Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

MULTITASK DEEP LEARNING FOR SOLUBILITY, PERMEABILITY, PROTEIN BINDING, AND CLEARANCE PREDICTION

Sven Larsson[1]*, Erik Johansson[1], Anna Nilsson[2]

1. ***Department of Pharmaceutical Informatics, Faculty of Medicine, Karolinska Institute, Stockholm, Sweden.***
2. ***Department of AI Drug Engineering, Faculty of Pharmacy, Lund University, Lund, Sweden.***

ARTICLE INFO

Received:
27 May 2025
Received in revised form:
02 September 2025
Accepted:
04 September 2025
Available online:
28 October 2025

Keywords: Multitask deep learning, ADME prediction, Solubility, Permeability, Protein binding, Clearance

ABSTRACT

Solubility, permeability, protein binding, and clearance together define the ADME profile of a drug candidate. These endpoints are often predicted by separate models, even though they depend on overlapping chemical determinants. Single-task ADME models do not fully exploit the information contained in correlated endpoints. They can also produce fragmented molecular profiles that are difficult to reconcile during lead optimization. This article describes a multitask deep neural network that learns a common molecular representation and simultaneously predicts aqueous solubility, membrane permeability, plasma protein binding, and metabolic clearance. Each output is paired with a task-specific uncertainty estimate to support model-informed decision-making. A molecular graph encoder, such as a graph attention network or graph isomorphism network, produces a shared fixed-length embedding for each compound. Four task-specific prediction heads map this representation to solubility, permeability, fraction unbound, and clearance outputs using heteroscedastic uncertainty-weighted learning.
Conceptually, the multitask model would be expected to improve information sharing across related ADME endpoints relative to equivalent single-task models. It could provide a coherent ADME profile from one molecular input without requiring separate model calls for each endpoint. A multitask approach to ADME prediction could simplify early pharmacokinetic screening by unifying several core developability endpoints in a single model. Such a framework would be especially useful for multi-parameter optimization, uncertainty-aware compound triage, and prospective medicinal chemistry design.

To Cite This Article: Larsson S, Johansson E, Nilsson A. Multitask Deep Learning for Solubility, Permeability, Protein Binding, and Clearance Prediction. Pharmacophore. 2025;16(5):1-9. https://doi.org/10.51847/rcSMQ7wfn7

Introduction

Absorption, distribution, metabolism, and excretion properties are central determinants of whether a compound can progress from biochemical activity to a viable drug candidate, and in silico ADME modelling has become a routine part of early discovery workflows [1]. Existing computational systems often treat individual endpoints as separate prediction problems, with dedicated models for properties such as plasma protein binding [2], lipophilicity-related solubility [3], and broader ADMET liability profiling [4]. This endpoint-by-endpoint approach is practical, but it can fragment the pharmacokinetic view of a molecule because each model learns from a different representation, training set, and calibration procedure. A model-oriented alternative is to construct one shared representation that supports multiple task-specific ADME outputs.
Single-task learning is inefficient when experimental data are sparse, noisy, or unevenly distributed across chemical series. Multitask deep learning was proposed in drug discovery partly because shared representations can exploit related structure–property relationships and act as regularizers when each task alone provides limited supervision [5]. Demystification studies of multitask neural networks have shown that the benefit of sharing depends on task relatedness, representation quality, and the avoidance of task dominance during training [6]. For ADME, the opportunity is especially clear because solubility, permeability, protein binding, and clearance all depend on partially overlapping features such as molecular size, polarity, ionization, hydrogen bonding, aromaticity, and lipophilicity.
Deep learning for molecular prediction has evolved from fingerprint-based multitask models toward graph neural networks, message-passing architectures, and transformer-style chemical language models. MoleculeNet established a benchmark-

Corresponding Author: Sven Larsson; Department of Pharmaceutical Informatics, Faculty of Medicine, Karolinska Institute, Stockholm, Sweden. E-mail: sven.larsson@gmail.com.

oriented view of molecular machine learning across multiple property datasets [7], while learned representation analyses demonstrated how graph encoders could capture property-relevant molecular patterns without relying only on handcrafted descriptors [8]. Multitask ADME-specific modelling has also been explored through deep featurization and integrated ADMET datasets, suggesting that shared molecular embeddings can support heterogeneous pharmacokinetic endpoints [9]. The remaining gap is a unified conceptual model that focuses specifically on four core ADME endpoints spanning aqueous exposure, membrane transport, systemic binding, and metabolic elimination.

The proposed MDL framework is a multitask deep neural network that jointly learns aqueous solubility, passive permeability, plasma protein binding, and metabolic clearance from a shared molecular representation. A graph-based or transformer-based encoder would provide the shared trunk, while task-specific heads would specialize the representation for logS, logPapp, fraction unbound, and intrinsic clearance outputs [10, 11]. The model is not intended to replace experimental ADME assays, but to provide a coherent, uncertainty-aware pharmacokinetic profile that can guide compound prioritization before resource-intensive measurements. In this sense, the architecture extends the logic of multitask molecular property prediction into an integrated ADME decision-support model [12, 13].

