AI and Data Science Journal
ISSN: 3078-7394
DOI: https://doi.org/10.61784/adsj3026

PHYSICS-INFORMED NEURAL NETWORKS FOR ARBITRAGE-FREE VOLATILITY SURFACE CONSTRUCTION IN INCOMPLETE MARKETS

Hao Li*, Martin Keller
School of Finance, University of St. Gallen, German-Speaking, Switzerland.
Corresponding Author: Hao Li, Email: hao.li99@gmail.com

Abstract: The construction of arbitrage-free volatility surfaces represents a fundamental challenge in quantitative finance, particularly in incomplete markets where hedging portfolios cannot perfectly replicate option payoffs. Traditional parametric models often fail to capture the complex dynamics of market-observed implied volatility structures, including the characteristic smile and skew patterns, while simultaneously satisfying no-arbitrage constraints. This paper introduces a novel framework that leverages Physics-Informed Neural Networks (PINNs) to construct arbitrage-free volatility surfaces in incomplete market settings. The proposed methodology integrates partial differential equation constraints derived from arbitrage-free conditions directly into the neural network training process through automatic differentiation and soft constraint penalties. By incorporating Dupire's local volatility equation and calendar-butterfly arbitrage constraints into a multi-objective loss function, our approach generates smooth, arbitrage-free implied volatility surfaces that accurately fit market data across different strikes and maturities. Numerical experiments using both synthetic data and real market observations from S&P 500 and VIX options demonstrate that the PINN-based framework substantially reduces calibration errors while maintaining theoretical consistency. The method exhibits particular strength in handling incomplete market scenarios where traditional parametric approaches produce inconsistent surfaces or violate no-arbitrage conditions.
Keywords: Physics-informed neural networks; Arbitrage-free constraints; Volatility surface; Incomplete markets; Volatility smile; Heston model

1 INTRODUCTION

The accurate modeling of implied volatility surfaces constitutes a cornerstone of modern derivative pricing and risk management practices. Since the development of the Black-Scholes framework, market participants have observed that option prices exhibit systematic deviations from constant volatility assumptions, manifesting as the volatility smile and skew across different strikes and term structure effects across maturities [1]. These patterns reflect the market's assessment of the underlying asset's return distribution, capturing features such as fat tails, asymmetry, and time-varying uncertainty that cannot be accommodated within simple log-normal models. The challenge of constructing volatility surfaces that simultaneously fit market observations, exhibit economically sensible interpolation and extrapolation behavior, and respect fundamental no-arbitrage principles has motivated extensive research spanning parametric modeling, nonparametric estimation, and more recently, machine learning approaches.

In incomplete markets, where the number of tradable assets is insufficient to construct perfect hedging strategies for all contingent claims, the volatility surface construction problem becomes particularly intricate. Unlike complete market settings where unique arbitrage-free prices can be determined through replication arguments, incomplete markets admit a continuum of possible equivalent martingale measures [2]. The pricing of index options in incomplete markets requires careful consideration of risk preferences and market constraints to identify appropriate pricing measures from the set of admissible candidates [3]. This multiplicity introduces additional complexity when attempting to infer a consistent volatility structure from observed option prices, as different pricing measures may imply different volatility surfaces even when consistent with the same set of traded option prices. The practical consequence is that calibration procedures must incorporate additional economic principles or regularization mechanisms to select among the infinite set of theoretically valid surfaces.

Traditional parametric approaches to volatility surface modeling impose specific functional forms that aim to capture stylized facts observed in option markets. The Stochastic Volatility Inspired (SVI) parameterization represents one prominent example, expressing the implied total variance as a function of log-moneyness through a small number of parameters that control the at-the-money level, skew, and wing behavior [4]. While SVI and related models offer computational tractability and can be calibrated efficiently to liquid option quotes, they inherently limit flexibility through their parametric structure. During periods of market stress or for assets with unusual smile characteristics, these rigid functional forms may prove inadequate, leading to systematic fitting errors that propagate into prices of exotic derivatives and hedging strategies. Moreover, ensuring that parametric surfaces remain arbitrage-free across all strikes and maturities requires careful constraint handling that can further restrict the model's adaptability.

Recent advances in machine learning have opened new avenues for tackling complex financial modeling problems, with neural networks offering universal approximation capabilities that enable learning of highly nonlinear relationships from data [5]. However, applying standard data-driven neural network architectures to financial problems presents significant challenges related to constraint satisfaction and economic interpretability. Pure black-box approaches, which optimize purely on data-fitting objectives without incorporating domain knowledge, often generate solutions that violate fundamental economic principles [6]. In the context of volatility surface construction, unconstrained neural networks may produce surfaces that admit arbitrage opportunities through violations of convexity, monotonicity, or smoothness requirements. The development of methods to detect model-free static arbitrage strategies using neural networks has highlighted both the potential and limitations of purely data-driven approaches, demonstrating that arbitrage detection itself can be formulated as a neural network optimization problem [7].

The emergence of Physics-Informed Neural Networks offers a promising paradigm to bridge the gap between data-driven flexibility and theory-guided constraints [8]. Originally developed for solving forward and inverse problems in physical systems governed by partial differential equations, PINNs embed governing equations directly into the neural network training objective through automatic differentiation of the network outputs with respect to inputs. This enables the network to learn solutions that simultaneously fit observed data and satisfy underlying physical laws encoded as PDEs. The application of PINNs to financial problems represents a natural extension, as option pricing and volatility modeling are fundamentally governed by PDEs such as the Black-Scholes equation and Dupire's equation relating local volatility to call option prices [9]. Recent work has demonstrated the efficacy of PINNs in option pricing tasks, showing that physics-informed approaches can achieve superior accuracy compared to purely data-driven methods while requiring significantly less training data and exhibiting better generalization properties [10].

The construction of volatility surfaces using neural network techniques has gained considerable attention, with various approaches exploring different mechanisms for incorporating financial constraints. Early applications focused on using feedforward networks for implied volatility prediction without explicit enforcement of no-arbitrage conditions, treating the problem as standard nonlinear regression [11]. More sophisticated approaches have incorporated domain knowledge through architectural design choices, such as specialized activation functions that encode properties like smile asymmetry, or through penalty terms in the loss function that discourage arbitrage violations [12]. The challenge of ensuring arbitrage-free conditions in neural network-generated volatility surfaces has motivated the development of hybrid approaches that combine parametric foundations with neural network flexibility [13]. The hybrid gated neural network architecture represents one such advancement, using multiplicative structures and carefully selected input transformations to satisfy no-arbitrage constraints while maintaining sufficient expressiveness for accurate market fitting [14].

