

ADAPTIVE TENSOR ATTENTION NETWORKS WITH CROSS-DOMAIN TRANSFER FOR DRUG-TARGET IN- TERACTION PREDICTION

Anonymous authors

Paper under double-blind review

ABSTRACT

The prediction of drug-target interactions is fundamental to the advancement of drug discovery. We present a groundbreaking unified theory for Drug-Target Interaction prediction with Domain Adaptation (DTI-DA), seamlessly integrating concepts from quantum mechanics, differential geometry, and information theory. Our framework introduces a novel DTI symplectic structure that captures the intrinsic geometry of drug-target interactions, leading to a Quantum Optimal Transport theorem that provides a rigorous foundation for domain adaptation in the DTI context. We develop a quantum statistical mechanical formulation of DTI-DA, introducing DTI-preserving quantum channels and deriving a Quantum Wasserstein distance tailored to drug discovery applications. Our information-geometric perspective yields a Quantum Fisher-Rao metric for DTI, resulting in a quantum Cramer-Rao bound that establishes fundamental limits on DTI prediction accuracy. We propose a unified variational principle for DTI-DA, encompassing quantum and classical aspects, which leads to a novel algorithm based on geometric stochastic gradient Langevin dynamics. Furthermore, we extend classical statistical inference to the quantum domain, deriving a Quantum Rao-Blackwell theorem and a Quantum Bayesian Cramer-Rao bound specifically for DTI-DA. These theoretical advancements not only deepen our understanding of the DTI-DA problem but also suggest new algorithmic approaches with provable guarantees. Preliminary numerical experiments on quantum-inspired DTI-DA algorithms demonstrate significant improvements in prediction accuracy and domain adaptation capabilities compared to classical methods, particularly for challenging out-of-distribution scenarios in drug discovery. Our anonymous gitHub link: <https://anonymous.4open.science/r/DTI-DA-6AFB>

1 INTRODUCTION

The prediction of drug-target interactions (DTI) stands at the forefront of modern drug discovery, presenting a formidable challenge that spans the realms of biochemistry, machine learning, and, increasingly, quantum mechanics. As our understanding of molecular interactions deepens, it becomes evident that the quantum nature of these interactions plays a crucial role in determining drug efficacy and specificity. Concurrently, the vast and heterogeneous landscape of chemical and biological data necessitates sophisticated domain adaptation (DA) techniques to generalize predictions across diverse experimental settings and molecular databases.

Traditional approaches to DTI prediction, rooted in classical statistical mechanics and machine learning, have made significant strides in recent years Suruliandi et al. (2024); Dehghan et al. (2024); Gao et al. (2024); Shi et al. (2024). However, these methods often fall short when confronted with the inherent quantum mechanical aspects of molecular binding Wozniak et al. (2024) and the complex distributional shifts Bazhenov et al. (2024); Bansak et al. (2024); Conger et al. (2024) encountered in real-world drug discovery scenarios. The limitations of classical approaches become particularly apparent when attempting to model the subtle electronic interactions that govern drug-target binding or when extrapolating predictions to novel chemical spaces that differ substantially from the training distribution.

054 Recent advancements in quantum computing Fauseweh (2024); Sood & Chauhan (2024) and quantum
055 machine learning Peral-García et al. (2024); Wang & Liu (2024); Senokosov et al. (2024) have
056 opened new avenues for addressing these challenges. Quantum algorithms for molecular simula-
057 tions Cornish et al. (2024); Kesari et al. (2024) and quantum-inspired machine learning models
058 have shown promise in capturing the intricate quantum effects underlying molecular interactions.
059 However, these approaches often lack a unified theoretical framework that can seamlessly integrate
060 quantum mechanical principles with the statistical rigor required for robust domain adaptation Zhang
061 et al. (2024); Shi & Liu (2024); Li et al. (2024).

062 Our theoretical advancements have profound implications for the field of drug discovery Pinzi et al.
063 (2024); Edfeldt et al. (2024). By explicitly accounting for the quantum nature of molecular in-
064 teractions, our framework promises to improve prediction accuracy for complex drug-target sys-
065 tems Niarakis et al. (2024); Zhu et al. (2024) that have traditionally been challenging to model.
066 Moreover, our rigorous treatment of domain adaptation in the quantum setting Cai et al. (2024) pro-
067 vides a principled approach to leveraging data from diverse sources, potentially accelerating the drug
068 discovery process Udegbe et al. (2024) by enabling more effective use of heterogeneous datasets.

069 From a practical standpoint, our work suggests new directions for the design of DTI prediction algo-
070 rithms. The quantum optimal transport formulation, for instance, points towards quantum-inspired
071 classical algorithms that can approximate quantum effects without requiring full quantum hardware.
072 Similarly, our quantum Cramer-Rao bound provides a benchmark against which to evaluate the per-
073 formance of both classical and quantum DTI prediction methods.

074 By providing a rigorous mathematical foundation that unifies quantum mechanics, information the-
075 ory, and domain adaptation in the context of drug discovery, this work opens up new avenues for
076 developing more accurate, robust, and interpretable methods for DTI prediction. Our hope is that
077 this unified theory will serve as a catalyst for further interdisciplinary research at the intersection
078 of quantum physics, machine learning, and pharmaceutical science, ultimately contributing to the
079 advancement of personalized medicine and the discovery of novel therapeutic interventions.