Background

The Four Endpoints and Their Molecular Determinants

Aqueous solubility reflects how readily a compound partitions into water, and it is strongly influenced by polarity, ionization, crystal packing, hydrogen bonding, and lipophilicity, as highlighted by graph-based solubility modelling studies [3, 14]. Passive permeability measured in systems such as Caco-2 or PAMPA depends on size, polar surface area, charge state, flexibility, and membrane partitioning, and interpretable permeability models have linked structural attention patterns to these transport-relevant determinants [15]. Plasma protein binding is commonly expressed through bound or unbound fractions, and prediction studies have emphasized hydrophobicity, acidity, aromaticity, and interaction-prone substructures as important drivers [2, 16, 17]. Metabolic clearance, often represented through intrinsic or hepatic clearance, is shaped by metabolic stability, enzyme recognition, ionization, and molecular exposure to biotransformation pathways, making it conceptually related to but not identical with the other ADME endpoints [18, 19]. **Table 1** summarizes how solubility, permeability, plasma protein binding, and metabolic clearance differ in their dominant molecular determinants while still offering opportunities for shared multitask representation learning.

Table 1. Mechanistic Rationale for Shared and Task-Specific Learning Across Four ADME Endpoints

ADME endpoint	Dominant molecular determinants	Why it can benefit from shared multitask learning	Why task-specific modeling remains necessary
Aqueous solubility	Polarity, ionization, hydrogen bonding, crystal packing, lipophilicity	Shares polarity, ionization, and lipophilicity signals with permeability and binding	Solid-state and crystal-packing effects may not be captured by other endpoints
Passive permeability	Size, polar surface area, charge state, flexibility, membrane partitioning	Benefits from shared structural features related to polarity, charge, and lipophilicity	Transport-system differences such as Caco-2 versus PAMPA can introduce assay-specific behavior
Plasma protein binding	Hydrophobicity, acidity, aromaticity, interaction-prone substructures	Shares hydrophobic and ionization-related features with solubility, permeability, and clearance	Binding-site affinity and protein-specific interactions require endpoint-specific interpretation
Metabolic clearance	Metabolic stability, enzyme recognition, ionization, biotransformation exposure	Can reuse shared structural encodings for lipophilicity, accessibility, and reactive substructures	Enzyme-mediated metabolism is mechanistically distinct and may create negative-transfer risk

Single-Task Deep Learning for ADME

Single-task deep learning models have been applied to individual ADME endpoints using graph convolution, message passing, recurrent SMILES encoders, and transformer-derived chemical representations. Graph attention mechanisms have been used to push molecular representation learning beyond fixed descriptors for medicinal chemistry tasks [10], and graph convolutional surveys have described their relevance to computational drug development more broadly [20]. Transformer and chemical language approaches, including large-scale molecular representation learning, provide a complementary route in which SMILES or molecular tokens are pre-trained before being adapted to property prediction [11]. These single-task approaches can be powerful, but each endpoint-specific model learns its own representation and may not use the cross-endpoint information available in related ADME assays.

Multitask and Multi-Output Learning in Drug Discovery

Multitask and multi-output learning share a common encoder across related prediction targets and then attach task-specific heads for individual outputs. Early pharmaceutical multitask work demonstrated the practical appeal of shared deep networks for drug discovery applications [5], while later QSAR analyses clarified that gains are not automatic and depend on how tasks interact within the representation [6]. In ADME modelling, multitask deep featurization has been proposed as a way to improve generalization by allowing related pharmacokinetic tasks to inform a common latent space [9]. However, negative transfer

remains a risk when one endpoint dominates the shared representation or when mechanistically distinct tasks force incompatible molecular features into the same trunk [13].

Heteroscedastic Uncertainty and Task Balancing

A multitask ADME model must combine losses from endpoints with different scales, noise levels, and label availability, so a fixed uniform weighting scheme may be inappropriate. Heterogeneous drug discovery data imputation work has shown that deep learning can operate over incomplete assay matrices when the loss is defined only for observed labels and the representation is shared across related outputs [21]. Adaptive auxiliary task selection provides another conceptual route, in which the model controls which tasks support a target endpoint rather than assuming that all tasks are equally helpful [13]. A heteroscedastic formulation would learn task-specific noise parameters so that noisier or less reliable endpoints contribute proportionally to the total objective while still benefiting from shared molecular information.

Datasets and Benchmarks for ADME Multitask Learning

ADME multitask learning requires curated datasets in which compound structures and endpoint labels can be aligned across solubility, permeability, protein binding, and clearance assays. Public benchmarks such as MoleculeNet have shaped evaluation practices in molecular machine learning [7], while ADMETlab 2.0 illustrates the value of broad integrated ADMET resources for practical property prediction [4]. Industrial-scale studies emphasize that real ADME data are often heterogeneous, sparse, and collected under varying experimental protocols, which complicates direct pooling across endpoints [19, 22]. Therefore, a conceptual multitask model should assume careful endpoint harmonization, scaffold-aware evaluation, and explicit separation between representation learning and performance reporting.