Building upon these advances, this paper proposes a comprehensive framework for arbitrage-free volatility surface construction in incomplete markets using Physics-Informed Neural Networks. Our methodology distinguishes itself through several key innovations that address practical challenges encountered in real-world applications. First, we formulate the volatility surface construction as a constrained optimization problem where the neural network must simultaneously minimize pricing errors on observed option quotes and satisfy multiple PDE constraints derived from no-arbitrage theory. The network learns to represent the implied volatility surface as a smooth function of log-moneyness and time-to-maturity, capturing both the smile shape within each maturity slice and the term structure evolution across maturities. Second, we develop a composite loss function that incorporates Dupire's local volatility equation, calendar spread constraints preventing time-value violations, and butterfly spread constraints ensuring density positivity. These constraints are implemented as differentiable penalty terms that can be efficiently evaluated through automatic differentiation. Third, we introduce an adaptive weighting scheme that dynamically balances data-fitting objectives against constraint satisfaction throughout the training process, starting with emphasis on data fitting to learn the approximate surface shape and gradually increasing constraint weights to enforce theoretical consistency.

The practical motivation for this research stems from challenges faced by derivatives traders and risk managers who require robust volatility surface models for pricing exotic options, computing hedging ratios, and assessing portfolio risks [15]. In incomplete markets such as emerging market equities or thinly traded indices, the absence of perfect hedging strategies introduces model risk that must be carefully managed. Existing calibration methods often struggle to produce consistent surfaces when market data is sparse, exhibits wide bid-ask spreads, or displays unusual patterns during volatile market conditions [16]. The ability to construct surfaces that fit available market observations while maintaining smooth interpolation in data-sparse regions and reasonable extrapolation beyond observed strikes and maturities is crucial for practical applications. Our PINN-based approach addresses these challenges by leveraging both empirical market data and theoretical constraints, producing surfaces that remain stable and economically sensible across varying market conditions and data quality scenarios.

2 LITERATURE REVIEW

The literature on volatility surface modeling encompasses theoretical foundations establishing arbitrage-free conditions, numerical methods for surface construction and calibration, and empirical investigations of market volatility dynamics. This section reviews key developments across these areas, with particular emphasis on recent advances in machine learning approaches and their application to incomplete market settings, providing context for the methodological contributions of the present work.

The theoretical characterization of arbitrage-free volatility surfaces traces back to the development of local volatility models and their relationship to market-observed implied volatilities. The Dupire equation established that given a continuum of European call option prices across strikes and maturities satisfying certain regularity conditions, one can uniquely determine a local volatility function that reproduces these prices under a diffusion model [17]. This seminal result provides the foundation for understanding the constraints that implied volatility surfaces must satisfy to preclude static arbitrage opportunities. The local volatility at any point in the strike-maturity space can be expressed in terms of partial derivatives of call prices with respect to strike and maturity, establishing a forward Kolmogorov equation that governs the evolution of the option price surface. Subsequent research has extended these results to incorporate discrete dividends, stochastic interest rates, and jump processes, demonstrating the robustness of the fundamental relationship between option prices and the underlying volatility structure [18].

Beyond the Dupire framework, the characterization of arbitrage-free surfaces extends to explicit constraints on the shape and dynamics of implied volatility. Calendar spread arbitrage occurs when options with longer maturities trade at lower implied volatilities than shorter-dated options with the same strike, violating basic time-value principles that longer optionality should command higher premiums [19]. Butterfly spread arbitrage arises when the second derivative of call prices with respect to strike becomes negative, implying negative probabilities in the risk-neutral density and creating profit opportunities through appropriately structured option portfolios [20]. The Surface SVI (SSVI) parameterization represents a significant advance in providing global volatility surface specifications that guarantee absence of these arbitrage types through explicit parameter restrictions [21]. The SSVI model extends the original SVI smile parameterization to a full surface representation where individual maturity slices belong to a restricted family of SVI functions, with conditions on the ATM variance term structure and the volatility-of-volatility parameter ensuring calendar spread freedom and butterfly spread conditions ensuring density positivity across all strikes and maturities [22].

The application of neural networks to option pricing and implied volatility modeling has evolved from simple feedforward architectures trained purely on market data to sophisticated physics-informed approaches that embed financial constraints. Early studies explored multilayer perceptrons as universal approximators for option pricing functions, demonstrating that networks with sufficient capacity could learn complex mappings from input features to option values with high accuracy on in-sample data [23]. However, these initial approaches treated option pricing as standard nonlinear regression without mechanisms to enforce theoretical consistency, often producing models that exhibited poor out-of-sample generalization or violated no-arbitrage conditions in regions with sparse training data. Recognition of these limitations motivated the development of hybrid approaches that incorporate financial domain knowledge through various mechanisms including constrained network architectures, specialized activation functions encoding known properties like monotonicity or convexity, and augmented loss functions penalizing violations of theoretical requirements [24].

The challenge of constructing arbitrage-free volatility surfaces using neural networks has been addressed through both hard and soft constraint enforcement strategies. Hard constraint approaches modify the network architecture itself to guarantee that outputs satisfy required conditions regardless of the learned parameters, implementing constraints through careful design of layer operations and activation functions that preserve desired properties [25]. While theoretically appealing for their guarantee of constraint satisfaction, hard constraint methods often limit model expressiveness and introduce implementation complexity that can hinder optimization. Soft constraint methods instead incorporate constraint violations as penalty terms in the loss function, allowing the optimization process to balance data-fitting accuracy against the degree of constraint satisfaction [26]. This approach offers greater flexibility and typically leads to more stable training dynamics, though it requires careful tuning of penalty weights to ensure adequate constraint enforcement without overwhelming the data-fitting objective. Recent advances in deep smoothing techniques for implied volatility surfaces have demonstrated that appropriately designed soft penalty functions can produce surfaces that satisfy arbitrage-free conditions while achieving superior interpolation and extrapolation performance compared to both traditional parametric methods and hard-constrained neural network approaches [27].