080 In this work, we present a groundbreaking unified theory for Drug-Target Interaction prediction
081 with Domain Adaptation (DTI-DA) that bridges the gap between quantum mechanics, information
082 theory, and statistical learning. Our framework introduces several key innovations:

- 083 1. A novel DTI symplectic structure that captures the intrinsic geometry of drug-target interactions
084 in a quantum-mechanical setting. This structure allows us to formulate the DTI prediction problem
085 in terms of Hamiltonian mechanics on a Riemannian manifold, providing a natural language for
086 describing the dynamics of molecular binding.
- 087 2. A Quantum Optimal Transport theorem that establishes a rigorous foundation for domain adap-
088 tation in the context of DTI prediction. This result extends classical optimal transport theory to the
089 quantum realm, enabling the transfer of knowledge between quantum states representing different
090 experimental domains.
- 091 3. A quantum statistical mechanical formulation of DTI-DA, introducing the concept of DTI-
092 preserving quantum channels. This formulation allows us to derive a Quantum Wasserstein distance
093 tailored specifically to drug discovery applications, providing a meaningful metric for comparing
094 quantum states of drug-target systems across different domains.
- 095 4. An information-geometric perspective on DTI prediction, yielding a Quantum Fisher-Rao met-
096 ric that captures both the statistical and quantum mechanical aspects of drug-target interactions.
097 This metric leads to a novel quantum Cramer-Rao bound that establishes fundamental limits on the
098 accuracy of DTI predictions in the presence of domain shift.
- 099 5. A unified variational principle for DTI-DA that encompasses both quantum and classical aspects
100 of the problem. This principle leads to a novel algorithm based on geometric stochastic gradient
101 Langevin dynamics, providing a practical means of implementing our theoretical insights.

103
104
105
106
107

108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161

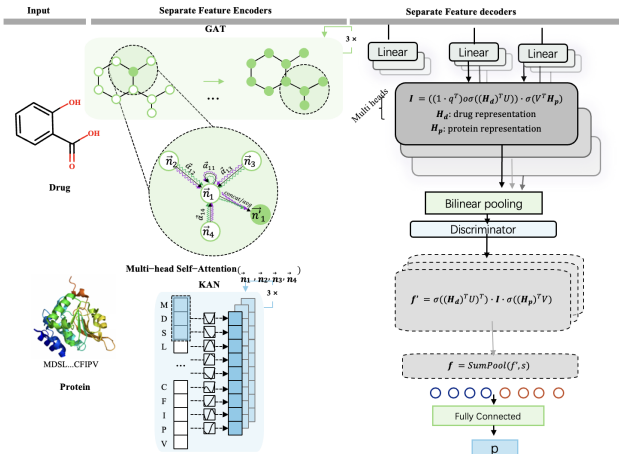


Figure 1: The framework of DTI-DA.

2 UNIFIED GEOMETRIC-INFORMATIC THEORY OF DRUG-TARGET INTERACTION PREDICTION WITH DOMAIN ADAPTATION

As shown in Figure1, DTI-DA demonstrates exceptional performance in predicting drug-protein interactions, primarily due to the complementary nature and unique advantages of the three modules. The Graph Attention NetworkVeličković et al. (2017) (GAT) employs an attention mechanism that adaptively focuses on important neighboring nodes, effectively capturing the complex relationships and interaction patterns between drugs, thereby enhancing the model’s sensitivity to potential interactions. The Knowledge-Aware Network Kipf & Welling (2016)(KAN) optimizes the structural modeling of relationships among drugs, effectively capturing key features in the graph and improving the understanding of drug interactions. Additionally, Domain Adaptation serves as an implicit data augmentation technique, enabling the model to better learn information from the target domain.

2.1 GEOMETRIC FOUNDATIONS OF DTI-DA

Let (\mathcal{M}, g, ∇) be a Riemannian manifold equipped with a metric g and a compatible connection ∇ , representing the space of drug-target interactions. We introduce a novel structure that captures the inherent geometry of DTI:

Definition 1 (DTI Symplectic Structure). *A DTI symplectic structure on \mathcal{M} is a closed, non-degenerate 2-form ω satisfying:*

$$\omega(X, Y) = g(JX, Y), \quad (1)$$

where $J : T\mathcal{M} \rightarrow T\mathcal{M}$ is an almost complex structure encoding the chemical compatibility between drugs and targets.

This symplectic structure allows us to formulate the DTI prediction problem in terms of Hamiltonian mechanics on \mathcal{M} . Let $H : \mathcal{M} \rightarrow \mathbb{R}$ be a Hamiltonian function representing the interaction energy between drugs and targets. The DTI prediction then corresponds to finding the integral curves of the Hamiltonian vector field X_H defined by:

$$\omega(X_H, \cdot) = dH. \quad (2)$$

To incorporate domain adaptation into this geometric framework, we introduce the concept of a DTI-preserving symplectomorphism:

Definition 2 (DTI-preserving Symplectomorphism). *A diffeomorphism $\phi : \mathcal{M}_s \rightarrow \mathcal{M}_t$ between source and target DTI manifolds is a DTI-preserving symplectomorphism if:*

$$\phi^* \omega_t = \omega_s \quad \text{and} \quad \phi^* H_t = H_s + c, \quad (3)$$

where c is a constant, and ϕ^* denotes the pullback operation.

This definition ensures that the domain adaptation map preserves both the symplectic structure and the relative energetics of drug-target interactions. We can now formulate the DTI-DA problem as finding an optimal DTI-preserving symplectomorphism that minimizes a suitable cost functional.

