Title

Optimal Transport Based Graph Kernels for Drug Property Prediction

Authors

Aburidi, Mohammed

Marcia, Roummel

Publication Date

2025

DOI

10.1109/ojemb.2024.3480708

Peer reviewed

Optimal Transport Based Graph Kernels for Drug Property Prediction

Mohammed Aburidi[1] **and Roummel Marcia**[1]

[1] University of California Merced, Applied Mathematics Department, Merced, 95343, USA

CORRESPONDING AUTHOR: Mohammed Aburidi (e-mail: maburidi@ucmerced.edu)

This research is partially supported by NSF Grant IIS 1741490 and DMS 1840265.

ABSTRACT —*Objective:* The development of pharmaceutical agents relies heavily on optimizing their pharmacodynamics, pharmacokinetics, and toxicological properties, collectively known as ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity). Accurate assessment of these properties during the early stages of drug development is challenging due to resource-intensive experimental evaluation and limited comprehensive data availability. To overcome these obstacles, there has been a growing reliance on computational and predictive tools, leveraging recent advancements in machine learning and graph-based methodologies. This study presents an innovative approach that harnesses the power of optimal transport (OT) theory to construct three graph kernels for predicting drug ADMET properties. This approach involves the use of graph matching to create a similarity matrix, which is subsequently integrated into a predictive model. *Results:* Through extensive evaluations on 19 distinct ADMET datasets, the potential of this methodology becomes evident. The OT-based graph kernels exhibits exceptional performance, outperforming state-of-the-art graph deep learning models in 9 out of 19 datasets, even surpassing the most impactful Graph Neural Network (GNN) that excels in 4 datasets. Furthermore, they are very competitive in 2 additional datasets. *Conclusion:* Our proposed novel class of OT-based graph kernels not only demonstrates a high degree of effectiveness and competitiveness but also, in contrast to graph neural networks, offers interpretability, adaptability and generalizability across multiple datasets.

INDEX TERMS Optimal tranpsort, ADMET properties, Wasserstein distance, Graph matching, Graph kernels.

IMPACT STATEMENT The proposed method advances drug discovery and development by employing a graph-based framework rooted in optimal transport theory. This approach facilitates enhanced ADMET drug prooperty predictions, potentially revolutionizing therapeutic advancements for targeted treatments and drug design precision.

I. INTRODUCTION

THE foundation of drug development is rooted in the discovery and refinement of therapeutic agents possessing a harmonious blend of pharmacodynamics, pharmacokinetics, and toxicological properties. These properties collectively constitute the critical realm of ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) properties [1]–[3]. They wield a profound influence over the effectiveness and safety of a drug, rendering the assessment of ADMET properties integral to the drug development process. However, navigating this terrain poses formidable challenges, particularly in the early stages of drug development. The resource-intensive and costly nature of experimental evaluations, coupled with limited accessible data, underscores the imperative need for innovative approaches and predictive models capable of streamlining drug discovery, minimizing risks, and optimizing cost-effectiveness [4], [5].

While millions of active compounds have been identified, the rate of approved new drugs has not seen significant growth in recent years. Beyond non-technical challenges, shortcomings in efficacy and safety are primary factors contributing to this stagnation, largely attributed to issues with ADMET properties. ADMET encompasses pharmacokinetic considerations, determining the ability of a drug molecule to reach its target protein within the body and its duration in the bloodstream. Concurrent assessment of the efficacy and pharmacological properties of drug candidates has become standardized, with studies on ADMET processes routinely conducted in the early stages of drug discovery to mitigate attrition rates.

In response to these challenges, computational and predictive tools have emerged as indispensable assets, harnessing the power of recent advances in machine learning and graph-based methodologies [6]–[16]. These tools hold the potential to illuminate the intricate landscape of ADMET properties, enabling more informed decision-making and facilitating a smoother transition from drug discovery to clinical realization [6], [7].