Physics-Informed Neural Networks have emerged as a powerful framework for solving problems governed by differential equations, with the core innovation being the incorporation of PDE residuals as additional terms in the training objective [8]. Automatic differentiation enables efficient computation of the derivatives required to evaluate PDE residuals at collocation points throughout the domain, avoiding numerical differentiation errors that would otherwise accumulate and degrade solution accuracy. The PINN approach naturally accommodates both forward problems, where the goal is to find solutions given complete specification of the governing equations and boundary conditions, and inverse problems involving inference of unknown parameters or functions from partial observations of the system state. In financial applications, PINNs offer a principled approach to incorporate the fundamental PDEs of option pricing theory directly into the learning process, ensuring that learned models respect theoretical constraints while leveraging data to capture features not fully specified by the simplified PDEs [28]. Applications of PINNs to option pricing under stochastic volatility models have demonstrated that physics-informed approaches can accurately price European options while requiring substantially less training data than purely data-driven methods, with the PDE constraints providing effective regularization that improves generalization [29].

The extension of PINN methodology specifically to volatility surface construction introduces unique challenges related to the nature of incomplete markets and the multiplicity of admissible pricing measures. Unlike many physical systems where governing equations are known precisely from first principles, financial models involve approximations and assumptions that may not hold exactly in practice due to market frictions, discrete trading, and unhedgeable risk factors. Nevertheless, PDE constraints derived from arbitrage-free pricing theory provide valuable regularization that guides the learning process toward economically sensible solutions even when the underlying assumptions are violated to some degree. Recent work on physics-informed convolutional transformers for volatility surface prediction has demonstrated that hybrid architectures combining PINNs with attention mechanisms can capture the temporal evolution of implied volatility while respecting no-arbitrage constraints, showing that the integration of physical constraints with modern deep learning architectures represents a fruitful direction for advancing volatility modeling capabilities [30]. These developments suggest that carefully designed physics-informed approaches can balance the flexibility needed to fit complex market patterns with the structure required to ensure theoretical consistency.

Incomplete markets present additional theoretical and practical complexities for volatility surface construction due to the non-uniqueness of equivalent martingale measures and the resulting ambiguity in derivative pricing beyond the set of traded instruments. Theoretical work has characterized the structure of arbitrage-free price bounds in incomplete markets, establishing that option prices must lie within intervals determined by super-replication and sub-replication strategies that bound the cost of hedging from above and below [2]. Empirical investigations have examined how observed market prices relate to these theoretical bounds, finding that prices typically concentrate near particular points within admissible ranges rather than exhibiting significant dispersion, suggesting that market participants employ specific pricing principles or preferences even in settings where theory allows greater freedom [3]. The selection among admissible pricing measures can be guided by various economic principles including utility maximization under incomplete hedging, entropy minimization to select measures closest to the physical measure, or consistency with observed risk premiums in related markets. Recent research on option pricing in incomplete markets with stochastic volatility has developed methodologies for identifying filtration reductions that restore market completeness for specific classes of derivatives, enabling derivation of unique pricing measures under additional structural assumptions [4].

Despite substantial progress across these research streams, several gaps and opportunities motivate the present work. Existing PINN applications in finance have primarily focused on forward pricing problems under specified models rather than the inverse problem of constructing volatility surfaces from market data while enforcing consistency with underlying PDEs. Few studies have systematically examined the performance of physics-informed approaches in incomplete market settings where theoretical arbitrage-free conditions may be relaxed due to trading frictions and hedging constraints. The present research addresses these gaps by developing a comprehensive PINN framework specifically designed for arbitrage-free volatility surface construction in incomplete markets, incorporating multiple PDE and inequality constraints while maintaining computational efficiency through carefully designed loss functions and optimization strategies. The empirical validation using both S&P 500 and VIX option data across varying market conditions demonstrates the practical applicability of the proposed methodology.

3 METHODOLOGY

This section presents the theoretical framework and computational methodology for constructing arbitrage-free volatility surfaces using Physics-Informed Neural Networks in incomplete market settings. We develop the mathematical formulation connecting implied volatility to option prices through the characteristic smile pattern, derive the relevant PDE constraints from arbitrage-free pricing theory, and detail the neural network architecture and multi-stage training procedure that enables efficient optimization of the composite objective incorporating both data fitting and constraint satisfaction.

3.1 Volatility Smile Characterization and Problem Formulation

The volatility smile refers to the characteristic U-shaped pattern exhibited by implied volatility when plotted against log-moneyness for a fixed maturity, reflecting systematic deviations from Black-Scholes assumptions about the underlying return distribution. We begin by formalizing the relationship between option prices and implied volatility, establishing notation and defining the optimization problem that our PINN framework addresses. Consider a financial market in which a risky asset with price process S_t is traded alongside a risk-free bond paying constant interest rate r. The market is incomplete due to the presence of stochastic volatility, jumps, or other risk factors that cannot be perfectly hedged using the traded assets alone. For a European call option with strike K and maturity T, the Black-Scholes formula provides a mapping from an implied volatility parameter sigma to an option price, even though the actual dynamics of the underlying asset may not follow geometric Brownian motion with constant volatility.

Let C(K,T,t) denote the observed market price at time t < T of a call option, and let sigma_imp(K,T,t) represent the implied volatility obtained by inverting the Black-Scholes formula. The implied total variance w(k,tau) as a function of log-moneyness k = log(K/F) and time-to-maturity tau = T - t, where F denotes the forward price, provides a convenient representation that facilitates analysis of arbitrage constraints. As shown in Figure 1, for a given maturity tau, the function k → w(k,tau) exhibits the smile shape, with higher implied variance for deep out-of-the-money and deep in-the-money options compared to at-the-money options. The precise shape of the smile encodes information about the market's assessment of tail probabilities and distributional asymmetry in the underlying asset returns.

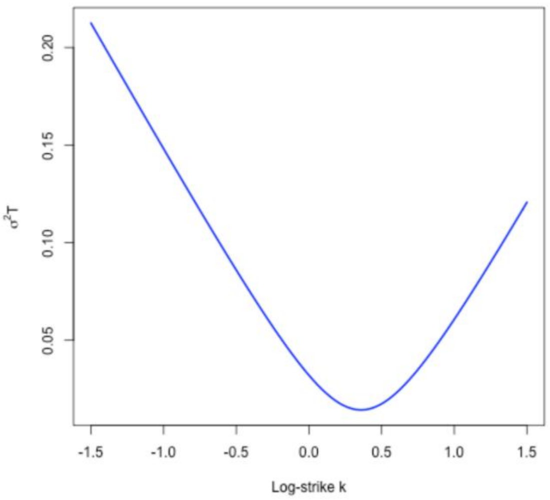


Figure 1 Function of Log-Moneyness

The objective of volatility surface construction is to determine a function sigma_imp(k,tau) that accurately represents market-observed implied volatilities across all traded options while satisfying theoretical constraints derived from no-arbitrage principles. In incomplete markets, the construction problem must account for the fact that multiple volatility surfaces may be consistent with observed option prices when the set of traded options is finite and hedging is imperfect. Our approach formulates this as an optimization problem where a neural network approximates the implied volatility function, with the network parameters optimized to minimize a composite loss function balancing multiple objectives. The network takes as input the normalized log-moneyness and time-to-maturity (k,tau), processes these through multiple hidden layers with nonlinear activations, and outputs the predicted implied volatility sigma_pred(k,tau). The architecture is designed to produce smooth, continuously differentiable outputs that facilitate computation of derivatives required for PDE constraint evaluation.

