Designing Monolayer Films for Nanoscale Lubrication Using Machine Learning

Ákánké D. Mason-Hogans[1], Co D. Quach[2,3], Justin B. Gilmer[3,4], Clare McCabe[2,3,5]
[1] Department of Chemical, Biological, and Bioengineering, North Carolina A&T State University, Greensboro, NC, 27411
[2] Department of Chemical and Biomolecular Engineering, Vanderbilt University, Nashville, TN, 37235, USA
[3] Multiscale Modeling and Simulation Center
[4] Interdisciplinary Materials Science, Vanderbilt University, Nashville, TN, 37235, USA
[5] Department of Chemistry, Vanderbilt University, Nashville, TN, 37235, USA

Introduction

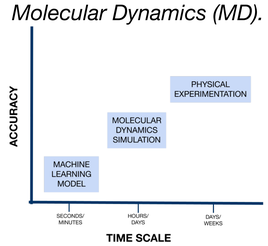
Motivation

Micro- and nano-scale devices have a wide variety of applications in technology, ranging from hard drives to medical & automotive equipment. The inadequate lubrication of the interacting surfaces within these systems often leads to significant surface damage which can prevent the entire system from operating at its full potential[1,2]. Current methods for identifying potential monolayers with ideal lubrication characteristics are computationally expensive and impractical because of the large design space[4]. Therefore, a new approach that screens certain tribological properties like Coefficient of Friction (COF) and Adhesive Force (F_0) is needed.

Primary Goals

Our primary goals are to use a Machine Learning (ML) model in order to:

- Efficiently screen through numerous monolayers and provide a rapid estimation of their tribological properties.
- Contribute to a greater overall understanding of what properties most influence the behavior of these monolayers,
- *Identify the top 20 monolayer systems to be further investigated through more thorough simulations like Molecular Dynamics (MD).*



Physical experimentation and MD simulations produce more reliable results than a ML model would, but an ML model takes significantly less time and is ideal for screening. candidate films.

References
1. Cummings PT., Docherty H, Iacovella CR., Singh JK., 2010
2. Vilt SG, Leng Z, Booth BD, McCabe C, Jennings GK. JPC C., 113, 2009;
3. Summers, A. Z., Gilmer, J. B., Iacovella, C. R., Cummings, P. T., McCabe, C. J. Chem.
Theory Comput. 16 (3), 1779–1793. 2020
4. Quach CD, et. al. in prep.
5. Davies M, Nowotka M, Papadatos G, Dedman N, Gaulton A, Atkinson F, Bellis L, Overington JP.— Nucleic Acids Res. 2015; 43(W1):W612-20. doi: 10.1093/nar/gkv352
6. Sklearn Random Forest Regressor ¶. scikit. (n.d.). https://scikit-learn.org/stable/modules/generated/
7. Landrum, G. 2010. "RDKit." Q2. https://www.rdkit.org/.

Methodology

Designing the Monolayer Systems

The ML model used for this project utilizes the Sklearn module of the random forest regressor algorithm[6] in order to predict the tribological properties of the potential lubricants that were designed. The systems consisted of two interacting surfaces:

- 1 homogenous monolayer film on the top
- A different homogenous monolayer film on the bottom

The 586 chemistries from the ChEMBL small molecules library[5] that we used to design these monolayers had:

- Molecular weights ranging from 4-99g/mol
- Exclude metallic elements/ complexes

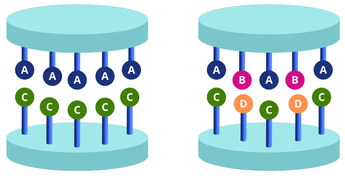


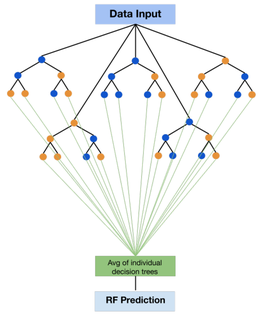
Figure 1: Homogeneous monolayers

Figure 2: Heterogeneous monolayers

This resulted in the creation of 171,698 unique systems that were screened through our model, that resemble Figure1. Figure2 represents heterogeneous monolayers that will be studied in future studies.

Training the ML Model

Our group has previously examined a set of 19 small molecules under similar conditions and calculated their properties via MD simulations[3]. Our ML model identifies patterns from those predictions, and provides estimates for the Coefficient of Friction (COF) and Adhesive Force (F_0) for each system using the Random Forest Regressor algorithm.



This algorithm constructs several decision trees simultaneously to get a variety of values, and takes the average of all of these values in order to get a general prediction.