Recent strides in machine learning have forged crucial links between molecular characteristics and ADMET properties, promising more efficient and accurate prediction [17], [18]. This pivotal connection has paved the way for in-depth explorations of distribution [19], [20], regulations involved in drug metabolism [21], [22], excretion [23], and other factors paramount to drug development.

Furthermore, the spotlight in the pharmaceutical research domain has turned towards Graph Machine Learning (GML). This burgeoning field has garnered considerable attention due to its remarkable capacity to model complex biomolecular structures and incorporate multi-omic datasets effectively [24]. Graph neural networks, in particular, have made substantial contributions to the prediction of ADMET properties for small-molecule drugs [25]–[27]. However, they come with inherent challenges such as interpretability, deployment complexity, training speed, and the demand for substantial training data.

Optimal transport (OT), a mathematical framework, has recently emerged as a promising and versatile tool in the machine learning community [28]–[33]. Its geometric understanding of sample distributions and applications in various fields, including biology [34]–[37], have opened up exciting possibilities for innovative research.

Moreover, the concept of graph kernels [38] has demonstrated remarkable effectiveness in handling the complexities of graph data structures [39], [40]. In this study, we present a novel approach for predicting drug ADMET properties.

A preliminary version of this work has been reported [41], where we utilized optimal transport and developed a graph kernel method based only on the Wasserstein distance. Here, we build upon this previous work and incorporate Gromov-Wasserstein and Fused Gromov-Wasserstein distances to construct graph kernels used to guide machine learning models in prediction tasks. By performing graph matching and generating a similarity matrix, we utilize this kernel in a predictive model. Our comprehensive evaluations on 19 ADMET datasets demonstrate the promise of our models. Specifically, our OT-based graph kernel outperforms state-of-the-art graph deep learning models in 9 out of 19 datasets and performs competitively in two others. These findings underscore the potential of such methodologies to advance pharmaceutical research. Our methodology based on Fused Gromov-Wasserstein shows improvements on accuracy when compared with kernels based on Wasserstein distance. This novel class of OT-based graph kernels not only demonstrates a high degree of effectiveness but also offers interpretability, in contrast to graph neural networks, and generalizability across multiple datasets.

In this study, our research objectives are to enhance the prediction of drug ADMET properties by developing and evaluating novel methodologies based on optimal transport and graph kernels. Specifically, our objectives are to:

1) **Develop Advanced Graph Kernels:** Build upon previous work by incorporating Gromov-Wasserstein and Fused Gromov-Wasserstein distances to create new graph kernels that improve the prediction of ADMET properties [41].
2) **Evaluate Model Performance:** Assess the performance of these OT-based graph kernels against state-of-the-art graph deep learning models on 19 ADMET datasets, focusing on accuracy and interpretability.
3) **Compare and Analyze:** Compare the effectiveness of the new graph kernels with state-of-the-art methods, and analyze the improvements in predictive accuracy and generalizability, particularly highlighting advancements in accuracy and interpretability [17], [25].

These objectives aim to address the challenges in drug development by providing more effective and interpretable predictive models for ADMET properties, ultimately advancing pharmaceutical research.

II. RESULTS

We constructed the cost matrix, C, by calculating the Euclidean distance between the features at the nodes of the drug pairs. This cost matrix is employed in the optimization problem to match the two distributions. Derived from the graph W/GW/FGW distances, a similarity matrix denoted as can be formulated to be utilized into a learning algorithm. We run two tasks, classification and regression. For the classification task, we utilized Support Vector Machine (SVM) and Kernel Multi Layer Perception using the indefinite kernel matrix $e^{-\lambda M}$, which is an instance of a Laplacian kernel and seen as a noisy observation of the true positive semidefinite kernel [43]. The parameter λ can be chosen between 1 and 3. For the regression task, we used Support Vector Regression (SVR) and different architecture of Kernel Multi Layer Perception, and we utilized the same kernel. We utilized a scaffold split to replicate this distant influence. The data are partitioned into a 7:1:2 ratio for the training, validation, and test sets, with the training and validation sets being scaffold shuffled five times to generate five independent runs. In binary classification tasks, we employ AUROC (Area Under the Receiver Operating Characteristic) when the data is balanced, and AUPRC (Area Under the Precision-Recall Curve) when there are fewer positive instances compared to negatives. For regression tasks, we utilize MAE (Mean Absolute Error) and Spearman correlation, especially in benchmarks where capturing the underlying trend is more critical than the absolute error.