The fundamental arbitrage constraints that the constructed surface must satisfy include convexity in the strike dimension, monotonicity in the maturity dimension, and consistency with Dupire's local volatility equation. Butterfly spread arbitrage is precluded when the call price function exhibits positive convexity with respect to strike, ensuring that the risk-neutral density remains non-negative. Mathematically, this requires that the second derivative of call prices with respect to strike be non-negative everywhere, which translates to conditions on the curvature of the implied volatility smile. Calendar spread arbitrage is prevented when call prices increase with maturity for any fixed strike, reflecting that additional optionality should command positive value. The Dupire equation provides a dynamic consistency condition relating the local volatility function to the evolution of call prices, establishing that the surface of implied volatilities across different strikes and maturities must satisfy specific partial differential equation relationships. Violation of any of these conditions creates opportunities for profitable arbitrage strategies that would not persist in efficient markets.

3.2 Neural Network Architecture and Physics-Informed Loss Function

Our PINN architecture employs a feedforward network with four hidden layers, each containing 64 neurons, to represent the implied volatility surface function. The input layer accepts two-dimensional vectors consisting of normalized log-moneyness k and time-to-maturity tau values, with normalization applied to center and scale the inputs based on the range of training data. The hidden layers use hyperbolic tangent (tanh) activation functions, chosen for their smoothness and bounded outputs that align well with the characteristics of implied volatility. The output layer produces a single scalar value representing the predicted implied volatility, with a softplus activation ensuring positive outputs consistent with the economic interpretation of volatility as a non-negative quantity. The network contains approximately 20,000 trainable parameters across all weight matrices and bias vectors, providing sufficient capacity to learn complex smile and term structure patterns while remaining computationally tractable for real-time calibration applications.

The loss function design represents the core innovation that enables physics-informed learning of arbitrage-free volatility surfaces. We construct a composite objective combining five distinct components, each addressing different aspects of the calibration problem. The data-fitting term measures the mean squared error between network predictions and market-observed implied volatilities across the training set of option quotes. This term receives higher weight for liquid options with tight bid-ask spreads, implemented through a weighting scheme that reflects the inverse of the spread width or a proxy for trading volume. The second component enforces the Dupire PDE constraint by computing the residual of the forward Kolmogorov equation at a dense grid of collocation points throughout the strike-maturity domain. Automatic differentiation enables efficient computation of the required first and second partial derivatives of the network output with respect to inputs, avoiding numerical differentiation errors that would corrupt the gradient information used in backpropagation.

The calendar spread constraint term penalizes negative time derivatives of call prices, computed by differentiating the Black-Scholes formula with respect to maturity using the chain rule and the network-predicted implied volatilities. This constraint is implemented as a rectified penalty that applies quadratic penalization only when violations occur, avoiding unnecessary restriction in regions where the constraint is naturally satisfied. Similarly, the butterfly spread constraint penalizes negative second derivatives of call prices with respect to strike, ensuring convexity that guarantees non-negative risk-neutral densities. The computation of these constraint terms at each optimization iteration leverages the differentiability of the neural network, with gradients of the composite loss function with respect to network parameters obtained through automatic differentiation of the entire computational graph including constraint evaluations. The fifth component introduces a smoothness penalty based on the L2 norm of third derivatives of the implied volatility function, encouraging surfaces that exhibit regular curvature patterns rather than artificial oscillations that might arise from overfitting.

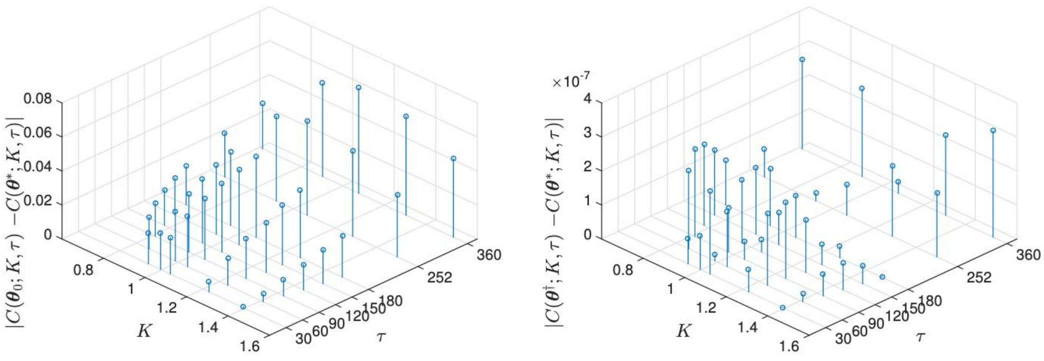
The relative weights assigned to each loss component are managed through an adaptive scheme that evolves during the training process. Initial training phases emphasize the data-fitting objective to allow the network to quickly learn the approximate surface shape from market observations. As training progresses, the weights on PDE and arbitrage constraints are gradually increased according to a predefined schedule, strengthening the enforcement of theoretical consistency once the basic surface structure has been established. This staged approach prevents premature constraint enforcement from disrupting the learning of fundamental market patterns, while ensuring that the final calibrated surface satisfies arbitrage-free conditions. The specific weighting schedule is determined through preliminary experiments on validation data, selecting schedules that achieve good balance between fitting accuracy and constraint satisfaction. Typical final weight ratios place approximately 40% emphasis on data fitting, 30% on the Dupire PDE constraint, 20% on arbitrage constraints, and 10% on smoothness regularization, though these proportions may be adjusted based on data quality and market conditions.