Model Development Overview

High-Level Architecture

The proposed architecture consists of a shared molecular encoder followed by four task-specific dense prediction heads. The encoder could be a graph attention network, a graph isomorphism network, a message-passing network, or a chemical transformer, reflecting the range of molecular representation strategies developed for property prediction [8, 10, 11]. The shared trunk learns features that may be useful across ADME endpoints, while separate heads allow solubility, permeability, protein binding, and clearance to preserve endpoint-specific structure–property relationships. This design follows the general principle of multitask deep learning in which common lower layers encode reusable chemical information and upper layers specialize to individual tasks [5, 12].

Figure 1 illustrates the proposed multitask ADME prediction architecture, showing how a standardized molecular input is encoded into a shared representation, specialized through four endpoint-specific prediction heads, and converted into an uncertainty-aware compound triage profile.

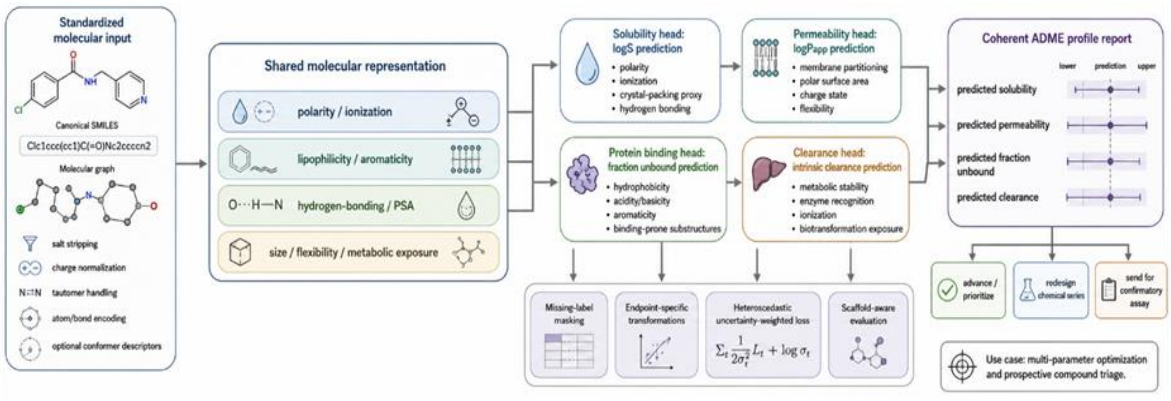

Figure 1. Multitask Deep Learning Architecture for Integrated ADME Prediction with Shared Molecular Representation, Endpoint-Specific Heads, and Uncertainty-Aware Compound Triage

Core Inputs and Outputs

The main input is a standardized molecular structure represented as a graph, canonical SMILES string, or both, and the model could optionally be extended to include conformer-derived descriptors where three-dimensional effects are important. The outputs are logS for aqueous solubility, logPapp for passive permeability, fraction unbound for plasma protein binding, and intrinsic clearance for metabolic stability or clearance profiling, each paired with a task-specific uncertainty estimate. Protein binding models focusing on fraction unbound [2, 17] and clearance-oriented developability models [19] demonstrate why endpoint-specific output transformations may be needed even when all predictions arise from one shared representation. The model would therefore treat each endpoint as a related but distinct regression problem rather than forcing all outputs into a single undifferentiated numerical space.

Design Principles

The model is guided by five design principles: a shared trunk, task-specific heads, missing-label masking, uncertainty-aware learning, and evaluation against appropriate single-task baselines. Multitask molecular studies indicate that the shared representation can improve generalization when related tasks provide complementary signal [5, 6], while ADME-focused multitask and adaptive task-selection approaches show why task relationships should be handled carefully rather than assumed [9, 13]. Missing-label handling is essential because most compounds would not have measurements for all four endpoints, and heterogeneous data imputation frameworks provide a conceptual precedent for training with incomplete assay matrices [21]. Dynamic weighting or learned task uncertainty would further prevent high-variance endpoints from overwhelming lower-noise tasks during optimization.

Data Sources and Endpoint-Specific Preprocessing

Curation of ADME Datasets

A multitask ADME dataset would be curated by combining public repositories, literature-derived measurements, and benchmark sources covering solubility, permeability, protein binding, and clearance. ChEMBL-scale modelling has demonstrated how large bioactivity and property collections can support deep learning when records are standardized and filtered [23], while ADMET platforms illustrate how diverse endpoint collections can be integrated for prediction workflows [4, 24]. Endpoint values should be harmonized conceptually into logS for solubility, log-transformed apparent permeability, fractional or transformed fraction unbound, and clearance-related measures such as intrinsic clearance. Because assay formats, species, protein matrices, and experimental conditions vary, the curation process should preserve metadata that can later be used for stratified evaluation or task-specific filtering [22].