In the context of classification and regression using Kernel Multi-Layer Perception (KMLP), the architecture comprises two fully connected layers with 64 and 32 nodes, respectively. A Rectified Linear Unit (ReLU) activation function is applied between these layers, and the output layer size varies based on the specific task, whether regression or classification. The optimization of network parameters was performed using the Adam optimizer [44] with a learning rate of 0.001. Both networks underwent training for 200 epochs, each with a batch size of 20.

The results are presented in Table I. Our proposed models, especially W- and FGW-based frameworks, demonstrate significantly improved performance relative to current state-of-the-art graph-based techniques and various deep learning models across 9 distinct datasets. Moreover, our methods ranked among the top three methods in 11 out of 19 datasets. as shown in Figure 1. Notably, the most impactful baseline method, the GNN employing context prediction masking (CPred), performs the best on three datasets. This substantiates the adaptability and generalization capabilities of our proposed models, which are highly significant attributes within the domain of drug discovery. Several deep learning methods, including CNN and MLP based on Morgan fingerprints, did not yield the best results in any of the datasets, accentuating the robustness of our

TABLE I: Average across five runs are reported. Arrows (↑,↓) indicate the direction of better performance. The best method is in bold and the second best is underlined. The (-) symbol denotes that the method is computationally intensive and experienced extended processing duration. Our results are showed in the last three columns utilized SVM and SVR as learning algorithms.

Raw Feature Type		Expert-Curated		SMILES	Molecular Graph-Based				Proposed		
Dataset	Metric	**Morgan**	**RDKit2D**	**CNN**	**GCN**	**AttFP**	**AttrM**	**CPred**	**W**	**GW**	**FGW**
Absorption											
caco2 ↓	MAE	0.908	0.393	0.446	0.599	0.401	0.546	0.502	0.390	0.521	**0.368**
HIA ↑	AUROC	0.807	0.972	0.869	0.936	0.974	**0.978**	0.975	0.928	0.911	0.945
Pgp ↑	AUROC	0.880	0.918	0.908	0.895	0.892	**0.929**	0.923	0.882	0.830	0.886
Bioav ↑	AUROC	0.581	0.672	0.613	0.566	0.632	0.577	0.671	0.748	0.646	**0.767**
Lipo ↓	MAE	0.701	0.574	0.743	0.541	0.572	0.547	**0.535**	0.809	-	-
AqSol ↓	MAE	1.203	0.827	1.023	0.907	**0.776**	1.026	1.040	0.992	-	-
Distribution											
PPBR ↓	MAE	12.848	9.994	11.106	10.194	9.373	10.075	9.445	**8.556**	8.733	8.584
BBB ↑	AUROC	0.823	0.889	0.781	0.842	0.855	0.892	**0.897**	0.857	-	-
VD ↑	Spearman	0.493	0.561	0.226	0.457	0.241	0.559	0.485	0.722	0.412	**0.729**
Metabolism											
cyp2d6_s ↑	AUPRC	0.671	0.677	0.485	0.617	0.574	0.704	0.736	0.784	0.575	**0.814**
cyp3d4_s ↑	AUROC	0.633	0.639	**0.662**	0.590	0.576	0.582	0.609	0.641	0.639	0.651
cyp2c9_s ↑	AUPRC	0.380	0.360	0.367	0.344	0.375	0.381	0.392	0.448	0.385	**0.476**
Excretion											
Half_Life ↑	Spearman	0.329	0.184	0.038	0.239	0.085	0.151	0.129	0.372	0.269	**0.414**
CL-Micro ↑	Spearman	0.492	**0.586**	0.252	0.532	0.365	0.585	0.578	0.512	0.533	0.552
CL-Hepa ↑	Spearman	0.272	0.382	0.235	0.366	0.289	0.413	**0.439**	0.341	0.314	0.324
Toxicity											
hERG ↑	AUROC	0.736	0.841	0.754	0.738	0.825	0.778	0.756	0.779	0.762	**0.853**
AMES ↑	AUROC	0.794	0.823	0.776	0.818	0.814	**0.842**	0.837	0.789	-	-
DILI ↑	AUROC	0.832	0.875	0.792	0.859	0.886	**0.919**	0.861	0.887	0.862	0.904
LD50 ↓	MAE	0.649	0.678	0.675	0.649	0.678	0.685	0.669	**0.648**	-	-
Top-Ranked Method		0/19	2/19	1/19	0/19	1/19	4/19	3/19	2/19	0/19	**7/19**