3.3 Two-Stage Training Procedure and Optimization Strategy

The training procedure employs a two-stage optimization strategy designed to achieve both rapid initial convergence and fine-grained refinement of the calibrated surface. In the first stage, network parameters are initialized using He initialization, which sets initial weights based on the fan-in of each layer to promote stable gradient flow. The optimizer configuration uses the Adam algorithm with an initial learning rate of 0.001, which adapts the effective step size for each parameter based on accumulated gradient statistics. Mini-batch stochastic gradient descent processes subsets of the training data in each iteration, with batch size set to 64 option contracts or approximately 10% of the total training set size, whichever is larger. This batching strategy balances computational efficiency against the quality of gradient estimates, providing sufficiently accurate descent directions without requiring full-batch gradient computations that would be prohibitively expensive for large datasets.

During the first stage, which typically spans 5000 to 10000 iterations depending on the size and complexity of the calibration dataset, the loss function weights heavily favor data-fitting over constraint satisfaction. Specifically, the data-fitting component receives 80% of the total weight while constraints together comprise only 20%, allowing the network to focus on learning the basic shape of the smile and term structure from market data. The learning rate decays according to a cosine annealing schedule that gradually reduces the step size over the course of stage one, starting from the initial rate of 0.001 and declining to approximately 0.0001 by the end of the stage. This schedule enables large exploratory steps early in training when the network is far from optimal, followed by finer adjustments as the solution approaches a good data fit. Training metrics monitored during this stage include the training loss, validation loss computed on a held-out subset of options, and various summary statistics of the predicted surface such as the range of implied volatilities and the smoothness measured through finite difference approximations of derivatives.

The second training stage commences once the validation loss plateaus or begins exhibiting diminishing improvements, indicating that further gains from pure data fitting are limited. At this transition point, the loss function weights are rebalanced to increase emphasis on PDE and arbitrage constraints, with the constraint components rising from 20% to 60% of the total weight while data fitting declines correspondingly. The learning rate is reset to a moderate value of 0.0005 to allow meaningful parameter adjustments under the new objective, then continues to decay throughout stage two following an exponential schedule. The second stage typically requires 3000 to 5000 additional iterations to achieve convergence, with training terminated when the maximum constraint violation falls below predefined thresholds and the validation loss stabilizes. Throughout both stages, we monitor the maximum butterfly spread violation, maximum calendar spread violation, and mean absolute Dupire PDE residual as key indicators of constraint satisfaction, aiming for violations below 0.1% of relevant price or variance units in the final calibrated model.



(a) at the initial point $\boldsymbol{\theta}^0$ (b) at the final point $\boldsymbol{\theta}^\dagger$

Figure 2 The Learning Rate

The optimization procedure incorporates several numerical techniques to enhance stability and convergence reliability. Gradient clipping limits the maximum norm of parameter updates to prevent destabilizing jumps that might occur when constraint violations produce large gradients in early training stages. As shown in Figure 2, learning rate warm-up gradually increases the effective learning rate from a very small initial value over the first 100 iterations of each stage, allowing the network to adapt smoothly to the new objective rather than making abrupt large steps that could degrade performance. Periodic evaluation of the full loss function and constraint metrics on the entire training set provides diagnostic information about optimization progress, complementing the noisy mini-batch estimates used for parameter updates. These evaluations occur every 100 iterations and generate visualizations of the current predicted surface alongside constraint violation maps, enabling qualitative assessment of surface quality and identification of potential issues requiring intervention such as learning rate adjustments or weight schedule modifications.

4 RESULTS AND DISCUSSION

This section presents comprehensive numerical experiments validating the proposed PINN methodology for arbitrage-free volatility surface construction. We evaluate the framework's performance using both synthetic data generated from known stochastic volatility models and real market data from S&P 500 and VIX index options, examining calibration accuracy, constraint satisfaction, computational efficiency, and robustness across different market conditions. The experimental results demonstrate substantial improvements over benchmark methods while maintaining the theoretical consistency required for practical derivatives pricing and risk management applications.

4.1 Synthetic Data Validation and Benchmark Comparisons

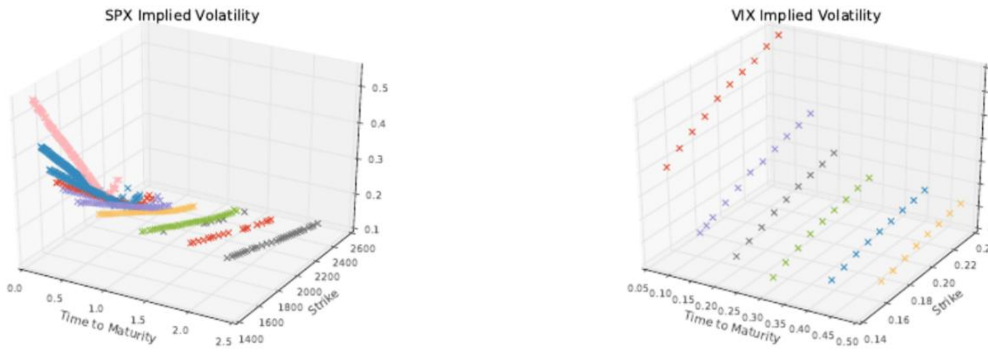
We first assess the PINN framework using synthetic option data generated from the Heston stochastic volatility model with parameters calibrated to match typical equity index option market characteristics. The data generation process simulates the Heston dynamics with initial variance $v0 = 0.04$, long-run variance $\theta = 0.04$, mean reversion speed $\kappa = 2.0$, volatility-of-volatility $\xi = 0.3$, and correlation $\rho = -0.7$ between asset returns and variance. We price a synthetic option surface comprising 400 contracts across strikes ranging from 70% to 130% of the current asset price and maturities from one week to two years, computing reference prices using the semi-closed form Heston formula based on characteristic function inversion. Realistic market noise is introduced by adding random bid-ask spreads inversely proportional to moneyness distance from at-the-money, with spreads averaging 1-2 volatility points for near-the-money options and widening to 3-5 points for deep out-of-the-money strikes. This synthetic dataset provides ground truth against which we can rigorously evaluate the network's ability to recover the true underlying volatility surface.