Handling Endpoint-Specific Missing Data

Most molecules in an integrated ADME matrix would have labels for only a subset of the four endpoints, so missingness must be treated as a modelling condition rather than as an exception. Deep imputation studies in heterogeneous drug discovery data show that neural networks can learn from incomplete target matrices when observed entries are used and missing entries are masked from the loss [21]. In the proposed MDL model, each task head contributes to the objective only when the corresponding experimental label is present, while the shared encoder receives gradients from all available endpoints for that molecule. This design avoids discarding partially labelled compounds and allows a solubility-only compound, for example, to still improve the shared chemical representation used by permeability, protein binding, and clearance heads.

Input Standardization

Input standardization begins with molecular canonicalization, salt stripping, charge normalization, tautomer handling, and removal of inconsistent or ambiguous structures. Learned molecular representation studies show that graph encoders and message-passing networks are sensitive to how atoms, bonds, charges, aromaticity, and stereochemical features are encoded [8, 25]. SMILES-based transformer approaches similarly depend on tokenization, canonical or randomized SMILES choices, and consistency between pre-training and downstream prediction representations [11]. For endpoints influenced by shape, ionization, or intramolecular interactions, conformer-aware extensions could be considered, but the base MDL framework remains centered on standardized two-dimensional structure to maintain broad applicability across datasets.

Multitask Deep Learning Architecture

Molecular Encoder

The molecular encoder is the shared backbone of the MDL model, converting each compound into a continuous embedding that can support all four ADME prediction heads. A graph isomorphism network, message-passing neural network, or graph attention model could encode atoms and bonds through iterative neighborhood aggregation, following the representation-learning principles developed for molecular property prediction [8, 10, 25]. Chemical language models such as large-scale transformer representations offer an alternative or complementary encoder in which molecular strings are converted into contextual embeddings before task-specific adaptation [11]. The shared embedding should capture features that generalize across endpoints, such as polarity, aromaticity, size, charge distribution, and hydrogen-bonding patterns, while leaving final endpoint specialization to the heads.

Task-Specific Heads and Loss Functions

Each task-specific head is a shallow neural module attached to the shared embedding, allowing the model to specialize its final mapping for solubility, permeability, protein binding, and clearance. Regression heads for solubility and permeability can use robust losses appropriate for continuous physicochemical endpoints, while clearance and protein binding heads may require transformations that reflect skewed distributions or bounded fractional values, as suggested by fraction-unbound modelling studies [2, 17]. Multitask ADME models should also allow task-specific activation functions or output transforms so that each endpoint remains chemically interpretable after inverse transformation [9, 19]. This separation between shared representation and task-specific output logic helps reduce negative transfer while preserving the efficiency of joint learning.

Table 2 defines how the proposed multitask model separates shared molecular learning from endpoint-specific ADME specialization across solubility, permeability, protein binding, clearance, and task uncertainty.

Table 2. Endpoint-Specific Learning Logic in the Proposed Multitask ADME Model

ADME endpoint	Primary prediction target	Dominant molecular determinants	Why shared representation is useful	Why task-specific specialization remains necessary	Recommended output transformation	Decision-use interpretation
Aqueous solubility	logS or experimentally harmonized solubility value	Polarity, ionization state, hydrogen bonding, lipophilicity, molecular size, crystal-packing-related structural proxies	Shares physicochemical signal with permeability and protein binding because polarity, charge, and lipophilicity influence multiple ADME behaviors	Solubility is strongly affected by solid-state and formulation-relevant factors that may not map directly onto membrane transport or metabolic clearance	Log-scale solubility prediction with endpoint-specific standardization	Identifies compounds likely to suffer from poor aqueous exposure before synthesis or assay prioritization
Passive permeability	logPapp, PAMPA permeability, or Caco-2-derived apparent permeability	Polar surface area, charge state, flexibility, molecular size, lipophilicity, membrane partitioning potential	Benefits from the same polarity–lipophilicity representation learned for solubility and protein binding	Permeability depends on transport-context assumptions and assay system differences that require endpoint-specific calibration	Log-transformed apparent permeability with assay-aware harmonization	Supports early ranking of compounds for absorption potential and flags structures needing permeability optimization
Plasma protein binding	Fraction unbound or transformed fraction unbound	Hydrophobicity, acidity/basicity, aromaticity, molecular volume, ionization, protein-interaction-prone substructures	Shares hydrophobicity, aromaticity, and charge-related features with permeability and clearance	Fraction unbound is bounded, matrix-dependent, and biologically distinct from solubility or passive transport	Bounded, logit-style, or otherwise stabilized fraction-unbound transformation	Helps estimate free exposure and prioritize compounds with interpretable binding-related liabilities
Metabolic clearance	Intrinsic clearance, hepatic clearance proxy, or metabolic-stability-derived clearance measure	Metabolic soft spots, enzyme recognition motifs, ionization, lipophilicity, molecular exposure to biotransformation pathways	Gains from shared chemical features such as lipophilicity, size, aromaticity, and ionization that also influence binding and permeability	Clearance may depend on enzyme-specific recognition, species, and assay context, making it the endpoint most vulnerable to negative transfer	Log-transformed clearance or normalized intrinsic-clearance output	Supports early elimination-risk assessment and selection of compounds for metabolic stability testing
Task uncertainty	Endpoint-specific uncertainty estimate paired with each prediction	Assay noise, chemical-domain coverage, missing-label density, endpoint heterogeneity, scaffold novelty	Allows all tasks to inform confidence patterns across related chemical regions	Each endpoint has different experimental variability and label completeness, so uncertainty must remain task-specific	Learned heteroscedastic task-noise parameter or calibrated predictive interval	Distinguishes confident prioritization candidates from compounds requiring confirmatory ADME measurement