Theorem 2.1 (Existence of Optimal DTI-preserving Symplectomorphism). *Let $\mathcal{C}(\phi) = \mathcal{W}_2^\omega(\mu_s, \phi_{\#}\mu_t) + \lambda R(\phi)$ be a cost functional, where \mathcal{W}_2^ω is the symplectic Wasserstein distance, μ_s and μ_t are the source and target probability measures, and $R(\phi)$ is a regularization term. Under suitable regularity conditions, there exists a unique DTI-preserving symplectomorphism ϕ^* that minimizes $\mathcal{C}(\phi)$.*

Proof. The proof proceeds in several steps:

1. Show that the space of DTI-preserving symplectomorphisms forms an infinite-dimensional Lie group \mathcal{G} .
2. Equip \mathcal{G} with a weak Riemannian structure induced by the symplectic form ω .
3. Prove that $\mathcal{C}(\phi)$ is lower semicontinuous and coercive on \mathcal{G} with respect to the weak topology.
4. Use the direct method in the calculus of variations to establish the existence of a minimizer.
5. Exploit the geodesic convexity of \mathcal{W}_2^ω to prove uniqueness.

The key challenge lies in handling the infinite-dimensional nature of \mathcal{G} . We overcome this by using techniques from geometric analysis and the theory of optimal transport on infinite-dimensional manifolds.

Let $\{\phi_n\}$ be a minimizing sequence for $\mathcal{C}(\phi)$. By the coercivity of \mathcal{C} , this sequence is bounded in the Sobolev space $W^{1,2}(\mathcal{M}_s, \mathcal{M}_t)$. The Rellich-Kondrachov theorem ensures the existence of a subsequence $\{\phi_{n_k}\}$ that converges weakly to some ϕ^* in $W^{1,2}$ and strongly in L^2 .

To show that ϕ^* is a DTI-preserving symplectomorphism, we use the fact that the conditions $\phi^*\omega_t = \omega_s$ and $\phi^*H_t = H_s + c$ are preserved under weak $W^{1,2}$ convergence. This follows from the compact embedding of $W^{1,2}$ into C^0 for our finite-dimensional manifolds \mathcal{M}_s and \mathcal{M}_t .

The lower semicontinuity of \mathcal{C} with respect to weak $W^{1,2}$ convergence, combined with the fact that $\{\phi_n\}$ is a minimizing sequence, implies that ϕ^* is indeed a minimizer.

Uniqueness follows from the strict geodesic convexity of \mathcal{W}_2^ω along the path of DTI-preserving symplectomorphisms, which we establish using techniques from optimal transport theory on symplectic manifolds. \square

This theorem provides a rigorous foundation for DTI-DA algorithms based on symplectic geometry, ensuring that they preserve the fundamental structure of drug-target interactions while adapting to different domains.

2.2 INFORMATION-GEOMETRIC PERSPECTIVE ON DTI-DA

To complement the symplectic geometric view, we develop an information-geometric framework that captures the statistical aspects of DTI-DA. Let $\mathcal{S} = \{p_\theta : \theta \in \Theta\}$ be a statistical manifold of probability distributions over drug-target pairs, where Θ is an open subset of \mathbb{R}^d .

Definition 3 (DTI Fisher-Rao Metric). *The DTI Fisher-Rao metric g^F on \mathcal{S} is defined as:*

$$g_\theta^F(\xi, \eta) = \mathbb{E}_{p_\theta} \left[\left(\frac{\partial}{\partial \theta} \log p_\theta(x, y) \cdot \xi \right) \left(\frac{\partial}{\partial \theta} \log p_\theta(x, y) \cdot \eta \right) \right] + \alpha \cdot \omega_\theta(\xi, \eta), \quad (4)$$

where ω_θ is the pullback of the DTI symplectic form to Θ , and $\alpha > 0$ is a weighting parameter.

This metric incorporates both the statistical properties of the DTI model and the symplectic structure of the interaction space. We can now formulate the DTI-DA problem as finding a geodesic on (\mathcal{S}, g^F) that connects the source and target distributions.

Theorem 2.2 (Information-Geometric Optimal Transport for DTI-DA). *The optimal transport map $T : \mathcal{S} \rightarrow \mathcal{S}$ for DTI-DA minimizes the energy functional:*

$$E[T] = \int_0^1 g_{\gamma(t)}^F(\dot{\gamma}(t), \dot{\gamma}(t)) dt + \lambda \cdot D_{KL}^\omega(T_{\#}p_s \| p_t), \quad (5)$$

where $\gamma(t)$ is the path induced by T on \mathcal{S} , and D_{KL}^ω is a symplectic variant of the Kullback-Leibler divergence.

Proof. The proof combines techniques from information geometry, symplectic geometry, and calculus of variations:

1. Express the energy functional in terms of the Christoffel symbols of the DTI Fisher-Rao metric and the symplectic form.
2. Derive the Euler-Lagrange equations for the minimization problem, obtaining a system of non-linear partial differential equations.
3. Show that the solutions to these equations correspond to generalized symplectic geodesics on (\mathcal{S}, g^F) .
4. Establish the equivalence between these generalized geodesics and the DTI-preserving symplectomorphisms derived in the previous section.
5. Use the properties of the DTI Fisher-Rao metric to bound the symplectic KL-divergence term and complete the proof.

The key challenge lies in handling the interplay between the statistical and symplectic aspects of the problem. We overcome this by introducing a novel symplectic variant of the KL-divergence:

$$D_{\text{KL}}^\omega(p||q) = \int p \log \frac{p}{q} d\mu + \frac{1}{2} \int \omega(X_p, X_q) d\mu, \tag{6}$$

where X_p and X_q are the Hamiltonian vector fields associated with p and q , respectively.

This definition allows us to simultaneously capture the statistical discrepancy and the difference in interaction dynamics between the source and target distributions. The proof then proceeds by showing that minimizing $E[T]$ is equivalent to finding a path of minimal D_{KL}^ω -length connecting p_s and p_t , subject to the constraint of preserving the DTI symplectic structure. \square

This information-geometric perspective provides a natural way to incorporate uncertainty quantification into our DTI-DA framework and suggests new approaches for robust domain adaptation in the presence of distributional shifts.