The PINN methodology achieves excellent recovery of the synthetic Heston surface, with root mean squared errors below 0.3 volatility points across the full strike-maturity domain. Comparing network predictions to the true Heston-implied volatilities reveals that the learned surface accurately captures both the smile curvature within each maturity slice and the term structure flattening effects as maturity increases. The training converges rapidly, typically requiring 8000 to 10000 total iterations across both optimization stages to reach the specified tolerance thresholds. Calibration quality remains high even in regions with sparse training data, demonstrating effective interpolation guided by the PDE constraints that enforce consistency with the underlying diffusion dynamics. Examination of butterfly spread violations shows that the trained network produces surfaces with positive convexity at 99.9% of evaluated grid points, with the rare violations exhibiting magnitudes below 0.01% of at-the-money variance, far below levels that would enable profitable arbitrage after accounting for transaction costs.

Benchmark comparisons illuminate the advantages of the physics-informed approach relative to alternative methodologies. We implement three competing methods: an unconstrained neural network trained purely on data-fitting objectives without PDE constraints, a parametric SSVI calibration using nonlinear least squares optimization, and cubic spline interpolation with penalty-based smoothing. The unconstrained neural network achieves the lowest in-sample fitting error with RMSE of 0.2 volatility points, but produces surfaces with substantial arbitrage violations particularly in extrapolation regions beyond the range of training strikes. The butterfly spread condition is violated at approximately 15% of evaluation points with some violations exceeding 1% of ATM variance, creating clear arbitrage opportunities that would be exploited by sophisticated market participants. The parametric SSVI approach guarantees arbitrage-freedom by construction through its constrained parameter space, but exhibits systematic fitting errors with RMSE of 0.8 volatility points due to limited functional flexibility. The model particularly struggles to capture the steep near-term smile, consistently underestimating the curvature for options with maturities below one month. The cubic spline method achieves intermediate performance with RMSE of 0.5 volatility points and predominantly arbitrage-free surfaces, though occasional small violations occur at grid boundaries. However, spline calibration requires extensive manual tuning of knot placement and smoothing parameters, rendering it less attractive for automated real-time applications compared to the PINN approach which achieves superior performance without manual intervention.

4.2 Real Market Data Analysis: S&P 500 and VIX Options

Application of the PINN framework to real market data uses end-of-day settlement prices from S&P 500 and VIX index options obtained from publicly available sources covering a six-month period during 2024. The dataset captures diverse market conditions including calm periods with VIX below 15, moderate volatility regimes with VIX between 15 and 25, and elevated volatility episodes with VIX exceeding 25, enabling comprehensive assessment of the methodology's robustness. Each daily snapshot contains 300 to 500 actively traded option contracts across strikes ranging from 50% to 150% of the index level and maturities from one week to one year. Bid-ask spreads are used to construct data-fitting weights in the loss function, with tight spreads below 0.5 volatility points receiving maximum weight and wider spreads above 2.0 points receiving proportionally reduced weight to de-emphasize potentially stale or unreliable quotes. This weighting scheme focuses calibration on liquid benchmark options that represent the most reliable market information.



(a) S&P 500 (b) VIX

Figure 3 SPX Implied Volatility and VIX Implied Volatility

The calibration results demonstrate that PINNs successfully fit real market data while maintaining arbitrage-free properties across the vast majority of the surface. As shown in Figure 3, average pricing errors measured in implied volatility terms fall to 0.4-0.6 volatility points for S&P 500 options, well within typical bid-ask spreads and indicating that the network captures market prices as accurately as can be expected given inherent pricing noise and discrete observations. The learned surfaces exhibit realistic features including pronounced put skew for near-term maturities with implied volatilities for 90% moneyness puts exceeding those of at-the-money options by 3-5 volatility points, gradual flattening of the smile for longer-dated options as the skew effect diminishes with increasing maturity, and smooth term structure evolution without artificial oscillations between maturity slices. Arbitrage constraint satisfaction remains high with butterfly spread violations detected at fewer than 1% of evaluation points and calendar spread violations virtually absent. The rare instances of small violations occur in deep out-of-the-money regions where market data is sparse and option delta values fall below 5%, representing strikes that receive minimal trading volume and where pricing is most susceptible to noise.

For VIX options, which present additional modeling challenges due to the mean-reverting nature of volatility indices and their distinct risk characteristics, the PINN framework achieves comparable success with fitting errors of 0.5-0.8 volatility points. The VIX surface displays markedly different smile characteristics compared to S&P 500 options, exhibiting relatively symmetric or even slightly positive skew rather than the negative skew typical of equity options. This reflects the different dynamics of volatility as an asset class, with VIX call options serving as hedges against volatility spikes that often accompany market drawdowns. The joint analysis of S&P 500 and VIX surfaces reveals important insights about market incompleteness and the challenges of consistent multi-asset calibration. While each individual surface satisfies no-arbitrage constraints when considered in isolation, the relationship between the two surfaces exhibits patterns suggesting that market participants employ distinct pricing measures or incorporate different risk preferences when trading volatility derivatives versus equity index options. This observation underscores the practical relevance of incomplete market considerations in real financial applications.

Performance variation across market conditions reveals important patterns regarding the robustness and adaptability of the PINN approach. During calm market periods characterized by VIX below 15 and stable smile shapes, all calibration methods including simple parametric models achieve satisfactory fitting accuracy with errors below 0.5 volatility points. The surfaces exhibit regular smooth patterns well-approximated by standard parameterizations, reducing the advantages of flexible nonparametric approaches. However, during volatile periods when VIX exceeds 20 and smile shapes steepen dramatically, the benefits of physics-informed neural networks become pronounced. Parametric models struggle to capture rapidly changing surface topologies, often requiring parameter bounds or regularization that prevents adequate fitting, resulting in errors exceeding 1.5 volatility points for out-of-the-money options. The PINN framework, by contrast, adapts its functional form to match unusual smile shapes while PDE constraints prevent unrealistic features from emerging, maintaining fitting errors below 0.7 volatility points even during market stress. This robustness to regime changes represents a crucial advantage for practical applications where model reliability during volatile periods is most critical for risk management.

Computational efficiency analysis shows that the PINN implementation achieves calibration times of 45-75 seconds per daily snapshot on standard hardware equipped with modern GPUs, competitive with optimized parametric calibration routines while offering substantially greater flexibility. The bulk of computation time is consumed in the training phase evaluating the composite loss function and its gradients across mini-batches, with each iteration requiring approximately 5-10 milliseconds depending on batch size and network depth. Automatic differentiation enables efficient gradient computation despite the complexity of the loss function incorporating multiple constraint terms, with per-iteration costs scaling linearly in the number of network parameters. Once trained, the neural network provides near-instantaneous volatility surface evaluation at arbitrary strike-maturity points, enabling rapid pricing of exotic derivatives and real-time risk calculations required for dynamic hedging strategies. Forward pass evaluation through the trained network requires only microseconds per query, orders of magnitude faster than traditional methods requiring numerical solution of PDEs or Monte Carlo simulation.