Uncertainty-Weighted Multitask Loss

The combined multitask objective can be expressed conceptually as $L = \sum\left(\frac{1}{2\sigma^2}L_t + log\sigma_t\right)$ here (L_t) is the observed-label loss for task (t) and (σ_t) is a learned task-noise parameter. This formulation allows the model to reduce the influence of noisier endpoints without manually fixing task weights, which is important when solubility, permeability, protein binding, and clearance differ in scale, assay variability, and label completeness. Heterogeneous assay learning and adaptive task-selection studies support the broader principle that multitask models should not assume equal contribution from every endpoint under every data condition [13, 21]. The resulting architecture would therefore learn both molecular representations and task-balancing parameters within a single end-to-end optimization framework.

Handling Heterogeneous Data and Missing Labels

Dealing with Various Scales and Distributions

The four ADME endpoints differ in numerical range, experimental noise, and distributional shape, so a shared model should transform outputs into forms that are suitable for neural regression. Solubility and permeability are commonly represented on logarithmic scales, while clearance may require log transformation to avoid domination by extreme values and protein binding may require a bounded or logit-style transformation when expressed as fraction unbound [2, 17]. Integrated pharmacokinetic modelling studies emphasize that harmonization is not only a numerical issue but also a biological one, because assay context, species, matrix, and endpoint definition can alter the meaning of each label [18, 22]. The MDL model should therefore learn

in standardized endpoint spaces while preserving the ability to map predictions back to experimentally interpretable ADME quantities.

Incorporating Task-Specific Uncertainty
Each prediction head should output both an endpoint estimate and an associated uncertainty term, allowing the model to distinguish confident predictions from those made in sparse or chemically unfamiliar regions. This is especially important for ADME profiling because experimental noise and coverage vary across endpoints, and heterogeneous drug discovery datasets often contain partially observed or assay-dependent labels [21]. A task-adaptive uncertainty framework would allow the model to express higher uncertainty for endpoints that are noisier, less represented, or less supported by related tasks, rather than forcing all predictions to appear equally reliable. In practical use, this uncertainty would guide whether a compound should be advanced, deprioritized, or selected for confirmatory measurement rather than treated as a definitive simulated result.

Multi-Fidelity Learning Opportunity
The same multitask architecture could be extended to multi-fidelity learning by combining high-confidence experimental measurements with lower-confidence sources such as approximate in vitro assays, legacy screens, or computational annotations. Broad ADMET prediction platforms and benchmarking resources show that property prediction increasingly depends on integrating evidence from diverse sources rather than relying on a single homogeneous dataset [4, 24, 26]. A multi-fidelity extension could attach metadata-aware losses or auxiliary heads so that gold-standard clearance, screening-level solubility, and literature-derived permeability values contribute differently to the shared representation. This strategy would be expected to improve coverage while still allowing the model to recognize that not all labels carry the same evidential weight.

Model Interpretability and Profiling
Understanding the Shared Representation
Interpretability is essential for a multitask ADME model because medicinal chemists need to know whether the same structural features are driving several endpoints or whether a predicted liability is endpoint-specific. Explainable molecular modelling work has shown how attribution methods can highlight substructures associated with preclinical properties [27], and graph attention approaches can expose atom- or bond-level patterns that influence property predictions [10]. In the proposed MDL framework, SHAP, integrated gradients, attention inspection, or counterfactual molecular edits could be applied separately to the shared encoder and each task head. This would help identify whether features such as aromatic hydrophobic groups, polar substituents, or ionizable centers are contributing consistently to solubility, permeability, protein binding, and clearance.

Generating an ADME Profile Report
For a candidate compound, the model would generate a concise ADME profile report containing predicted solubility, permeability, fraction unbound, clearance, and uncertainty estimates for each endpoint. ADMET-AI and related platforms illustrate the value of packaging molecular property prediction into accessible workflows for large-scale library evaluation [26], while multi-objective optimization tools show how ADMET outputs can be connected to compound prioritization logic [15]. The proposed report should not present predictions as experimental substitutes, but as decision-support estimates that identify favorable regions, likely liabilities, and uncertainty-driven follow-up needs. By presenting all four endpoints together, the report would support medicinal chemistry trade-off analysis more directly than isolated single-endpoint predictions.