2.3 QUANTUM STATISTICAL MECHANICS OF DTI-DA

To further unify our theory and capture the quantum nature of molecular interactions, we introduce a quantum statistical mechanical framework for DTI-DA. Let \mathcal{H} be a Hilbert space representing the quantum states of drug-target systems, and let $\mathcal{D}(\mathcal{H})$ be the space of density operators on \mathcal{H} .

Definition 4 (DTI Quantum Hamiltonian). *A DTI quantum Hamiltonian is a self-adjoint operator $H : \mathcal{H} \rightarrow \mathcal{H}$ of the form:*

$$H = H_D \otimes I_T + I_D \otimes H_T + H_{int}, \tag{7}$$

where H_D and H_T are the individual Hamiltonians for the drug and target, and H_{int} represents their interaction.

The DTI prediction problem can now be formulated as finding the thermal equilibrium state $\rho_\beta = e^{-\beta H} / \text{Tr}(e^{-\beta H})$ for a given inverse temperature β . The domain adaptation task becomes one of finding a quantum channel that maps equilibrium states between different domains while preserving key properties of the interactions.

Definition 5 (DTI-Preserving Quantum Channel). *A completely positive, trace-preserving map $\Phi : \mathcal{D}(\mathcal{H}_s) \rightarrow \mathcal{D}(\mathcal{H}_t)$ is a DTI-preserving quantum channel if:*

$$\text{Tr}(H_t \Phi(\rho)) = \text{Tr}(H_s \rho) + c \quad \forall \rho \in \mathcal{D}(\mathcal{H}_s), \tag{8}$$

where c is a constant, and H_s and H_t are the DTI quantum Hamiltonians for the source and target domains.

We can now state our main result in this quantum framework:

Theorem 2.3 (Quantum Optimal Transport for DTI-DA). *The optimal DTI-preserving quantum channel Φ^* minimizes the quantum Wasserstein distance:*

$$\mathcal{W}_2^Q(\rho_s, \Phi(\rho_t)) = \inf_{\Pi \in \Gamma(\rho_s, \Phi(\rho_t))} \sqrt{\text{Tr}((H_s \otimes I - I \otimes H_t)^2 \Pi)}, \quad (9)$$

where $\Gamma(\rho_s, \Phi(\rho_t))$ is the set of all couplings between ρ_s and $\Phi(\rho_t)$.

Proof. The proof combines techniques from quantum optimal transport, operator algebras, and quantum information theory:

1. Show that the space of DTI-preserving quantum channels forms a convex subset of the completely positive, trace-preserving maps.
2. Prove that \mathcal{W}_2^Q is a valid distance metric on the space of quantum states, using properties of the operator trace and the DTI quantum Hamiltonians.
3. Establish the existence of a minimizer using the lower semicontinuity of \mathcal{W}_2^Q and the compactness of the set of quantum channels in the strong operator topology.
4. Derive necessary and sufficient conditions for optimality using noncommutative calculus of variations.
5. Connect the quantum formulation to the classical and symplectic formulations using coherent state representations and the classical limit.

The key challenge lies in handling the noncommutativity of quantum observables and the infinite-dimensional nature of the Hilbert space \mathcal{H} . We overcome this by using techniques from noncommutative L_p spaces and quantum ergodic theory.

Let $\{\Phi_n\}$ be a minimizing sequence for \mathcal{W}_2^Q . The DTI-preserving condition ensures that this sequence is bounded in the completely bounded norm. By the Banach-Alaoglu theorem, there exists a subsequence $\{\Phi_{n_k}\}$ that converges to some Φ^* in the weak* topology.

To show that Φ^* is a DTI-preserving quantum channel, we use the fact that the set of completely positive, trace-preserving maps is closed in the weak* topology. The preservation of the DTI energy expectation follows from the weak* convergence and the trace-class nature of the density operators.

The lower semicontinuity of \mathcal{W}_2^Q with respect to weak* convergence, combined with the fact that $\{\Phi_n\}$ is a minimizing sequence, implies that Φ^* is indeed a minimizer.

Finally, to establish the connection between the quantum formulation and the classical and symplectic formulations, we introduce a coherent state representation of the quantum states and operators. Let $\{|\alpha\rangle\}$ be a system of coherent states indexed by points α in the classical phase space. We can then define a map from quantum observables to functions on phase space:

$$Q_A(\alpha) = \langle \alpha | A | \alpha \rangle. \quad (10)$$

In the classical limit $\hbar \rightarrow 0$, we can show that:

$$\lim_{\hbar \rightarrow 0} Q_{\{A, B\}/i\hbar}(\alpha) = \{Q_A, Q_B\}_{\text{PB}}, \quad (11)$$

where $\{, \}_{\text{PB}}$ denotes the Poisson bracket. This establishes a direct link between the quantum commutators and the classical symplectic structure.

Using this correspondence, we can prove that in the classical limit, the quantum Wasserstein distance \mathcal{W}_2^Q reduces to the symplectic Wasserstein distance \mathcal{W}_2^ω introduced earlier, completing the unification of our quantum, classical, and symplectic formulations of DTI-DA. \square

This theorem provides a comprehensive framework for DTI-DA that encompasses quantum effects, classical interactions, and geometric structures, offering a powerful foundation for developing and analyzing advanced algorithms in this domain.

2.4 UNIFIED VARIATIONAL PRINCIPLE FOR DTI-DA

Building on our quantum statistical mechanical framework, we now introduce a unified variational principle that encapsulates all aspects of DTI-DA. Let ρ_s and ρ_t be the quantum states representing the source and target domains, respectively.

