To-do **Important points (value)** General text

1. Introduction/angle?
   1. Current Machine Learning and issues of interpretability

Over the past two decades, advances in machine learning (ML) have enabled its application to a wide variety of problems.1 Machine learning has been used to predict properties,2 search for novel materials,3,4 and even guide experimental procedures.5 Critically, ML can be applied in situations where theoretical understanding is incomplete or prohibitively complex. Examples of this type of application include crystal structure prediction,5 structure-property relationships [????]

Despite the impressive accuracy in the predictive power of ML methods, ML algorithms are still treated largely as black boxes. Not only does this make trusting any given ML model difficult, but also prevents a translation of said model into physical understanding.

* 1. Description of current stance toward failure and learning
  2. **We care more about the knowledge than the data. Data is useless**
  3. Inverse problem: People have used physics to build models, might make sense to learn in the opposite way
  4. Interpretability? Move to make ML for MSE more interpretable….
  5. Benefits beyond raw predictive power
     1. Quick exposure of anomalous data points
     2. Interpretability methods can give guidelines for local spaces and sometimes more globally
  6. Final Thesis:

In this work, we use density data for both crystalline and amorphous metallic glasses to show how careful development of machine learning models can provide physical insight into the system in question.

1. Description of methods
   1. Dataset

Two primary datasets were employed in this work. Densities for ~ 700 metallic glasses were collected from literature. Although the dataset is small for most machine learning applications, it provides a convenient initial sandbox. For comparison, crystalline data computed by the Materials Project was collected. To constrain this database, we took only compounds within the ternaries featured in our metallic glass database.

* 1. Features

Critical to this approach is the featurization method. All compounds were featurized using the matminer python library.6 [Explain features and how model fits all of them initially]

* + 1. Virtue in simple features: quick, vectorizable. Matminer good for this
    2. Captures information/trends made apparent from decades of theoretical studies. (periodic table)
  1. Models
     1. Various models can and do work. RF works for this system, and highly interpretable.
     2. Do other models invalidate these approaches? Interpretability isn’t the greatest in other models
        1. For the purposes of understanding what ML is discovering, choosing an easily interpretable model is expedient
        2. Alternate methods exist, such as Global surrogate (training an interpretable model on the output of an uninterpretable one), LIME,
           1. <https://christophm.github.io/interpretable-ml-book/lime.html>

1. Results
   1. MG info/models
      1. ~~By knowing physics, can capture knowledge (density effectively capturing 34 features)~~
      2. When the physics is lacking, it is possible to turn to ML. *But physics matters*
      3. ML also captures things slightly differently… Vegards law is not exactly reproduced
      4. ~~Building on top of existing models is akin to a boosting, but not on top of an existing ML model.~~

In building machine learning models, it is good to have a performance benchmark to compare any models to. For better understood phenomenon, perhaps traditional models provide adequate descriptions. In the case of density, Vegard’s law can be used to estimate the density of an arbitrary compound given the density of pure elements. Throughout this work we will make this comparison. Initial attempts to fit a random forest model to the metallic glass density data perform slightly better than Vegard’s law (Figure A) once outliers have been addressed (see supplemental), although only slightly. As random forest models are easily interpretable, the features used can be ranked by their importance. Because the features provided by matminer are derived physical quantities (atomic weight, covalent radius, melting temperature), this provides us the opportunity to gain insight into the physical system in question. For example, examining the first 10 features (Table 1a) might suggest that the density metallic glasses are primarily dependent on the atomic weight of their constituent elements. This makes intuitive sense for metallic glasses, which lack long range order as amorphous materials. While the specific case example of amorphous metals may not be particularly interesting, it provides a framework for gaining physical insight from other machine learning models.

[why do we care for small improvements? Look at features and interpretability?]