Integration Into Drug Discovery Workflow
Early Triage and Multi-Parameter Optimization
The MDL model could be deployed as a batch-screening tool or web service that scores virtual libraries for balanced solubility, permeability, protein binding, and clearance profiles. Multi-objective ADMET optimization frameworks demonstrate how predicted properties can guide compound design when multiple developability constraints must be considered simultaneously [15, 28]. In early triage, compounds with favorable predicted profiles and low uncertainty could be prioritized for synthesis or testing, while compounds with conflicting endpoints could be examined through structural attribution before redesign. This use case emphasizes ranking, profiling, and hypothesis generation rather than claiming that the model alone can determine compound success.

Interface with Physiologically-Based Pharmacokinetic Models
Predicted ADME endpoints from the MDL model could also serve as upstream inputs for physiologically based pharmacokinetic or pharmacometric simulations. Pharmacokinetic developability models and industrial ADMET workflows have highlighted the need to connect molecular property prediction with later decision frameworks rather than treating machine learning outputs as isolated scores [19, 29]. Solubility and permeability may inform absorption assumptions, fraction unbound may influence distribution and free exposure, and clearance predictions may support elimination-related parameterization. Such a pipeline would remain conceptual unless prospectively validated, but it would provide a coherent bridge from chemical structure to mechanistic pharmacokinetic reasoning.

Evaluation Strategy

Per-Task Performance Metrics

The model should be evaluated separately for each endpoint using metrics appropriate to continuous ADME prediction, such as RMSE, MAE, R^2, Pearson correlation, or rank-based measures when prioritization is the primary goal. Benchmarking studies such as MoleculeNet established the importance of standardized splits and endpoint-specific reporting in molecular machine learning [7], while large-scale comparisons of deep learning and conventional methods show why multiple metrics are needed to avoid overinterpreting one performance view [30]. Scaffold-based and temporal splits should be preferred over random splits where possible because they better reflect prospective chemical generalization. The evaluation should compare the MDL architecture conceptually against matched single-task models, classical baselines, and published benchmark approaches without reporting invented numerical outcomes.

Multitask Benefit Analysis

A multitask benefit analysis should ask whether each endpoint improves, remains unchanged, or suffers when trained jointly with the other ADME tasks. Previous multitask QSAR studies show that shared learning can help when tasks are related but can also fail when task relationships are weak or imbalanced [5, 6]. ADME-specific multitask and industrial-scale studies further suggest that the benefit of sharing may depend on endpoint pairing, chemical overlap, assay heterogeneity, and the amount of missing data [9, 29]. The analysis should therefore include endpoint-wise comparisons, ablation of auxiliary tasks, and checks for negative transfer rather than assuming that a larger multitask model is automatically superior.

Table 3 consolidates the key analytical safeguards required to evaluate whether a multitask ADME model is reliable, interpretable, uncertainty-aware, and suitable for prospective medicinal chemistry decision support.

Table 3. Analytical Safeguards for Multitask ADME Prediction and Deployment Readiness

Analytical risk	Why it matters in multitask ADME prediction	Diagnostic analysis to include	Model-design safeguard	Deployment implication for medicinal chemistry
Negative transfer across endpoints	Shared learning may improve one endpoint while degrading another, especially if clearance or protein binding depends on features not useful for solubility or permeability	Compare each multitask head against matched single-task baselines; perform auxiliary-task ablation; evaluate endpoint-pair effects	Use deeper task-specific heads, adaptive task weighting, selective sharing, or task-gating mechanisms	Do not assume the integrated model is superior unless each endpoint shows stable or interpretable benefit
Missing and uneven endpoint labels	ADME matrices are often sparse, and many compounds may have labels for only one or two endpoints	Report label density by endpoint, overlap across tasks, missingness patterns, and scaffold coverage	Use observed-label masking so each head contributes only when its label is present while the encoder learns from all available data	Allows partially labeled compounds to support representation learning without inventing missing experimental values
Endpoint scale imbalance	Solubility, permeability, fraction unbound, and clearance occupy different numerical ranges and noise regimes	Examine loss contribution per endpoint, residual distribution, and gradient dominance during training	Apply endpoint-specific transformations and heteroscedastic uncertainty-weighted loss	Prevents high-variance or large-scale endpoints from dominating the shared encoder
Assay heterogeneity	Permeability, protein binding, and clearance values can depend on assay system, species, matrix, and protocol	Stratify performance by assay type, species, source, and measurement context when metadata are available	Preserve assay metadata; use harmonized endpoint definitions; consider metadata-aware auxiliary inputs	Predictions should be interpreted within the assay context represented in training data
Poor uncertainty calibration	A model may generate precise-looking predictions even for unfamiliar scaffolds or noisy endpoints	Compare prediction intervals with observed errors; use calibration curves and error-versus-confidence analysis	Calibrate uncertainty estimates per task and stratify by scaffold novelty or endpoint coverage	High-uncertainty predictions should trigger confirmatory assays rather than direct advancement
Weak prospective generalization	Random splits may overestimate performance by allowing close analogues across train and test sets	Use scaffold-based, temporal, and chemical-series-aware validation splits	Benchmark against single-task deep learning, classical QSAR, and simple descriptor-based baselines	Supports realistic lead-optimization use rather than retrospective memorization
Limited interpretability of shared embedding	Medicinal chemists need to know whether ADME liabilities arise from shared or endpoint-specific structural drivers	Apply SHAP, integrated gradients, graph attention inspection, and counterfactual molecular edits per endpoint	Separate shared-encoder attribution from head-specific attribution	Enables chemically actionable redesign rather than opaque compound ranking
Overreliance on in silico output	ADME predictions can guide triage but cannot replace experimental measurement	Track whether model recommendations align with later experimental assays in prospective use	Present outputs as decision-support estimates with uncertainty, not definitive assay substitutes	Positions the model as a prioritization and hypothesis-generation tool within experimental ADME workflows