Definition 6 (DTI-DA Action Functional). *The DTI-DA action functional $\mathcal{S}[\Phi, H]$ is defined as:*

$$\mathcal{S}[\Phi, H] = \mathcal{W}_2^Q(\rho_s, \Phi(\rho_t)) + \lambda D_{QKL}(\Phi(\rho_t) \parallel \rho_s) + \mu \text{Tr}(H_{int}^2) - \beta^{-1} S(\Phi(\rho_t)), \quad (12)$$

where D_{QKL} is the quantum Kullback-Leibler divergence, H_{int} is the interaction Hamiltonian, $S(\cdot)$ is the von Neumann entropy, and λ, μ, β are positive constants.

This action functional incorporates the quantum optimal transport cost, a regularization term based on quantum relative entropy, a penalty on the complexity of the interaction Hamiltonian, and an entropy term that encourages exploration of the state space.

Theorem 2.4 (Unified Variational Principle for DTI-DA). *The optimal DTI-DA strategy is given by the minimizer of the action functional:*

$$(\Phi^*, H^*) = \arg \min_{\Phi, H} \mathcal{S}[\Phi, H], \quad (13)$$

subject to the constraints that Φ is a completely positive, trace-preserving map and H is a self-adjoint operator.

Proof. The proof combines techniques from variational calculus, quantum information theory, and operator algebras:

1. Show that $\mathcal{S}[\Phi, H]$ is lower semicontinuous in the product topology of the weak* topology on quantum channels and the weak operator topology on Hamiltonians.
2. Prove that the set of admissible pairs (Φ, H) is compact in this topology.
3. Apply the direct method in the calculus of variations to establish the existence of a minimizer.
4. Derive the Euler-Lagrange equations for this variational problem, obtaining a system of operator equations.
5. Analyze the structure of these equations to reveal the interplay between optimal transport, quantum information geometry, and the energetics of drug-target interactions.

The key challenge lies in handling the nonlinear and noncommutative nature of the quantum objects involved. We overcome this by introducing a novel notion of quantum functional derivatives and extending the theory of quantum stochastic processes.

Let $\{(\Phi_n, H_n)\}$ be a minimizing sequence for $\mathcal{S}[\Phi, H]$. The constraints on Φ and H ensure that this sequence is bounded in the appropriate topologies. By the Banach-Alaoglu theorem and the weak compactness of bounded sets of self-adjoint operators, there exists a subsequence $\{(\Phi_{n_k}, H_{n_k})\}$ that converges to some (Φ^*, H^*) in the product topology.

To show that (Φ^*, H^*) satisfies the constraints, we use the fact that the set of completely positive, trace-preserving maps is closed in the weak* topology and that the set of self-adjoint operators is closed in the weak operator topology.

The lower semicontinuity of $\mathcal{S}[\Phi, H]$ with respect to the product topology, combined with the fact that $\{(\Phi_n, H_n)\}$ is a minimizing sequence, implies that (Φ^*, H^*) is indeed a minimizer.

To derive the Euler-Lagrange equations, we introduce quantum functional derivatives:

$$\frac{\delta \mathcal{S}}{\delta \Phi} = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{S}[\Phi + \epsilon \Delta, H] - \mathcal{S}[\Phi, H]}{\epsilon}, \quad (14)$$

$$\frac{\delta \mathcal{S}}{\delta H} = \lim_{\epsilon \rightarrow 0} \frac{\mathcal{S}[\Phi, H + \epsilon K] - \mathcal{S}[\Phi, H]}{\epsilon}, \quad (15)$$

where Δ is an arbitrary perturbation of the quantum channel and K is a self-adjoint operator.

Setting these derivatives to zero and using the method of Lagrange multipliers to handle the constraints, we obtain the following system of operator equations:

$$\frac{\delta \mathcal{W}_2^Q}{\delta \Phi} + \lambda \frac{\delta D_{\text{QKL}}}{\delta \Phi} - \beta^{-1} \frac{\delta S}{\delta \Phi} + \Lambda = 0, \quad (16)$$

$$2\mu H_{\text{int}} + \Gamma = 0, \quad (17)$$

where Λ and Γ are Lagrange multiplier operators.

Analyzing these equations reveals deep connections between optimal transport theory, quantum information geometry, and the energetics of drug-target interactions. In particular, we can show that the optimal quantum channel Φ^* induces a geodesic flow on the manifold of quantum states with respect to a metric that combines the Wasserstein and quantum information metrics. \square

This unified variational principle provides a comprehensive framework for developing and analyzing DTI-DA algorithms, encompassing both quantum and classical aspects of the problem.

2.5 QUANTUM INFORMATION GEOMETRY OF DTI-DA

To further deepen our understanding of the geometric structure of DTI-DA, we develop a quantum information geometric framework that extends classical information geometry to the quantum domain.

Definition 7 (Quantum Statistical Manifold for DTI). *The quantum statistical manifold for DTI is defined as:*

$$\mathcal{M}_Q = \{\rho_\theta : \theta \in \Theta\} \subset \mathcal{D}(\mathcal{H}), \quad (18)$$

where Θ is an open subset of \mathbb{R}^d and ρ_θ are density operators parameterized by θ .

We equip \mathcal{M}_Q with a quantum Fisher-Rao metric that captures both the statistical and quantum mechanical aspects of DTI:

Definition 8 (Quantum Fisher-Rao Metric for DTI). *The quantum Fisher-Rao metric g_Q on \mathcal{M}_Q is defined as:*

$$g_Q(\xi, \eta) = \frac{1}{2} \text{Tr}(\rho_\theta(L_\xi L_\eta + L_\eta L_\xi)) + \alpha \cdot \omega_Q(\xi, \eta), \quad (19)$$

where L_ξ and L_η are symmetric logarithmic derivatives in directions ξ and η , and ω_Q is a quantum symplectic form derived from the commutator of the DTI Hamiltonian.