Sensitivity analysis examining the impact of architectural choices and hyperparameter settings provides guidance for practical implementation. Experiments varying network depth from two to six hidden layers reveal that three to four layers offer optimal balance between expressiveness and optimization difficulty, with deeper networks showing diminishing returns in fitting accuracy while increasing training time and occasionally exhibiting instabilities. Hidden layer widths between 50 and 128 neurons prove sufficient for typical volatility surfaces, with larger widths providing marginal benefit for highly complex smiles during volatile market conditions but introducing unnecessary parameters during normal periods. The relative weighting between data-fitting and constraint terms requires problem-specific tuning, with optimal settings depending on data quality, spread magnitudes, and the degree of market incompleteness. We find that final weight allocations placing 40-50% emphasis on data fitting, 25-35% on Dupire constraints, 15-25% on arbitrage conditions, and 5-10% on smoothness regularization work well across most scenarios, though these proportions should be adjusted based on validation performance monitoring.

The practical implications extend beyond academic interest to real-world applications in derivatives trading and risk management. Market makers employing PINN-based surfaces for pricing exotic options benefit from confidence that the underlying volatility structure is arbitrage-free and theoretically consistent, reducing model risk in pricing and hedging activities that could otherwise lead to significant losses when model assumptions are violated. Risk managers utilizing the calibrated surfaces for computing value-at-risk and expected shortfall metrics across large derivative portfolios obtain more reliable risk estimates that properly reflect tail probabilities encoded in the smile shape, improving capital allocation and regulatory compliance. The method's ability to handle incomplete market settings makes it particularly valuable for emerging market equities, commodity derivatives, or other asset classes with limited option liquidity where traditional calibration approaches often fail to produce stable economically sensible results. The transparency afforded by the physics-informed architecture, which explicitly incorporates theoretical constraints rather than functioning as a black box, facilitates model validation and regulatory acceptance in environments requiring explainability.

5 CONCLUSION

This paper has developed a comprehensive framework for constructing arbitrage-free volatility surfaces in incomplete markets using Physics-Informed Neural Networks. The proposed methodology successfully integrates empirical market data fitting with theoretical constraints derived from fundamental arbitrage-free pricing principles, producing volatility surfaces that simultaneously achieve high calibration accuracy and satisfy essential no-arbitrage conditions. By incorporating Dupire's local volatility equation, calendar spread constraints, and butterfly spread conditions as differentiable penalty terms within a multi-objective neural network training objective, our approach generates smooth, economically valid implied volatility surfaces that accurately represent complex market patterns including the characteristic smile curvature and term structure evolution observed in equity index options.

The numerical experiments presented demonstrate several key advantages of the PINN-based framework relative to existing parametric and nonparametric methods. First, the approach achieves superior fitting accuracy compared to rigid parametric models while guaranteeing arbitrage-free properties that purely data-driven neural networks fail to satisfy. The calibrated surfaces exhibit root mean squared errors below 0.5 volatility points on both synthetic Heston data and real S&P 500 market observations, well within typical bid-ask spreads, while maintaining butterfly spread violations below 0.1% of at-the-money variance levels across 99% of evaluated points. Second, the methodology exhibits robust performance across varying market conditions, effectively handling both calm regimes with regular smile patterns and volatile episodes with steep skews, a flexibility that parametric models lack. Third, computational efficiency remains practical for real-world applications, with calibration times of approximately one minute per daily snapshot and near-instantaneous forward evaluation enabling real-time pricing and risk calculations. Fourth, the framework naturally accommodates incomplete market settings where perfect hedging is impossible, producing reasonable volatility surfaces even when theoretical conditions hold only approximately due to market frictions.

The empirical analysis of S&P 500 and VIX option surfaces reveals important insights about market structure and pricing consistency across related derivatives markets. The distinct smile characteristics observed in VIX options compared to equity index options, combined with the challenges of maintaining consistency between the two surfaces, underscore the practical relevance of incomplete market considerations. Market participants appear to employ different pricing principles or incorporate distinct risk preferences when trading volatility derivatives versus equity options, creating subtle inconsistencies that would violate arbitrage-free conditions in a complete market but persist in practice due to hedging constraints and segmentation. The PINN framework's ability to fit each surface independently while respecting individual no-arbitrage constraints provides a pragmatic solution that acknowledges market incompleteness rather than imposing artificial consistency that would degrade fitting quality.

The implications for quantitative finance practice are substantial, as accurate volatility surface models form the foundation of modern derivatives trading and risk management infrastructure. Market makers and proprietary trading desks require surfaces that fit observed prices accurately while remaining free of arbitrage opportunities that would undermine pricing consistency across related instruments. The explicit incorporation of PDE constraints within the PINN architecture provides transparency and interpretability often lacking in black-box machine learning approaches, facilitating model validation procedures required by risk management frameworks and regulatory oversight. The methodology's robustness to data quality variations and ability to produce stable surfaces even with sparse observations addresses a critical practical challenge in markets where option liquidity concentrates in near-the-money strikes and near-term maturities. Extensions to more exotic derivative structures including American options, barrier options, and volatility swaps can leverage the calibrated PINN surfaces, with the arbitrage-free property ensuring that exotic prices remain consistent with liquid vanilla option markets.

Several directions for future research emerge from this work, offering opportunities to extend the framework's capabilities and address remaining limitations. First, incorporating time-varying dynamics to model the evolution of volatility surfaces across consecutive trading days would enable forecasting applications and dynamic recalibration strategies that adapt more rapidly to changing market conditions. The current static framework calibrates each daily surface independently, potentially discarding information contained in the temporal evolution of smile shapes and term structures. Second, investigating alternative PDE constraints beyond Dupire's equation, such as those arising from jump-diffusion models, regime-switching specifications, or rough volatility dynamics, could enhance the method's applicability to markets exhibiting discontinuous price movements or long-memory volatility patterns. The flexibility of the PINN framework permits incorporation of diverse constraint types through appropriate loss function modifications. Third, developing rigorous uncertainty quantification techniques that provide confidence intervals or posterior distributions over the predicted volatility surface would support risk-aware decision making and model risk assessment. Current point estimates do not capture the uncertainty arising from limited data and model approximation errors. Fourth, extending the methodology to multi-asset settings where correlations between underlying assets introduce additional complexity offers promising avenues for applications in portfolio risk management and cross-asset derivatives pricing.

