Multi-Task Learning Approaches for Predicting ADMET Properties: A Review

Nadezhda Diadkina[1]
[1] *University of Lisbon, Lisbon, Portugal*
ndiadkina@edu.ulisboa.pt

Abstract

Accurate prediction of ADMET properties (absorption, distribution, metabolism, excretion, and toxicity) is crucial in early drug discovery. While machine learning (ML) has become central to this task, traditional single-task approaches remain limited in generalization and scalability. Multi-task learning (MTL) offers an integrative solution by enabling joint modeling of multiple pharmacokinetic and toxicological endpoints through shared representations.

This review summarizes recent MTL architectures for ADMET prediction, including shared-layer neural networks, graph-based models, and transformer-based pipelines. We analyze commonly used datasets, endpoints (e.g., CYP inhibition, hERG toxicity, metabolic stability), and evaluation practices, highlighting architectural trends, methodological gaps, and unresolved challenges. This review aims to serve as a practical reference for researchers building robust multi-endpoint predictive models in pharmaceutical applications.

Keywords: ADMET prediction; multi-task learning; deep learning; drug discovery; pharmacokinetics; graph neural networks

1. Introduction

Accurate prediction of ADMET properties is essential for early-phase drug discovery. It has been reported that ADMET-related issues accounted for up to 50% of drug failures in the 1990s [1]. ML offers scalable alternatives to reduce cost and time in early drug discovery [2], [3], [4]. However, single-task learning (STL) approaches are often limited by fragmented datasets, poor generalization, and narrow endpoint focus [5], [6], [7]. MTL, by modeling multiple outputs simultaneously, offers a more data-efficient and scalable solution [5], [6].

This review surveys and compares recent MTL models applied to ADMET prediction. We focus on architectural strategies, datasets, targeted endpoints, and performance trade-offs, aiming to identify emerging patterns and open challenges. This review focuses on supervised MTL architectures and does not cover few-shot or reinforcement learning approaches, which are emerging in ADMET prediction.

2. Scope and Methodology

We selected and analyzed 8 representative MTL models (2014–2025), covering:

Architectures: shared-layer DNNs, GCNs, transformers, adaptive gating mechanisms, uncertainty estimation.

Datasets: Merck, Tox21, public ADMET sets, in-house pharma panels, 10M+ unlabeled pretraining sets.

Endpoints: CYP450 inhibition, hERG toxicity, metabolic clearance, oral bioavailability, logP, half-life, and more.

We summarized key attributes in a timeline and comparative table.

3. Comparative Summary

To illustrate the diversity of architectural strategies and task configurations in practice, we summarize representative MTL models for ADMET prediction developed over the past decade. These models vary in complexity, target endpoints, input modalities, and training strategies, reflecting the field's rapid evolution. The following models represent key milestones in the transition from simple STL baselines to modern, multi-endpoint MTL architectures (Table 1).

Table 1. MTL models for ADMET property prediction

Model	Architecture	Datasets	ADMET Endpoints
Merck MTL DNN [8]	Fully-connected DNN (shared hidden layers, multi-output)	Merck Kaggle challenge data (chemical descriptors)	15 assays (ADME-related properties like solubility, permeability, etc. in one model)
DeepTox [9]	Fully-connected deep neural nets (ensemble)	Tox21 challenge dataset (12k compounds, descriptors)	12 toxicity pathways (nuclear receptor and stress response assays) predicted together
Wenzel et al., 2019 [10]	Fully-connected DNN (optimized hyperparameters; "response maps" for interpretation)	Public & corporate ADME data (~50k compounds; molecular descriptors)	~5 endpoints: microsomal metabolic stability (multiple species), Caco-2 cell permeability, logD (distribution coefficient)
gTPP model [11]	Graph Convolutional Network (GCN); single model with shared graph embedding	In-house pharma dataset (enterprise-wide ADME assays)	18 "early ADME" properties (in vitro absorption, distribution, metabolism metrics across discovery pipeline)

Model	Architecture	Datasets	ADMET Endpoints
HelixADMET [12]	Self-supervised pre-trained model (molecular graph Transformer) + multitask fine-tuning	Combined public ADMET datasets; pretraining on 10M unlabelled molecules	Dozens of ADMET endpoints (flexible/extendable; e.g. CYP enzyme inhibition, clearance, toxicity, etc.), integrated via multi-task and multi-stage training
Rodríguez-Pérez et al., 2023 [11]	Multi-task graph neural network (GCN) with uncertainty estimation	In vitro clearance data for multiple species (e.g. human, rat microsomal clearance)	Intrinsic clearance in multiple species (multi-species metabolic clearance regression tasks), learned jointly
MTGL-ADMET [13]	Adaptive multi-task graph learning (GNN with primary-task-centric gating mechanism)	Multiple ADMET datasets; status theory used for task selection	Mixed endpoints (e.g. classification of CYP450 inhibition, toxicity, etc., and regression of PK properties like logP, half-life). Uses "one primary, multiple auxiliaries" paradigm
ADME-DL [14]	Sequential multi-task training pipeline (order-aware; uses GNN/Transformer foundation models)	Aggregated ADME datasets; drug-like and non-drug compound libraries	21 ADME property tasks (e.g. human absorption fraction, volume of distribution, clearance, half-life, etc.) trained in A→D→M→E sequence; plus final drug-likeness classification as evaluation task

4. Key Findings

4.1 Architectural Trends

Early models (Merck MTL, DeepTox) used simple shared-layer DNNs.