If we are to leverage feature importances to gaining physical insight, models must be built and optimized with this in mind. To this end, we refine our metallic glass density model with two methods: feature elimination[?] and (stacking models) a form of boosting.[?] Feature elimination is a necessary step for interpretable models, as it both improves model accuracy (reduces over-fitting) [ref about overfitting?] and simplifies interpretation of feature importances. By removing invariant features and performing recursive feature elimination,[] the number of relevant features for the metallic glass model is reduced from 152 to 14. Models trained with this subset of features show an improvement in R2 of 0.3% (Figure Ab) along with a much richer set of feature importances (Table Ab). Here the atomic number remains the most important feature, but other important quantities such as the melting point and electronegativity are revealed.

A second interesting method for model refinement is a stacked model. Typically referred to as “boosting,” this involves training the model on the errors from past models. While in traditional machine learning this is performed on past models[?], in this case we can build upon models grounded in experimental or first-principles theory. For the case of density, we can perform boosting using Vegard’s law, essentially asking the model to understand why some densities do not follow a simple law of mixing. (it is important to note that feature elimination should be performed again here). The stacked model shows a 4.3% improvement in R2 over previous generations (Figure Ac), and requires 110 features to form the most accurate model (Table Ac). The most important features are largely unique to those seen in direct predictions, and feature many quantities derived from multiple elements (average deviation in Column, mean bandgap). These ensemble quantities might suggest interactions between elements are the main cause of deviations from Vegard’s simple rule of mixtures, but critically the machine learning model is able to capture this behavior.

* 1. CR on MG data

While these models have worked well on amorphous materials, crystalline materials are a more interesting test case. As expected, the metallic glass model does not perform well when predicting the density of crystalline materials, implying a different subset of features must be relevant. A new model was built using the same methodology as the amorphous glass case, incorporating outlier removal, feature elimination, and finally a stacked model. The results and feature importances are summarized in Figure B and Table B. With Vegard’s law performing significantly worse on a crystalline dataset, building and understanding a machine learning model is even more important.

[comments about about features, which are important, why this is interesting?]

* + 1. ~~Whatever vegards law captures does not work as well on crystals.~~
    2. Look at all the models? Look at the best MG model on CR data?
    3. ~~Motivation to follow same flowchart~~
    4. Would be nice to link features to some established theory for BMG density. Basic intuition is that densities should be lower
  1. CR models/info
     1. **Outliers suggest vegards law is doing poorly. 🡪 Vegards law clearly does not capture open structures 🡪 physics insight**
     2. Maybe poke around at outliers specifically here
     3. Maybe vegards doesn’t do as well but ML models do well overall?
  2. Acknowledge that all of this is contingent on:
     1. Matminer allowing us to featurize the insights of the periodic table
     2. RF allows us to interpret this model
  3. Establish extensibility to other properties.

1. Typographical errors
   1. Outliers can skew some models ability to fit. RF should be rather agnostic.
   2. Possibility of ML model being able to reveal errors, as in scatter plot
   3. *Get deviation and define outliers in terms of deviation 🡪 justification for placing outliers in a class of its own.* 
      1. *Could say vegards law made deviation more obvious*
   4. This seems to be a bit of a given. As in this should happen before any model, and is not a novel contribution?
      1. Is the schema/categorization novel? Maybe, most errors seem to assume data is perfect (TPR, FPR, ROC, R2, etc). Could be worth emphasizing this in the materials science space
         1. If this is the case, we could propose other ways of mitigating (downsampling, weighting)
2. Physics/modeling errors
   1. **Features can be examined and learned from 🡪 Interpretability**
      1. Feature importances are “gini importance”
         1. Purity at each node.
      2. Feature elimination can be efficiently performed using available packages
         1. Vital if assessing feature importances of any sort (as in this work)
      3. LIME allows for local perturbations , but global importance can be more difficult to figure out
   2. **Predicting difference between a simple model and real data can be useful**
      1. Can improve model significantly (as in density/vegards)
      2. Models perform differently, how to incorporate the two?
         1. Track various outliers and see where they end up 🡪 can we combine performances 🡪 deviation model
      3. RF is thus being forced to use information from vegard’s law
      4. Can identify features that are different 🡪 point to physical reasons why we might deviate from a naïve model
   3. We see similar effects on data gathered from Materials project for density
      1. Suggests extensibility
   4. Vegard’s law does very well for metallic glasses, can we think about them as hard spheres? What does vegard’s law really mean. Physical insight here…
      1. ML appears to do well, but in a different way
      2. Vegard’s law takes the mixture of essentially close packed/hard sphere densities.
3. Conclusions/caveats
   1. This methodology requires interpretable models. These are useful but not always the most powerful. One could make the case that some less interpretable models are more effective (NN’s?) but strides are being made [?].