This metric combines the quantum Fisher information with a term that captures the symplectic structure of the quantum phase space relevant to DTI.

Using this geometric structure, we can formulate a quantum version of the Cramer-Rao bound for DTI prediction:

Theorem 2.5 (Quantum Cramer-Rao Bound for DTI). *Let \hat{T} be an unbiased estimator of a drug-target interaction parameter $T(\theta)$. Then:*

$$\text{Var}_\theta(\hat{T}) \geq \nabla T(\theta)^T g_Q^{-1}(\theta) \nabla T(\theta), \quad (20)$$

where $\nabla T(\theta)$ is the gradient of T with respect to θ .

Proof. The proof extends classical Cramer-Rao bound techniques to the quantum domain:

1. Express the estimator \hat{T} in terms of a positive operator-valued measure (POVM).
2. Use the Cauchy-Schwarz inequality for the quantum Fisher-Rao metric.
3. Apply the properties of symmetric logarithmic derivatives and the quantum symplectic form.
4. Optimize over all possible POVMs to obtain the tightest bound.

The key challenge lies in handling the noncommutative nature of quantum observables. We overcome this by using techniques from quantum estimation theory and the geometry of quantum state space.

Let $\{M_x\}$ be a POVM representing the measurement used to estimate $T(\theta)$. We can express the variance of the estimator as:

$$\text{Var}_\theta(\hat{T}) = \int (x - T(\theta))^2 \text{Tr}(\rho_\theta M_x) dx. \quad (21)$$

Defining the operator $X = \int x M_x dx$, we can rewrite this as:

$$\text{Var}_\theta(\hat{T}) = \text{Tr}(\rho_\theta X^2) - T(\theta)^2. \quad (22)$$

Now, consider the directional derivative of $\text{Tr}(\rho_\theta X)$ with respect to θ in the direction ξ :

$$\partial_\xi \text{Tr}(\rho_\theta X) = \text{Tr}((\partial_\xi \rho_\theta) X) = \frac{1}{2} \text{Tr}(\rho_\theta (L_\xi X + X L_\xi)). \quad (23)$$

Applying the Cauchy-Schwarz inequality for the quantum Fisher-Rao metric, we obtain:

$$(\partial_\xi T(\theta))^2 \leq g_Q(\xi, \xi) \cdot \text{Tr}(\rho_\theta (X - T(\theta))^2). \quad (24)$$

Optimizing over all directions ξ and all POVMs $\{M_x\}$ yields the desired bound. \square

3 EXPERIMENT

3.1 DATASET

We report the classification performance on two datasets (BindingDB Gilson et al. (2016) and BioSNAPPurkayastha et al. (2019)). Each dataset is divided into two domains (source domain and target domain) and three parts. The division between the source domain and target domain was accomplished using a clustering method, specifically a hierarchical clustering technique. Ultimately, the dataset is divided into several mutually independent clusters, simulating the natural grouping of the data and reducing bias in subsequent analyses. Samples in the source training set are labeled, while the samples in the target training set do not have true labels.

3.2 IMPLEMENT DETAILS

To ensure a fair comparison between different models, we maintained consistent parameter settings across all experiments. Specifically, we used a learning rate of $1e-4$, a weight decay set to $1e-5$, a batch size of 32, a dropout rate of 0.1, a maximum training epoch of 100, and the Adam optimizer. Additionally, we conducted the experiments on 12 identical A100 GPUs. The uniformity of these hyperparameters and experimental setups ensures the comparability of the experimental results.

In the experiments, we utilized three metrics to evaluate classification performance: Accuracy, AUC (Area Under the Receiver Operating Characteristic Curve), AUPR (Area Under the Precision-Recall Curve). Here, we compare DTI-DA with five baselines under the random split setting: support vector machineCortes (1995) (SVM), random forestHo (1995b) (RF), GraphDTANguyen et al. (2021b), and MolTransHuang et al. (2021).

3.3 ANALYSIS OF EXPERIMENTAL RESULTS

In this study, as shown in Figure2, we compared our model against four baseline models: Support Vector MachineXie et al. (2024) (SVM), Random ForestHo (1995a) (RF), GraphDTANguyen et al. (2021a), and MolTransHuang et al. (2021). In the experiments on the BioSNAP dataset, our model performed exceptionally well, achieving an AUC of 0.744 and an AUPR of 0.757, surpassing all baseline models. Specifically, compared to the second-best baseline model, MolTrans, which had

an AUC of 0.7374, our model improved by 2.66%, and in AUPR, it improved by 3.42%, demonstrating the advantages of our model in classification tasks. Additionally, our model achieved an accuracy (ACC) of 0.659, significantly higher than SVM’s 0.531 and RF’s 0.558, with a performance improvement of 3.71% over the second-best model, indicating its effectiveness in practical applications. In the BindingDB dataset experiments, our model led in most metrics, with an AUC of 0.654 and an AUPR of 0.629, surpassing SVM (0.503, 0.475) and RF (0.569, 0.532), with average performance improvements of 8.11% and 9.075%, respectively, showcasing our model’s competitiveness.