proposed approach. The performance of our proposed method was particularly noticeable in predicting properties related to metabolism, distribution, and toxicity.

The three methods based on OT kernels exhibited diverse performance, with the GW method unexpectedly demonstrating the lowest performance among the three methods. Conversely, the model based on FGW outperformed the other two, and the Wasserstein-based model was generally somewhere in between. These findings underscore the significance of integrating both atomic-level features and the topological features of molecules, along with information on how atoms are connected, when constructing a graph-based learning model.

In our experiments, we applied the Wasserstein distance-based kernel to all 19 datasets. However, we noted that for the GW and FGW-based kernels, obtaining distances was computationally prohibitive and infeasible for large datasets, namely: Lipo, AqSol, BBB, AMES, and LD50. As a result, we omitted these computations as shown blank ("-") in Table I. The variability in the performance of our methods across datasets is expected. In the TDC study [42], the authors observed that "the [machine learning state-of-the-art] models do not consistently perform well" across the ADMET Benchmark Group. They highlighted the contrasting approaches of GNN models, which focus on local substructures of molecular graphs, and descriptors, which consider global biochemical features. Furthermore, they suggested that integrating these diverse signals could lead to improved model performance in the future. Our experiments provide validation for this recommendation.

III. DISCUSSION

We attribute the performance of our method to the inherent strength of optimal transport theory, which offers a robust framework for capturing chemical and structural similarities between small molecule drugs. The optimal transport outputs a robust kernel that effectively discriminates between instances. Moreover, our results indicate that considering atomic-level features significantly enhances the predictive model's discriminatory capacity. This enhancement becomes evident when