Messages:

-applying ML gives insights, show how you get insights.

-Best models are not always most useful. **Paper shows how failures of ML models are important to extracting physics insights.**

-“can a failed (ML) model teach you physics?”

Literature, machine learning

1. G. Pilania, R. Ramprasad, Accelerating materials property predictions using machine learning, Scientific Reports 2013
   1. Material -> Fingerprint vector -> Chemical differences -> Machine learning
   2. Fingerprint vector similar to matminer features, possibly.
   3. Kernel Ridge regression (KRR)
2. Yi Wang, W.; Li, J.; Liu, W.; Liu, Z. K. Integrated Computational Materials Engineering for Advanced Materials: A Brief Review. *Comput. Mater. Sci.* **2019**, *158* (September 2018), 42–48.
   1. Interesting timeline
3. Ribeiro, M. T.; Singh, S.; Guestrin, C. “Why Should I Trust You?” In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16*; ACM Press: New York, New York, USA, 2016; pp 1135–1144.
   1. Idea of explanations, making models interpretable
   2. Local Fidelity is important, and simplification provides trust
   3. Fits a separate model to local area of item in question, not attempting to approximate globally
   4. Focus on trust as main benefit of technique
4. Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533* (7601), 73–76.
   1. Use ML trained on failures to improve prediction of reaction success
   2. MOF (Metal organic frameworks)
   3. Converts SVM to decision tree for interpretability
5. Dolgirev, P. E.; Kruglov, I. A.; Oganov, A. R. Machine Learning Scheme for Fast Extraction of Chemically Interpretable Interatomic Potentials. *AIP Adv.* **2016**, *6* (8).
   1. Proposes combined scheme with ML and physical models, attempting to create a more robust and interpretable model.
6. Mueller Reviews recent advances in ML for MSE
   1. https://books.google.com/books?hl=en&lr=&id=dRqRCgAAQBAJ&oi=fnd&pg=PA186&dq=machine+learning+materials+science&ots=q8vTFFaqmy&sig=2vWwPvtNQ9GEtt5Pe8EXvDXH\_9k#v=onepage&q=machine%20learning%20materials%20science&f=false

(1) Jordan, M. I.; Mitchell, T. M. Machine Learning: Trends, Perspectives, and Prospects. *Science (80-. ).* **2015**, *349* (6245), 255–260.

(2) Zeng, M.; Kumar, J. N.; Zeng, Z.; Savitha, R.; Chandrasekhar, V. R.; Hippalgaonkar, K. Graph Convolutional Neural Networks for Polymers Property Prediction. **2018**.

(3) Ren, F.; Ward, L.; Williams, T.; Laws, K. J.; Wolverton, C.; Hattrick-Simpers, J.; Mehta, A. Accelerated Discovery of Metallic Glasses through Iteration of Machine Learning and High-Throughput Experiments. *Sci. Adv.* **2018**, *4* (4), eaaq1566.

(4) Jonayat, A. S. M.; van Duin, A. C. T.; Janik, M. J. Discovery of Descriptors for Stable Monolayer Oxide Coatings Through Machine Learning. *ACS Appl. Energy Mater.* **2018**, *1*, acsaem.8b01261.

(5) Raccuglia, P.; Elbert, K. C.; Adler, P. D. F.; Falk, C.; Wenny, M. B.; Mollo, A.; Zeller, M.; Friedler, S. A.; Schrier, J.; Norquist, A. J. Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* **2016**, *533* (7601), 73–76.

(6) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; et al. Matminer: An Open Source Toolkit for Materials Data Mining. *Comput. Mater. Sci.* **2018**, *152* (April), 60–69.