Recent approaches (HelixADMET, MTGL-ADMET, ADME-DL) adopt GNNs and transformers with self-supervised pretraining, dynamic task weighting, and sequential pipelines.

Gating and task-specific heads enhance modeling of heterogeneous endpoints.

4.2 Endpoint and Dataset Diversity

Most models address 5–20 ADMET endpoints, e.g., metabolic stability, CYP450 inhibition, hERG toxicity, and clearance.

Datasets are often noisy or imbalanced; recent models use transfer learning to overcome data scarcity.

4.3 Performance and Limitations

MTL outperforms STL for correlated tasks. Key limitations include negative transfer, task imbalance, and data sparsity.

Few models explicitly model task uncertainty or resolve inter-task conflicts. Architectural diversity continues to grow, but benchmarking remains inconsistent.

5. Conclusion

MTL models are increasingly deployed in early-stage screening pipelines by pharmaceutical companies, where simultaneous prediction of multiple endpoints accelerates lead optimization and reduces experimental overhead. MTL models for ADMET prediction show growing architectural sophistication and improved generalization.

Challenges remain in task balancing, dataset quality, and evaluation standards.

Future directions include molecular graph pretraining, adaptive task weighting, and integration of omics/structure-based features.

This review outlines current progress and offers guidance for designing scalable MTL frameworks in pharmaceutical ML.

References

[1] I. Kola and J. Landis, 'Can the pharmaceutical industry reduce attrition rates?', *Nat. Rev. Drug Discov.*, vol. 3, no. 8, pp. 711–715, Aug. 2004, doi: 10.1038/nrd1470.

[2] N. Berdigaliyev and M. Aljofan, 'An Overview of Drug Discovery and Development', *Future Med. Chem.*, vol. 12, no. 10, pp. 939–947, May 2020, doi: 10.4155/fmc-2019-0307.

[3] 'Computer-aided drug discovery: From traditional simulation methods to language models and quantum computing: Cell Reports Physical Science'. Accessed: Aug. 02, 2025. [Online]. Available: https://www.cell.com/cell-reports-physical-science/fulltext/S2666-3864(24)00648-9?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS2666386424006489%3Fshowall%3Dtrue

[4] J. Vamathevan *et al.*, 'Applications of machine learning in drug discovery and development', *Nat. Rev. Drug Discov.*, vol. 18, no. 6, pp. 463–477, June 2019, doi: 10.1038/s41573-019-0024-5.

[5] Z. Zhang, W. Yu, M. Yu, Z. Guo, and M. Jiang, 'A Survey of Multi-task Learning in Natural Language Processing: Regarding Task Relatedness and Training Methods'.

[6] Y. Zhang and Q. Yang, 'A Survey on Multi-Task Learning', Mar. 29, 2021, *arXiv*: arXiv:1707.08114. doi: 10.48550/arXiv.1707.08114.

[7] 'An Overview of Multi-Task Learning in Deep Neural NetworksThis paper originally appeared as a

blog post at http://sebastianruder.com/multi-task/index.html on 29 May 2017.', ar5iv. Accessed: Aug. 03, 2025. [Online]. Available: https://ar5iv.labs.arxiv.org/html/1706.05098

[8] G. E. Dahl, N. Jaitly, and R. Salakhutdinov, 'Multi-task Neural Networks for QSAR Predictions', June 04, 2014, *arXiv*: arXiv:1406.1231. doi: 10.48550/arXiv.1406.1231.

[9] A. Mayr, G. Klambauer, T. Unterthiner, and S. Hochreiter, 'DeepTox: Toxicity Prediction using Deep Learning', *Front. Environ. Sci.*, vol. 3, Feb. 2016, doi: 10.3389/fenvs.2015.00080.

[10] J. Wenzel, H. Matter, and F. Schmidt, 'Predictive Multitask Deep Neural Network Models for ADME-Tox Properties: Learning from Large Data Sets', *J. Chem. Inf. Model.*, vol. 59, no. 3, pp. 1253–1268, Mar. 2019, doi: 10.1021/acs.jcim.8b00785.

[11] M. Venkataraman, G. C. Rao, J. K. Madavareddi, and S. R. Maddi, 'Leveraging machine learning models in evaluating ADMET properties for drug discovery and development', *ADMET DMPK*, vol. 13, no. 3, p. 2772, June 2025, doi: 10.5599/admet.2772.

[12] A. A. Abdelwahab, M. A. Elattar, and S. A. Fawzi, 'Advancing ADMET prediction for major CYP450 isoforms: graph-based models, limitations, and future directions', *Biomed. Eng. OnLine*, vol. 24, no. 1, p. 93, July 2025, doi: 10.1186/s12938-025-01412-6.

[13] B.-X. Du, Y. Xu, S.-M. Yiu, H. Yu, and J.-Y. Shi, 'ADMET property prediction via multi-task graph learning under adaptive auxiliary task selection', *iScience*, vol. 26, no. 11, p. 108285, Nov. 2023, doi: 10.1016/j.isci.2023.108285.

[14] D. Bang, J. Kim, H. Song, and S. Kim, 'ADME-drug-likeness: enriching molecular foundation models via pharmacokinetics-guided multi-task learning for drug-likeness prediction', *Bioinformatics*, vol. 41, no. Suppl 1, pp. i352–i361, July 2025, doi: 10.1093/bioinformatics/btaf259.