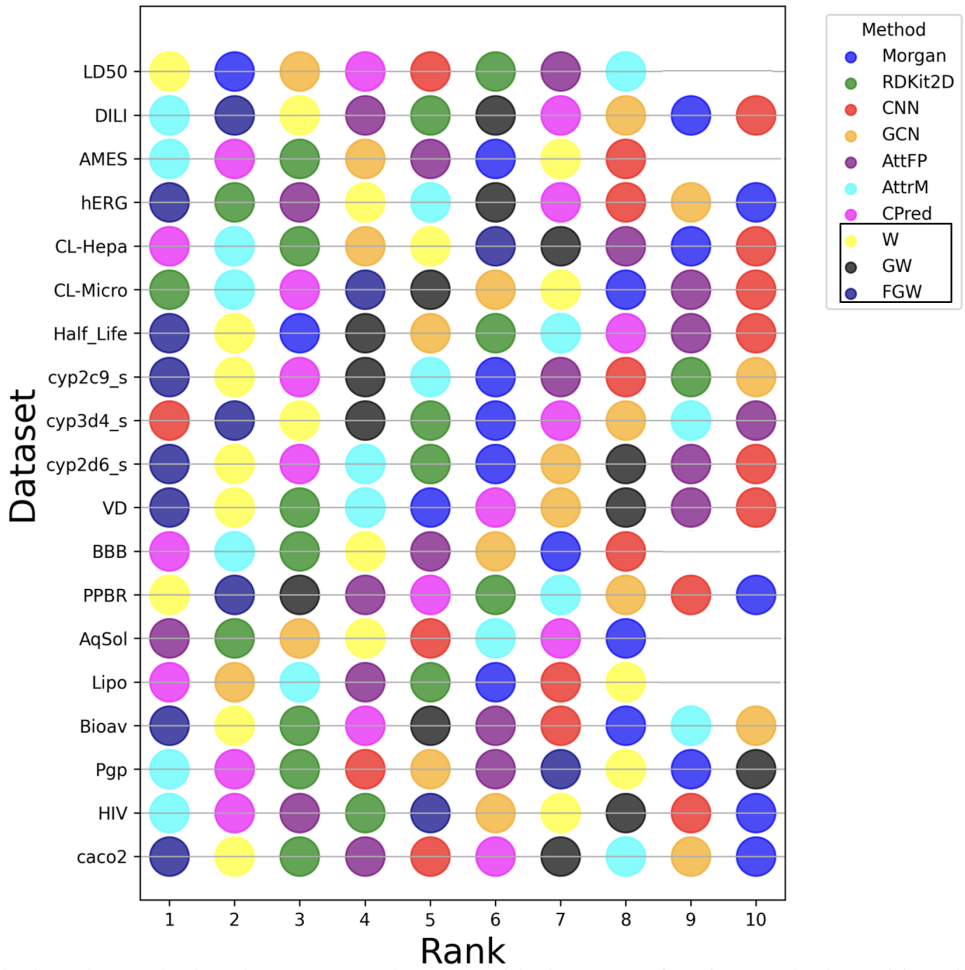

Fig. 1: This figure depicts the method ranks across 19 datasets, with the top-performing methods positioned on the left (ranked one) and the lower-performing ones gradually appearing towards the right (ranked ten).

comparing our method's outcomes with graph-level feature-based models such as MLP + Morgan features or MLP + RDKit2D. A notable limitation of the FGW-based model is its computational intensity, rendering it impractical for some large datasets. This constraint is evident in Table I, where entries for datasets requiring the solution of the quadratic optimization problem tens of thousands of times remain blank. In contrast, the Wasserstein-based model exhibited a more reasonable runtime and successfully processed all datasets, consistently delivering very competitive results. The GW and FGW distances are generally more computationally intensive than the Wasserstein distance due to its additional complexity in comparing and aligning not just the probability distributions, but also the underlying metric structures or graphs associated with these distributions. Finding the Wasserstein distance involves solving a linear programming problem, and it is computationally demanding but feasible. Whereas, finding GW and FGW introduce additional complexity, as it requires comparing the geometry or structure of the spaces, not just their mass distributions.

IV. CONCLUSIONS

Our innovative approach, harnessing optimal transport (OT) theory to formulate a graph kernel for predicting drug ADMET properties, demonstrates substantial potential and overall competitiveness compared to existing state-of-the-art methodologies. Notably, the model showcases remarkable adaptability and generalization capabilities, suggesting its capacity to significantly enhance pharmaceutical development. This study marks a significant advancement in the integration of machine learning and graph-based methodologies for the early assessment of drug properties, offering insights that pave the way for more efficient and cost-effective drug development processes. Future endeavors will focus on further developing OT-based methods to address intricate challenges in drug discovery, encompassing aspects such as protein-protein interactions, drug-drug interactions, and drug-target interactions.

V. MATERIALS AND METHODS

Graphs are an effective representation of small molecules, in which atoms are depicted as nodes, and the chemical bonds between them are represented as edges. This graph-based depiction adeptly encapsulates both the structural and chemical insights of these molecules, enabling the application of diverse

graph algorithms and machine learning methodologies to forecast their properties, interactions, and conduct within biological systems. Our approach unfolds across successive stages: (1) transformation of each drug molecule into an assembly of node embeddings based on the atomic properties of the compound, (2) quantifying the Wasserstein distance between all graph pairs, and (3) creation of a similarity matrix that to be utilized within the learning algorithm.

