Lu  | 72 Hf  |
| 7      | 87 Fr    | 88 Ra | 89 Ac | 90 Th | 91 Pa | 92 U  | 93 Np | 94 Pu | 95 Am | 96 Cm | 97 Bk | 98 Cf | 99 Es | 100 Fm | 101 Md | 102 No | 103 Lr | 104 Rf |



# The *MatFold* Python Module



| Options                  | Abbr. | Possibilities  |
|--------------------------|-------|--|
| Data Fraction            | $D$   | $\mathbb{R} \in (0, 1]$  |
| Default Train Assignment | $T$   | {None, Elemental, Binary, Ternary, ...}  |
| Split Method             | $S$   | { $K$ -fold, $(K, L)$ -fold}<br>$K, L \in \mathbb{N}^+$ (fixed or LOO)   |
| Criteria (outer)         | $C_K$ | {Random, Structure, Composition, Chemsys, Element, PT Group, PT Row, Space Group, Point Group, Crystal System} |
| Criteria (inner)         | $C_L$ | {Random, $C_K$ }   |

- **Featurization-agnostic** chemical/structural splits
- **Lightweight**: pymatgen and pandas are the only dependencies
- **Easily installable** and **open-source**: `pip install MatFold`
- JSON settings file → **Reproducible** splits

# Group Projects

Studying the OOD Performance on Matbench datasets & models

## Sign-up Sheet:

<https://docs.google.com/spreadsheets/d/1t5ng06gqtRehIXX5sPPvjQBLRUJYs8R08H5M7KeGByw/>

## Roles:

### *Leaders:*

Install ML algorithm on *Explorer* HPC and run models on dataset splits.  
If successful and results will be published → Co-author on publication.

### *Members:*

Create dataset splits with MatFold for several Matbench datasets.  
Help with analysis of final results.

# Group Projects – Timeline

## **10/23: Group Project Teams Finalized**

→ Models selected and dataset splits distributed

## **11/6: Group Leader Check-in 1**

→ Ensure that things are set up on Explorer HPC

## **11/20: Group Check-in 2**

→ All splits have been generated and supplied to leaders

→ Models are working on Explorer HPC

## **12/5: Group Project Deliverables are Due**

# Group Projects – Deliverables

## GitHub Repository:

- ML code and split generation JSON files
- Description of how environment can be set up on *Explorer*

## Splits and Trained Models:

Dataset splits and trained model files uploaded to external hosting (e.g., Onedrive)

## Analysis Summary:

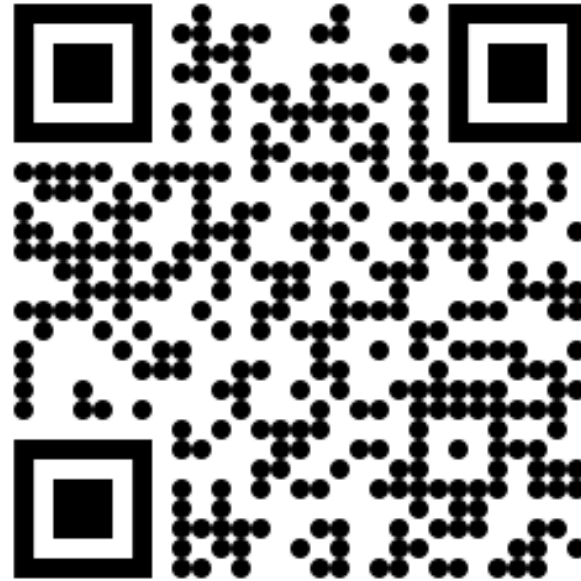
- 1-2 page summary report of performance across different OOD splits
- JSON file summarizing all main metrics



# Timeline of Remainder of Semester

|    |            |   |     |  |
|----|------------|---|-----|--|
| 7  | Tue, 10/14 | <i>Project Discussions</i>  | HW3 |  |
| 7  | Thu, 10/16 | Computational Materials Properties 1                                |     |  |
| 8  | Tue, 10/21 | Computational Materials Properties 2                                | HW4 | ← 10/23: Group Project Teams Finalized       |
| 8  | Thu, 10/23 | Deep Learning   |     | ← 10/24: Paper selected                      |
| 9  | Tue, 10/28 | Machine Learning Interatomic Potentials (MLIPs)                     | HW5 |  |
| 9  | Thu, 10/30 | Large-Language Models (LLMs) and Generative AI in Materials Science |     | ← 10/30: Final Project CP 1                  |
| 10 | Tue, 11/04 | <b>Student Paper Presentations</b>                                  | HW6 |  |
| 10 | Thu, 11/06 | <b>Student Paper Presentations</b>                                  |     | ← 11/6: Group Leader Check-in 1 (Explorer)   |
| 11 | Tue, 11/11 | Veteran's Day (no class)  |     |  |
| 11 | Thu, 11/13 | Prof. Peter Traveling (no class)                                    |     | ← 11/14: Final Project CP 2                  |
| 12 | Tue, 11/18 | <i>Guest Speaker: Prof. Robin Walters</i>                           |     |  |
| 12 | Thu, 11/20 | <i>Guest Speaker</i>  |     | ← 11/20: Group Check-in 2 (Splits Finalized) |
| 13 | Tue, 11/25 | <i>Guest Speaker</i>  |     |  |
| 13 | Thu, 11/27 | Thanksgiving recess (no class)                                      |     | ← 12/1: Final Project Submission             |
| 14 | Tue, 12/02 | <b>Final project presentations (in-class)</b>                       |     |  |
| 14 | Thu, 12/04 | <b>Final project presentations (in-class)</b>                       |     | ← 12/5: Group Projects Due                   |
| 15 | Tue, 12/09 | <b>Final project presentations (in-class)</b>                       |     | ← 12/10: Final Project Revisions Due         |

# Lecture Feedback



Please, scan the QR code and take a minute to let me know how the lecture was and mention any **feedback/questions**

This form is **anonymous!**