Uncertainty Calibration

Uncertainty calibration should be assessed by comparing predicted uncertainty with observed prediction error across each ADME endpoint and across relevant chemical subgroups. This is conceptually aligned with modern therapeutic machine

learning frameworks that emphasize not only point predictions but also reliable decision support under uncertainty [31]. Calibration plots, residual stratification, and error-versus-confidence analyses could indicate whether the model is appropriately cautious for poorly represented scaffolds, unusual ionization states, or endpoint-specific assay regimes. For clearance and protein binding, calibration should be interpreted carefully because experimental context can contribute uncertainty that may not be fully captured by molecular structure alone [16].

Limitations

Negative Transfer across Tasks

Negative transfer is a central limitation of multitask ADME modelling because a shared encoder may learn features that help one endpoint while obscuring features needed for another. Multitask neural network analyses have shown that task relatedness, training imbalance, and representation sharing strongly influence whether joint learning improves or harms prediction [6]. In the proposed model, permeability and solubility may share physicochemical determinants, while clearance may depend on metabolic recognition patterns that are less directly aligned with passive transport or protein binding. Monitoring task-specific residuals, adding deeper task-specific layers, or using adaptive task selection would be necessary safeguards against harmful sharing.

Data Scarcity for Rare Endpoints

Rare clearance pathways, transporter-mediated phenomena, species-specific metabolism, and specialized binding contexts may be underrepresented in integrated ADME datasets. ADMET modelling reviews and prospective industrial validation studies emphasize that dataset composition, assay quality, and domain applicability can limit the reliability of machine learning predictions even when model architecture is sophisticated [1, 22]. The MDL model would therefore be most appropriate for chemical regions and assay contexts that are represented in its training data, and predictions outside this domain should be accompanied by higher uncertainty. Prospective validation remains essential before such a model is used to influence high-value lead optimization decisions.

Conclusion

A multitask deep learning model for ADME prediction would provide a unified framework for estimating solubility, permeability, plasma protein binding, and clearance from a single molecular representation. By combining a shared encoder with task-specific prediction heads, the model could capture common chemical determinants while preserving endpoint-specific specialization.

The main strengths of the proposed approach are data efficiency, coherent ADME profiling, uncertainty-aware prediction, and direct relevance to early discovery workflows. Instead of consulting separate models for each property, medicinal chemists could obtain a single integrated profile that supports multi-parameter optimization.

Important challenges remain, including negative transfer, heterogeneous assay definitions, incomplete labels, and limited coverage of rare pharmacokinetic mechanisms. These issues mean that the model should be used as decision support rather than as a replacement for experimental ADME testing.

Open-source implementations, transparent benchmark datasets, and prospective validation studies would accelerate adoption of multitask ADME modelling. A standardized evaluation culture would also make it easier to compare graph-based, transformer-based, and hybrid representations for integrated pharmacokinetic prediction.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Ferreira LL, Andricopulo AD. ADMET modeling approaches in drug discovery. Drug Discov Today. 2019;24(5):1157–65.
2. Watanabe R, Esaki T, Kawashima H, Natsume-Kitatani Y, Nagao C, Ohashi R, et al. Predicting fraction unbound in human plasma from chemical structure: improved accuracy in the low value ranges. Mol Pharm. 2018;15(11):5302–11.
3. Wieder O, Kuenemann M, Wieder M, Seidel T, Meyer C, Bryant SD, et al. Improved lipophilicity and aqueous solubility prediction with composite graph neural networks. Molecules. 2021;26(20):6185.
4. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, et al. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. Nucleic Acids Res. 2021;49(W1):W5–14.