Now, let us delve into the formalization of the attributed graph matching problem. In a more rigorous representation, we deal with undirected labeled graphs, which can be denoted as tuples in the following structure: $\mathcal{G}(\mathcal{V},\mathcal{E},l_f,l_s)$. Here, $(\mathcal{V},\mathcal{E})$ constitutes the set of vertices and edges within the graph. Function l_f serves as a labeling mechanism, assigning each vertex $v_i \in \mathcal{V}$ a feature $a_i = l_f(v_i)$ within a specific feature metric space. Similarly, l_s maps a vertex v_i from the graph to its structure representation $s_i = l_s(v_i)$ in some structure space specific to each graph. In our particular application, this feature vector encapsulates atomic properties in a chemical compound.

Our proposal involves enhancing the aforementioned graph by introducing a histogram. This histogram is designed to convey the relative significance of vertices within the graph. To implement this, assuming the graph consists of N vertices, we assign individual weights h_i to each of these vertices. Consequently, our graph adopts the format $\mathcal{G}(\mathcal{V},\mathcal{E},l_f,l_s,h_\mathcal{G})$, where $h_\mathcal{G}$ operates as a function that links a weight to each vertex, such that $h_i = h_\mathcal{G}(v_i)$. This definition allows us to portray the graph as a probability measure characterized by comprehensive support across the feature space. In cases where all weights are equal, denoted as $h_i = \frac{1}{N}$, every vertex assumes an equivalent degree of relative importance.

Our goal is to establish a matching distance metric between two graphs, denoted as $\mathcal{G}_1$ and $\mathcal{G}_2$, with N and K vertices, respectively. Each of these graphs is uniquely characterized by its respective probability measure, $h_{\mathcal{G}_1}$ and $h_{\mathcal{G}_2}$. Our next step then involves evaluating the pairwise distance between the molecules. This process initiates with the computation of ground distances for each pair of nodes.

The mathematical formulation of the proposed methods is discussed in the supplementary material.

VI. Conflict of Interest Statement

The authors assert that they have no conflicts of interest.

VII. Author Contributions Statement

MA was responsible for coding, conducting the experiments, and performing the analysis. Both MA and RM contributed to the design of the methodology. MA and RM co-wrote the initial draft of the manuscript. All authors provided critical revisions to the manuscript and approved the final version for publication.

References

[1] J. Drews, 'Drug discovery: a historical perspective', Science, vol. 287, no. 5460, pp. 1960–1964, 2000.

[2] S. Ekins, B. J. Ring, J. Grace, D. J. McRobie-Belle, and S. A. Wrighton, 'Present and future in vitro approaches for drug metabolism', Journal of Pharmacological and Toxicological Methods, vol. 44, no. 1, pp. 313–324, 2000.

[3] R. E. White, 'High-throughput screening in drug metabolism and pharmacokinetic support of drug discovery', Annual Review of Pharmacology and Toxicology, vol. 40, no. 1, pp. 133–157, 2000.

[4] I. Kola and J. Landis, 'Can the pharmaceutical industry reduce attrition rates?', Nature Reviews Drug Discovery, vol. 3, no. 8, pp. 711–716, 2004.

[5] T. Kennedy, 'Managing the drug discovery/development interface', Drug Discovery Today, vol. 2, no. 10, pp. 436–444, 1997.

[6] H. Van De Waterbeemd and E. Gifford, 'ADMET in silico modelling: towards prediction paradise?', Nature Reviews Drug Discovery, vol. 2, no. 3, pp. 192–204, 2003.

[7] D. Stepensky, 'Prediction of drug disposition on the basis of its chemical structure', Clinical Pharmacokinetics, vol. 52, pp. 415–431, 2013.

[8] F. Sun, J. Sun, and Q. Zhao, 'A deep learning method for predicting metabolite–disease associations via graph neural network', Briefings in bioinformatics, vol. 23, no. 4, p. bbac266, 2022.

