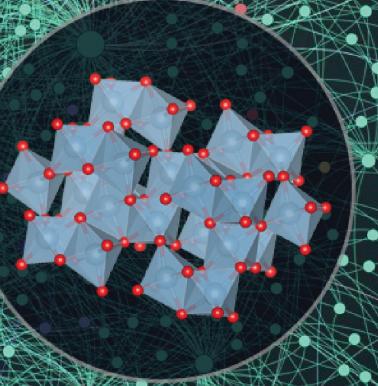
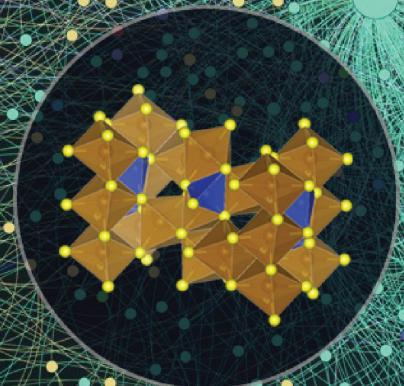


# materials informatics best practices



# We recently published the definitive best practices guide for ML in MSE



Search text, DOI, authors, etc.



My Activity

Publications



RETURN TO ISSUE

< PREV

METHODS/PROTOCOLS

NEXT >

## Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices

Anthony Yu-Tung Wang, Ryan J. Murdock, Steven K. Kauwe, Anton O. Oliynyk, Aleksander Gurlo, Jakoah Brgoch, Kristin A. Persson, and Taylor D. Sparks\*

Cite this: *Chem. Mater.* 2020, 32, 12, 4954–4965

Publication Date: May 19, 2020

<https://doi.org/10.1021/acs.chemmater.0c01907>

Copyright © 2020 American Chemical Society

[RIGHTS & PERMISSIONS](#)

Article Views

7626

Altmetric

48

Citations

45

[LEARN ABOUT THESE METRICS](#)

Share

Add to

Export



[Read Online](#)

[PDF \(2 MB\)](#)

[open URL](#)

Supporting Info (1) »

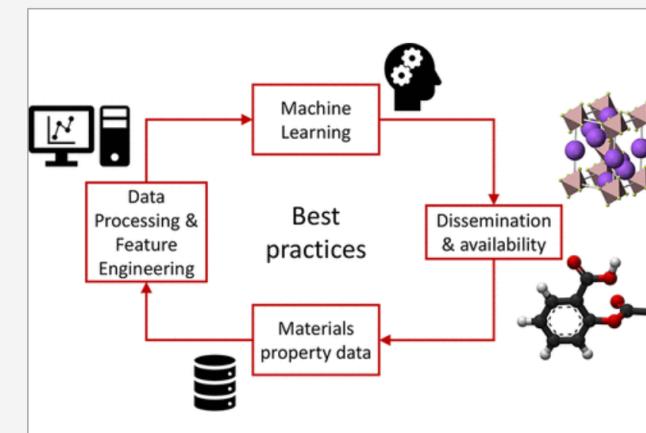
SUBJECTS: Materials science, ▾



Chemistry of Materials

### Abstract

This Methods/Protocols article is intended for materials scientists interested in performing machine learning-centered research. We cover broad guidelines and best practices regarding the obtaining and treatment of data, feature engineering, model training, validation, evaluation and comparison, popular repositories for materials data and benchmarking data sets, model and architecture sharing, and finally publication. In addition, we include interactive Jupyter notebooks with example Python code to demonstrate some of the concepts, workflows, and best practices discussed. Overall, the data-driven methods and machine learning workflows and considerations are presented in a simple way, allowing interested readers to more intelligently guide their machine learning research using the suggested references, best practices, and their own materials domain expertise.



# Jupyter notebooks are up on GitHub and in our worked examples folder

The screenshot shows a GitHub repository page for 'BestPractices' (Public) owned by anthony-wang. The repository has 1 issue, 1 pull request, and 28 forks. The code tab is selected, showing a list of files and their commit history:

File	Description	Last Commit
data	Pluto notebooks 1-3 completed. Some testing provided to ensure reaso...	13 months ago
notebooks	Improve imputing of missing values with column medians	14 months ago
pluto_notebooks	Pluto notebooks 1-3 completed. Some testing provided to ensure reaso...	13 months ago
.gitignore	Initial commit	2 years ago
LICENSE.md	Initial commit	2 years ago
README.md	Pluto notebooks 1-3 completed. Some testing provided to ensure reaso...	13 months ago
conda-env.yml	Update pandas-profiling package required version	14 months ago

The README.md file contains the following content:

## BestPractices

Things that you should (and should *not*) do in your Materials Informatics research.

This is a repository containing the relevant Python code and Jupyter notebooks to the publication "Machine Learning for Materials Scientists: An Introductory Guide toward Best Practices".

These notebooks are included to illustrate a hypothetical Machine Learning project in materials science created following best practices. The goal of this project is to predict the heat capacity of materials given a chemical composition and condition (the measurement temperature).

**About**

Things that you should (and should not) do in your Materials Informatics research.

[doi.org/10.1021/acs.chemmater.0c01907](https://doi.org/10.1021/acs.chemmater.0c01907)

python data-science machine-learning  
jupyter best-practices neural-networks  
materials-science materials-informatics  
jupyter-notebooks example-code  
common-pitfalls interactive-notebooks

Readme MIT License 61 stars 5 watching 28 forks

Releases No releases published

Packages No packages published

# ensemble techniques