5. Ramsundar B, Liu B, Wu Z, Verras A, Tudor M, Sheridan RP, et al. Is multitask deep learning practical for pharma? J Chem Inf Model. 2017;57(8):2068–76.
6. Xu Y, Ma J, Liaw A, Sheridan RP, Svetnik V. Demystifying multitask deep neural networks for quantitative structure–activity relationships. J Chem Inf Model. 2017;57(10):2490–504.
7. Wu Z, Ramsundar B, Feinberg EN, Gomes J, Geniesse C, Pappu AS, et al. MoleculeNet: a benchmark for molecular machine learning. Chem Sci. 2018;9(2):513–30.
8. Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing learned molecular representations for property prediction. J Chem Inf Model. 2019;59(8):3370–88.
9. Feinberg EN, Joshi E, Pande VS, Cheng AC. Improvement in ADMET prediction with multitask deep featurization. J Med Chem. 2020;63(16):8835–48.
10. Xiong Z, Wang D, Liu X, Zhong F, Wan X, Li X, et al. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. J Med Chem. 2019;63(16):8749–60.
11. Ross J, Belgodere B, Chenthamarakshan V, Padhi I, Mroueh Y, Das P. Large-scale chemical language representations capture molecular structure and properties. Nat Mach Intell. 2022;4(12):1256–64.
12. Wenzel J, Matter H, Schmidt F. Predictive multitask deep neural network models for ADME-Tox properties: learning from large data sets. J Chem Inf Model. 2019;59(3):1253–68.
13. Du BX, Xu Y, Yiu SM, Yu H, Shi JY. ADMET property prediction via multi-task graph learning under adaptive auxiliary task selection. iScience. 2023;26(11):108211.
14. Panapitiya G, Girard M, Hollas A, Sepulveda J, Murugesan V, Wang W, et al. Evaluation of deep learning architectures for aqueous solubility prediction. ACS Omega. 2022;7(18):15695–710.
15. Yi JC, Yang ZY, Zhao WT, Yang ZJ, Zhang XC, Wu CK, et al. ChemMORT: an automatic ADMET optimization platform using deep learning and multi-objective particle swarm optimization. Brief Bioinform. 2024;25(2):bbae008.
16. Li J, Yanagisawa K, Yoshikawa Y, Ohue M, Akiyama Y. Plasma protein binding prediction focusing on residue-level features and circularity of cyclic peptides by deep learning. Bioinformatics. 2022;38(4):1110–7.
17. Riedl M, Mukherjee S, Gauthier M. Descriptor-free deep learning QSAR model for the fraction unbound in human plasma. Mol Pharm. 2023;20(10):4984–93.
18. Wang X, Liu M, Zhang L, Wang Y, Li Y, Lu T. Optimizing pharmacokinetic property prediction based on integrated datasets and a deep learning approach. J Chem Inf Model. 2020;60(10):4603–13.
19. Beckers M, Sturm N, Sirockin F, Fechner N, Stiefl N. Prediction of small-molecule developability using large-scale in silico ADMET models. J Med Chem. 2023;66(20):14047–60.
20. Sun M, Zhao S, Gilvary C, Elemento O, Zhou J, Wang F. Graph convolutional networks for computational drug development and discovery. Brief Bioinform. 2020;21(3):919–35.
21. Irwin BW, Levell JR, Whitehead TM, Segall MD, Conduit GJ. Practical applications of deep learning to impute heterogeneous drug discovery data. J Chem Inf Model. 2020;60(6):2848–57.
22. Fang C, Wang Y, Grater R, Kapadnis S, Black C, Trapa P, et al. Prospective validation of machine learning algorithms for absorption, distribution, metabolism, and excretion prediction: an industrial perspective. J Chem Inf Model. 2023;63(11):3263–74.
23. Mayr A, Klambauer G, Unterthiner T, Steijaert M, Wegner JK, Ceulemans H, et al. Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. Chem Sci. 2018;9(24):5441–51.
24. Tian H, Ketkar R, Tao P. ADMETboost: a web server for accurate ADMET prediction. J Mol Model. 2022;28(12):408.
25. Heid E, Greenman KP, Chung Y, Li SC, Graff DE, Vermeire FH, et al. Chemprop: a machine learning package for chemical property prediction. J Chem Inf Model. 2023;64(1):9–17.
26. Swanson K, Walther P, Leitz J, Mukherjee S, Wu JC, Shivnaraine RV, et al. ADMET-AI: a machine learning ADMET platform for evaluation of large-scale chemical libraries. Bioinformatics. 2024;40(7):btae416.
27. Jiménez-Luna J, Skalic M, Weskamp N, Schneider G. Coloring molecules with explainable artificial intelligence for preclinical relevance assessment. J Chem Inf Model. 2021;61(3):1083–94.
28. Fralish Z, Chen A, Skaluba P, Reker D. DeepDelta: predicting ADMET improvements of molecular derivatives with deep learning. J Cheminform. 2023;15(1):101.
29. Walter M, Borghardt JM, Humbeck L, Skalic M. Multi Task ADME/PK prediction at industrial scale: leveraging large and diverse experimental datasets. Mol Inform. 2024;43(10):e202400079.
30. Korotcov A, Tkachenko V, Russo DP, Ekins S. Comparison of deep learning with multiple machine learning methods and metrics using diverse drug discovery data sets. Mol Pharm. 2017;14(12):4462–75.
31. Huang K, Fu T, Gao W, Zhao Y, Roohani Y, Leskovec J, et al. Artificial intelligence foundation for therapeutic science. Nat Chem Biol. 2022;18(10):1033–6.