[9] T. Wang, J. Sun, and Q. Zhao, 'Investigating cardiotoxicity related with hERG channel blockers using molecular fingerprints and graph attention mechanism', Computers in biology and medicine, vol. 153, p. 106464, 2023.

[10] Z. Chen, L. Zhang, J. Sun, R. Meng, S. Yin, and Q. Zhao, 'DCAMCP: A deep learning model based on capsule network and attention mechanism for molecular carcinogenicity prediction', Journal of cellular and molecular medicine, vol. 27, no. 20, pp. 3117–3126, 2023.

[11] J. Wang et al., 'Predicting drug-induced liver injury using graph attention mechanism and molecular fingerprints', Methods, vol. 221, pp. 18–26, 2024.

[12] L. Liu, Y. Wei, Q. Zhang, and Q. Zhao, 'SSCRB: Predicting circRNA-RBP interaction sites using a sequence and structural feature-based attention model', IEEE Journal of Biomedical and Health Informatics, 2024.

[13] X. Yang et al., 'Multi-task aquatic toxicity prediction model based on multi-level features fusion', Journal of Advanced Research, 2024, https://doi.org/10.1016/j.jare.2024.06.002

[14] M. Aburidi, M. Banuelos, S. Sindi, and R. Marcia, 'Genetic Variant Detection Over Generations: Sparsity-Constrained Optimization Using Block-Coordinate Descent', in 2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2023, pp. 1–5.

[15] Y. Hang, M. Aburidi, B. Husain, A. R. Hickman, W. L. Poehlman, and F. A. Feltus, 'Exploration into biomarker potential of region-specific brain gene co-expression networks', Scientific Reports, vol. 10, no. 1, p. 17089, 2020.

[16] G. Nolte, M. Aburidi, and A. K. Engel, 'Robust calculation of slopes in detrended fluctuation analysis and its application to envelopes of human alpha rhythms', Scientific reports, vol. 9, no. 1, p. 6339, 2019.

[17] M. W. B. Trotter and S. B. Holden, 'Support vector machines for ADME property classification', QSAR & Combinatorial Science, vol. 22, no. 5, pp. 533–548, 2003.

[18] Y. Sakiyama, 'The use of machine learning and nonlinear statistical tools for ADME prediction', Expert Opinion on Drug Metabolism & Toxicology, vol. 5, no. 2, pp. 149–169, 2009.

[19] V. K. Gombar and S. D. Hall, 'Quantitative structure–activity relationship models of clinical pharmacokinetics: clearance and volume of distribution', Journal of Chemical Information and Modeling, vol. 53, no. 4, pp. 948–957, 2013.

[20] B. Louis and V. K. Agrawal, 'Prediction of human volume of distribution values for drugs using linear and nonlinear quantitative structure pharmacokinetic relationship models', Interdisciplinary Sciences: Computational Life Sciences, vol. 6, pp. 71–83, 2014.

[21] F. Cheng et al., 'Classification of cytochrome P450 inhibitors and noninhibitors using combined classifiers', Journal of Chemical Information and Modeling, vol. 51, no. 5, pp. 996–1011, 2011.

[22] H. Sun, H. Veith, M. Xia, C. P. Austin, and R. Huang, 'Predictive models for cytochrome P450 isozymes based on quantitative high throughput screening data', Journal of Chemical Information and Modeling, vol. 51, no. 10, pp. 2474–2481, 2011.

[23] A. Sedykh et al., 'Human intestinal transporter database: QSAR modeling and virtual profiling of drug uptake, efflux and interactions', Pharmaceutical Research, vol. 30, pp. 996–1007, 2013.

[24] T. Gaudelet et al., 'Utilizing graph machine learning within drug discovery and development', Briefings in Bioinformatics, vol. 22, no. 6, p. bbab159, 2021.

[25] D. Duvenaud et al., 'Convolutional networks on graphs for learning molecular fingerprints', in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2, Montreal, Canada, 2015, pp. 2224–2232.