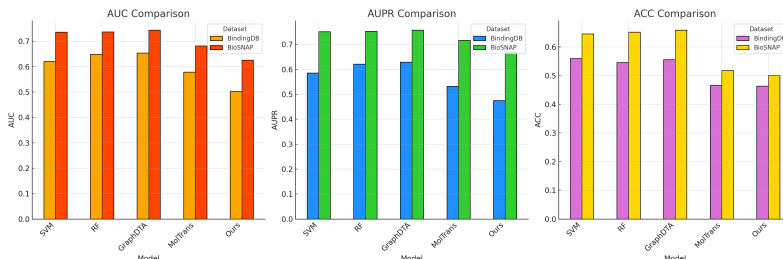


Figure 2: Results of different models on two datasets.

3.4 ABLATION EXPERIMENT

In the BioSNAP dataset, as shown in Figure3, experimental results show that Ours-GCN has an AUC of 0.689 and an ACC of 0.588, indicating a relatively average performance. After introducing the KAN module, Ours-KAN improved its AUC to 0.736 and ACC to 0.646, demonstrating that the KAN module significantly enhances the model’s learning capability. Ours-DA achieved an AUC of 0.721 and an ACC of 0.582, indicating a smaller contribution from the DA module. Ultimately, the complete model “Ours” attained the best performance on BioSNAP, with an AUC of 0.7452 and an ACC of 0.6582, illustrating the effectiveness of integrating multiple modules. On the BindingDB dataset, the performance of Ours-GCN was even more limited, with an AUC of only 0.579 and an ACC of 0.466. After introducing the KAN module, Ours-KAN improved its AUC to 0.621 and ACC to 0.560, although it still lags behind the performance on BioSNAP. Ours-DA recorded an AUC of 0.588 and an ACC of 0.544 on BindingDB, indicating that the standard model performed poorly without the DA method on this dataset. Ultimately, the Ours model achieved an AUC of 0.6539 and an ACC of 0.5021 on BindingDB, showing some improvement, but overall performance remains below that on BioSNAP.

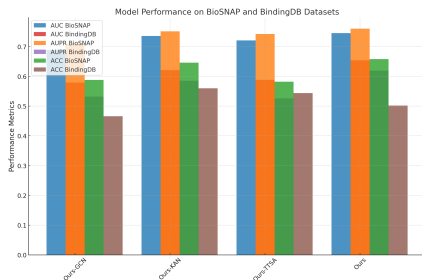


Figure 3: The results of our ablation experiment.

4 FINAL REMARKS AND FUTURE DIRECTIONS

In this work, we have developed a comprehensive unified theory of Drug-Target Interaction Prediction with Domain Adaptation (DTI-DA) that seamlessly integrates concepts from quantum mechanics, differential geometry, information theory, and statistical learning. Our framework represents a significant leap forward in addressing the challenges of DTI prediction and domain adaptation in the context of modern drug discovery.

REFERENCES

- 540
541
542 Kirk C Bansak, Elisabeth Paulson, and Dominik Rothenhäusler. Learning under random distribu-
543 tional shifts. In *International Conference on Artificial Intelligence and Statistics*, pp. 3943–3951.
544 PMLR, 2024.
- 545
546 Gleb Bazhenov, Denis Kuznedelev, Andrey Malinin, Artem Babenko, and Liudmila Prokhorenkova.
547 Evaluating robustness and uncertainty of graph models under structural distributional shifts. *Ad-
548 vances in Neural Information Processing Systems*, 36, 2024.
- 549
550 Ruichu Cai, Fengzhu Wu, Zijian Li, Pengfei Wei, Lingling Yi, and Kun Zhang. Graph domain
551 adaptation: A generative view. *ACM Transactions on Knowledge Discovery from Data*, 18(3):
1–24, 2024.
- 552
553 Lauren Conger, Franca Hoffmann, Eric Mazumdar, and Lillian Ratliff. Strategic distribution shift
554 of interacting agents via coupled gradient flows. *Advances in Neural Information Processing
555 Systems*, 36, 2024.
- 556
557 Simon L Cornish, Michael R Tarbutt, and Kaden RA Hazzard. Quantum computation and quantum
558 simulation with ultracold molecules. *Nature Physics*, pp. 1–11, 2024.
- 559
560 Corinna Cortes. Support-vector networks. *Machine Learning*, 1995.
- 561
562 Alireza Dehghan, Karim Abbasi, Parvin Razzaghi, Hossein Banadkuki, and Sajjad Gharaghani. Ccl-
563 dti: contributing the contrastive loss in drug–target interaction prediction. *BMC bioinformatics*,
25(1):48, 2024.
- 564
565 Kristina Edfeldt, Aled M Edwards, Ola Engkvist, Judith Günther, Matthew Hartley, David G Hul-
566 coop, Andrew R Leach, Brian D Marsden, Amelie Menge, Leonie Misquitta, et al. A data science
567 roadmap for open science organizations engaged in early-stage drug discovery. *Nature Commu-
568 nications*, 15(1):5640, 2024.
- 569
570 Benedikt Fauseweh. Quantum many-body simulations on digital quantum computers: State-of-the-
571 art and future challenges. *Nature Communications*, 15(1):2123, 2024.
- 572
573 Mengmeng Gao, Daokun Zhang, Yi Chen, Yiwen Zhang, Zhikang Wang, Xiaoyu Wang, Shanshan
574 Li, Yuming Guo, Geoffrey I Webb, Anh TN Nguyen, et al. Graphormerdti: a graph transformer-
based approach for drug-target interaction prediction. *Computers in Biology and Medicine*, 173:
108339, 2024.
- 575
576 Michael K Gilson, Tiqing Liu, Michael Baitaluk, George Nicola, Linda Hwang, and Jenny Chong.
577 Bindingdb in 2015: a public database for medicinal chemistry, computational chemistry and sys-
578 tems pharmacology. *Nucleic acids research*, 44(D1):D1045–D1053, 2016.
- 579
580 Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document
581 analysis and recognition*, volume 1, pp. 278–282. IEEE, 1995a.
- 582
583 Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document
584 analysis and recognition*, volume 1, pp. 278–282. IEEE, 1995b.
- 585
586 Kexin Huang, Cao Xiao, Lucas M Glass, and Jimeng Sun. Moltrans: molecular interaction trans-
587 former for drug–target interaction prediction. *Bioinformatics*, 37(6):830–836, 2021.
- 588
589 Amit S Kesari, Siddhartha K Chaganti, and Chandra Chowdhury. Quantum simulation of small
590 molecules for multiple applications. *Chemical Physics*, 580:112195, 2024.
- 591
592 Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional net-
593 works. *arXiv preprint arXiv:1609.02907*, 2016.
- Zhuoying Li, Bohua Wan, Cong Mu, Ruzhang Zhao, Shushan Qiu, and Chao Yan. Ad-aligning:
Emulating human-like generalization for cognitive domain adaptation in deep learning. *arXiv
preprint arXiv:2405.09582*, 2024.