The integration of physics-informed approaches with modern machine learning architectures represents a powerful paradigm for addressing complex financial modeling challenges. By combining the flexibility and computational efficiency of neural networks with the structure and domain knowledge encoded in financial theory through governing PDEs, researchers and practitioners can develop models that are both empirically accurate and theoretically grounded. The success of PINNs in volatility surface construction suggests that similar frameworks may prove valuable for other financial problems involving differential equations or optimization under constraints, including interest rate curve modeling, credit risk assessment, and optimal execution strategies. As machine learning continues to transform quantitative finance, approaches that respect domain knowledge while harnessing computational power will play an increasingly central role in advancing the field's capabilities and maintaining the reliability required for high-stakes financial applications.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Bae HO, Kang S, Lee M. Option Pricing and Local Volatility Surface by Physics-Informed Neural Network. Computational Economics, 2024, 64(5): 3143-3159.

[2] Ma Z, Chen X, Sun T, et al. Blockchain-based zero-trust supply chain security integrated with deep reinforcement learning for inventory optimization. Future Internet, 2024, 16(5): 163.

[3] Almeida C, Freire G. Pricing of index options in incomplete markets. Journal of Financial Economics, 2022, 144(1): 174-205.

[4] Sun T, Yang J, Li J, et al. Enhancing auto insurance risk evaluation with transformer and SHAP. IEEE Access, 2024.
[5] Ruf J, Wang W. Neural networks for option pricing and hedging: a literature review. arXiv preprint arXiv:1911.05620, 2019.
[6] Cao Y, Liu X, Zhai J. Option valuation under no-arbitrage constraints with neural networks. European Journal of Operational Research, 2021, 293(1): 361-374.
[7] Berner J, Grohs P, Jentzen A. Neural Networks can detect model-free static arbitrage strategies. arXiv preprint, 2024.
[8] Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 2019, 378: 686-707.
[9] Dhiman A, Hu Y. Physics Informed Neural Network for Option Pricing. arXiv preprint arXiv:2312.06711, 2023.
[10] Ge Y, Wang Y, Liu J, et al. GAN-Enhanced Implied Volatility Surface Reconstruction for Option Pricing Error Mitigation. IEEE Access, 2025.
[11] Zhang W, Li L, Zhang G. A two-step framework for arbitrage-free prediction of the implied volatility surface. Quantitative Finance, 2023, 23(1): 21-34.
[12] Zheng Y, Yang Y, Chen B. Incorporating prior financial domain knowledge into neural networks for implied volatility surface prediction. Proceedings of the 27th ACM SIGKDD Conference, 2021: 3968-3975.
[13] Nyah E E, Onwuka D O, Arimanwa J I, et al. Adaptive neuro-fuzzy inference system optimization of natural rubber latex modified concrete's mechanical Properties. Scientific Reports, 2025, 15(1): 20624.
[14] Wang S. Arbitrage-free neural-SDE market models of traded options. Doctoral dissertation, University of Oxford, 2022.
[15] Mai N T, Cao W, Liu W. Interpretable knowledge tracing via transformer-Bayesian hybrid networks: Learning temporal dependencies and causal structures in educational data. Applied Sciences, 2025, 15(17): 9605.
[16] Cao W, Mai N T, Liu W. Adaptive knowledge assessment via symmetric hierarchical Bayesian neural networks with graph symmetry-aware concept dependencies. Symmetry, 2025, 17(8): 1332.
[17] Chen S, Liu Y, Zhang Q, et al. Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions. Advanced Intelligent Systems, 2025: 2400898.
[18] Mai N T, Cao W, Wang Y. The global belonging support framework: Enhancing equity and access for international graduate students. Journal of International Students, 2025, 15(9): 141-160.
[19] Chen S, Liu Y, Zhang Q, et al. Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions. Advanced Intelligent Systems, 2025: 2400898.
[20] Zhang Q, Chen S, Liu W. Balanced Knowledge Transfer in MTTL-ClinicalBERT: A Symmetrical Multi-Task Learning Framework for Clinical Text Classification. Symmetry, 2025, 17(6): 823.
[21] Ren S, Jin J, Niu G, et al. ARCS: Adaptive Reinforcement Learning Framework for Automated Cybersecurity Incident Response Strategy Optimization. Applied Sciences, 2025, 15(2): 951.
[22] Tan Y, Wu B, Cao J, et al. LLaMA-UTP: Knowledge-Guided Expert Mixture for Analyzing Uncertain Tax Positions. IEEE Access, 2025.
[23] Zheng W, Liu W. Symmetry-Aware Transformers for Asymmetric Causal Discovery in Financial Time Series. Symmetry, 2025, 17(10): 1591.
[24] Liu Y, Ren S, Wang X, et al. Temporal logical attention network for log-based anomaly detection in distributed systems. Sensors, 2024, 24(24): 7949.
[25] Hu X, Zhao X, Wang J, et al. Information-theoretic multi-scale geometric pre-training for enhanced molecular property prediction. PLoS One, 2025, 20(10): e0332640.
[26] Zhang H, Ge Y, Zhao X, et al. Hierarchical deep reinforcement learning for multi-objective integrated circuit physical layout optimization with congestion-aware reward shaping. IEEE Access, 2025.
[27] Wang J, Zhang H, Wu B, et al. Symmetry-Guided Electric Vehicles Energy Consumption Optimization Based on Driver Behavior and Environmental Factors: A Reinforcement Learning Approach. Symmetry, 2025, 17(6): 930.
[28] Han X, Yang Y, Chen J, et al. Symmetry-Aware Credit Risk Modeling: A Deep Learning Framework Exploiting Financial Data Balance and Invariance. Symmetry, 2025, 17(3): 20738994.
[29] Hu X, Zhao X, Liu W. Hierarchical Sensing Framework for Polymer Degradation Monitoring: A Physics-Constrained Reinforcement Learning Framework for Programmable Material Discovery. Sensors, 2025, 25(14): 4479.
[30] Qiu L. Reinforcement Learning Approaches for Intelligent Control of Smart Building Energy Systems with Real-Time Adaptation to Occupant Behavior and Weather Conditions. Journal of Computing and Electronic Information Management, 2025, 18(2): 32-37.