[26] T. N. Kipf and M. Welling, 'Semi-Supervised Classification with Graph Convolutional Networks', CoRR, vol. abs/1609.02907, 2016.

[27] Z. Xiong et al., 'Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism', Journal of Medicinal Chemistry, vol. 63, no. 16, pp. 8749–8760, 2019.
[28] D. Alvarez-Melis, T. S. Jaakkola, and S. Jegelka, 'Structured Optimal Transport', arXiv [stat.ML]. 2017.
[29] M. Aburidi and R. Marcia, 'CLOT: Contrastive Learning-Driven and Optimal Transport-Based Training for Simultaneous Clustering', in 2023 IEEE International Conference on Image Processing (ICIP), 2023, pp. 1515–1519.
[30] M. Aburidi and R. Marcia, 'Optimal Transport and Contrastive-Based Clustering for Annotation-Free Tissue Analysis in Histopathology Images', in 2023 International Conference on Machine Learning and Applications (ICMLA), 2023, pp. 301–307.
[31] G. D. Canas and L. A. Rosasco, 'Learning probability measures with respect to optimal transport metrics', in Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 2, Lake Tahoe, Nevada, 2012, pp. 2492–2500.
[32] G. Peyré, M. Cuturi, and Others, 'Computational optimal transport: With applications to data science', Foundations and Trends® in Machine Learning, vol. 11, no. 5–6, pp. 355–607, 2019.
[33] M. Arjovsky, S. Chintala, and L. Bottou, 'Wasserstein generative adversarial networks', in International Conference on Machine Learning, 2017, pp. 214–223.
[34] G.-J. Huizing, G. Peyré, and L. Cantini, 'Optimal transport improves cell–cell similarity inference in single-cell omics data', Bioinformatics, vol. 38, no. 8, pp. 2169–2177, 02 2022.
[35] K. Cao, Y. Hong, and L. Wan, 'Manifold alignment for heterogeneous single-cell multi-omics data integration using Pamona', Bioinformatics, vol. 38, no. 1, pp. 211–219, Dec. 2021.
[36] P. Demetci, R. Santorella, B. Sandstede, W. S. Noble, and R. Singh, 'Gromov-Wasserstein optimal transport to align single-cell multi-omics data', bioRxiv, 2020.
[37] G.-J. Huizing, L. Cantini, and G. Peyré, 'Unsupervised Ground Metric Learning using Wasserstein Singular Vectors', arXiv [stat.ML]. 2022.
[38] S. V. N. Vishwanathan, N. N. Schraudolph, R. Kondor, and K. M. Borgwardt, 'Graph kernels', Journal of Machine Learning Research, vol. 11, pp. 1201–1242, 2010.
[39] N. Shervashidze, P. Schweitzer, E. J. van Leeuwen, K. Mehlhorn, and K. M. Borgwardt, 'Weisfeiler-Lehman Graph Kernels', Journal of Machine Learning Research, vol. 12, no. 77, pp. 2539–2561, 2011.
[40] C. Morris, N. M. Kriege, K. Kersting, and P. Mutzel, 'Faster kernels for graphs with continuous attributes via hashing', in 2016 IEEE 16th International Conference on Data Mining (ICDM), 2016, pp. 1095–1100.
[41] M. Aburidi and R. Marcia, 'Wasserstein Distance-Based Graph Kernel for Enhancing Drug Safety and Efficacy Prediction', in 2024 IEEE First International Conference on Artificial Intelligence for Medicine, Health and Care (AIMHC), 2024, pp. 113–119.
[42] K. Huang et al., 'Therapeutics Data Commons: Machine Learning Datasets and Tasks for Therapeutics', CoRR, vol. abs/2102.09548, 2021.
[43] R. Luss and A. d'Aspremont, 'Support vector machine classification with indefinite kernels', Advances in Neural Information Processing Systems, vol. 20, 2007.
[44] D. Kingma and J. Ba, 'Adam: A Method for Stochastic Optimization', International Conference on Learning Representations, 12 2014.