Using Cheminformatics

The input for the training of the ML model comes from the RDKit cheminformatics library[7], which generates many different molecular descriptors of the small molecules from the ChEMBL library[5]. After training the random forrest algorithm, the model returned predictions for each system's tribological properties. It identified patterns based from the similarities it found to the training set & estimated a COF and F_0 for each system.

Screening the Monolayer Systems

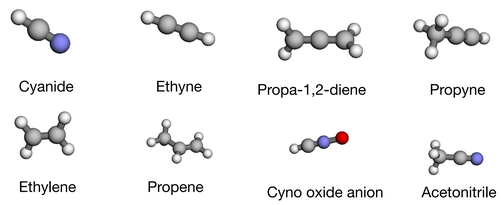
After the ML model was built, all 171,698 systems were iterated through and the properties were estimated, the monolayers were screened according to their projected performance as lubricants. The top 20 systems, identified by having the lowest combined COF & F_0, were identified and their features were further examined in order to begin designing better monolayers for nanoscale lubricants.

Results

Below is a list of the top 20 systems determined by our ML model. Notice that several terminal groups are repeated in more than one system, suggesting that they have qualities worthy of further exploration.

	Molecule in Top Monolayer	Molecule in Bottom Monolayer	COF	F0 (nN)
1	Ethylene	Cyanide	0.103716	0.776285
2	Propane	Cyanide	0.114494	0.725015
3	Cyclopropane	Cyanide	0.115051	0.463096
4	Methane	Cyanide	0.115282	0.54772
5	Ethylene	Ethylene	0.117885	0.777512
6	1,1-difluoroethene	Ethyne	0.118004	0.769939
7	Ethane	Cyanide	0.12066	0.649091
8	Acetonitrile	Ethylene	0.123267	0.790122
9	Cyclopropane	Cyno oxide anion	0.12364	0.712391
10	1,1-difluoroethene	Ethylene	0.124396	0.730926

	Molecule in Top Monolayer	Molecule in Bottom Monolayer	COF	F0 (nN)
11	Vinyl chloride	Propyne	0.125261	0.786784
12	Propionitrile	Ethylene	0.125718	0.800954
13	Methane	Cyno oxide anion	0.125959	0.770377
14	Vinyl chloride	Propa-1,2-diene	0.126333	0.803459
15	1,1-difluoroethane	Cyanide	0.126488	0.728034
16	Ethane	Ethyne	0.127037	0.724671
17	1,1-difluoroethene	Propene	0.127129	0.728936
18	Propyne	Ethylene	0.127331	0.79702
19	2,2-difluoro acetamide	Propane	0.128004	0.742966
20	Propane	Ethyne	0.128101	0.777746

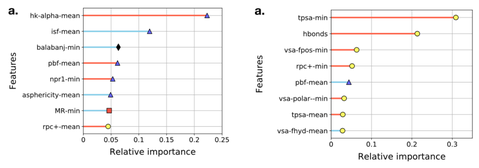


Note: The above illustrations do not show double/ triple bonds that are present in some of the molecules.

Above is a set of visualizations of some of the molecules present in the top 20 systems. Note that the majority of these molecules are relatively small, straight, and pointed. They are also not expected to be very polar.

Our findings were consistent with what was hypothesized based on the previous study:

- Larger molecules with bulkier shapes like [list molecules] exhibited the highest COFs.
- More polar molecules with a greater charge distribution exhibited the highest F_0.
- Smaller/ cylindrical, less polar molecules like [list molecules] performed the best, exhibiting low COF & F_O values.



The above graphs[3] illustrate the importance of various features as they relate to the performance of the monolayers that were used in the training set of 19 small molecules. Our findings were consistent with these results.

Conclusions & Discussion

The current project was restricted to the limitations of what could be measured and studied over the course of 10 weeks, but the scope of the project can be easily increased to continue to design better nanoscale lubricants.

The current study only screened small nonmetallic molecules with MW ranging from 4-99g/mol, but larger dataframes can easily be loaded into our predictions model. In the future, our model will be quickly adapted to accommodate heterogeneous monolayer films, and can screen through any ratio of molecules present within these films.

After estimating the COF and F_0 for the top 20 monolayer systems, *the next step is to simulate them with more thorough methods and confirm the findings of this project*. By performing Molecular Dynamics (MD) simulations on these monolayer films and physically experimenting with them, we will be able to further substantiate the estimations made by our ML model.

Acknowledgments
This research is supported by the National Science Foundation through awards NSF-DMR 1852157 & OAC-1835874