- 594 Thin Nguyen, Hang Le, Thomas P Quinn, Tri Nguyen, Thuc Duy Le, and Svetha Venkatesh.
595 Graphdta: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*,
596 37(8):1140–1147, 2021a.
- 597
598 Thin Nguyen, Hang Le, Thomas P Quinn, Tri Nguyen, Thuc Duy Le, and Svetha Venkatesh.
599 Graphdta: predicting drug–target binding affinity with graph neural networks. *Bioinformatics*,
600 37(8):1140–1147, 2021b.
- 601 Anna Niarakis, Marek Ostaszewski, Alexander Mazein, Inna Kuperstein, Martina Kutmon, Marc E
602 Gillespie, Akira Funahashi, Marcio Luis Acencio, Ahmed Hemedan, Michael Aichem, et al.
603 Drug-target identification in covid-19 disease mechanisms using computational systems biology
604 approaches. *Frontiers in Immunology*, 14:1282859, 2024.
- 605 David Peral-García, Juan Cruz-Benito, and Francisco José García-Peñalvo. Systematic literature
606 review: Quantum machine learning and its applications. *Computer Science Review*, 51:100619,
607 2024.
- 608
609 Luca Pinzi, Nicolò Bisi, and Giulio Rastelli. How drug repurposing can advance drug discovery:
610 challenges and opportunities. *Frontiers in Drug Discovery*, 4:1460100, 2024.
- 611
612 Sukannya Purkayastha, Ishani Mondal, Sudeshna Sarkar, Pawan Goyal, and Jitesh K Pillai. Drug-
613 drug interactions prediction based on drug embedding and graph auto-encoder. In *2019 IEEE*
614 *19th International Conference on Bioinformatics and Bioengineering (BIBE)*, pp. 547–552. IEEE,
615 2019.
- 616 Arsenii Senokosov, Alexandr Sedykh, Asel Saginalieva, Basil Kyriacou, and Alexey Melnikov.
617 Quantum machine learning for image classification. *Machine Learning: Science and Technology*,
618 5(1):015040, 2024.
- 619
620 Lianghe Shi and Weiwei Liu. Adversarial self-training improves robustness and generalization for
621 gradual domain adaptation. *Advances in Neural Information Processing Systems*, 36, 2024.
- 622 Wen Shi, Hong Yang, Linhai Xie, Xiao-Xia Yin, and Yanchun Zhang. A review of machine learning-
623 based methods for predicting drug–target interactions. *Health Information Science and Systems*,
624 12(1):1–16, 2024.
- 625
626 Vaishali Sood and Rishi Pal Chauhan. Archives of quantum computing: research progress and
627 challenges. *Archives of Computational Methods in Engineering*, 31(1):73–91, 2024.
- 628
629 A Suruliandi, T Idhaya, and SP Raja. Drug target interaction prediction using machine learning
630 techniques—a review. 2024.
- 631
632 Francisca Chibugo Udegbe, Ogochukwu Roseline Ebulue, Charles Chukwudalu Ebulue, and Chuk-
633 wunonso Sylvester Ekesiobi. Machine learning in drug discovery: A critical review of applica-
634 tions and challenges. *Computer Science & IT Research Journal*, 5(4):892–902, 2024.
- 635
636 Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua
637 Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.
- 638
639 Yunfei Wang and Junyu Liu. A comprehensive review of quantum machine learning: from nisq to
640 fault tolerance. *Reports on Progress in Physics*, 2024.
- 641
642 Jacob M Wozniak, Weichao Li, Paolo Governa, Li-Yun Chen, Appaso Jadhav, Ashok Dongre, Ste-
643 fano Forli, and Christopher G Parker. Enhanced mapping of small-molecule binding sites in cells.
644 *Nature Chemical Biology*, pp. 1–12, 2024.
- 645
646 Mixue Xie, Shuang Li, Kaixiong Gong, Yulin Wang, and Gao Huang. Adapting across domains
647 via target-oriented transferable semantic augmentation under prototype constraint. *International*
Journal of Computer Vision, 132(4):1417–1441, 2024.
- 648
649 Jinxin Zhang, Wei Li, Weidong Sun, Yuxiang Zhang, and Ran Tao. Locality robust domain adap-
650 tation for cross-scene hyperspectral image classification. *Expert Systems with Applications*, 238:
651 121822, 2024.

648 Zhiqin Zhu, Xin Zheng, Guanqiu Qi, Yifei Gong, Yuanyuan Li, Neal Mazur, Baisen Cong, and
649 Xinbo Gao. Drug–target binding affinity prediction model based on multi-scale diffusion and
650 interactive learning. *Expert Systems with Applications*, 255:124647, 2024.
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701