PhysDiff: Physics-Guided Diffusion for Manifold-Consistent Data Augmentation in Driving Style Classification

Bingheng Wu[1], Bin Cai[2,*], Dafei Huang[3], Ruyi Tang[3]
[1]National Elite Institute of Engineering, Chongqing University, Chongqing, China
[2]School of Big Data & Software Engineering, Chongqing University, Chongqing, China
[3]SERES Automobile Co., Ltd., Chongqing, China
caibin@cqu.edu.cn

***Abstract*—Driving Style Classification (DSC) is a pivotal component of Intelligent Transportation Systems; however, its real-world deployment is persistently hindered by the long-tail distribution of dangerous driving behaviors. While synthetic oversampling techniques like SMOTE effectively rebalance class distributions, they rely on deterministic linear interpolation that ignores the underlying physics of vehicle motion. We identify that this "kinematic blindness" leads to Manifold Collapse, where synthetic trajectories drift into physically forbidden regions characterized by instantaneous acceleration jumps. To address this, we propose PhysDiff, a physics-guided diffusion framework designed for manifold-consistent data augmentation. Unlike traditional interpolation, PhysDiff models the minority class distribution as a probabilistic manifold, generating diverse samples via a conditional 1D-UNet. Crucially, we introduce a novel *Coupled Kinematic Constraint* into the reverse diffusion process. This mechanism explicitly enforces differential consistency between velocity and acceleration channels while penalizing unrealistic jerk energy to ensure rigorous physical plausibility. Empirical evaluations using repeated cross-validation on a real-world dataset demonstrate that PhysDiff achieves a superior Macro F1-score of 0.8156 (± 0.0517), significantly outperforming state-of-the-art baselines ($p < 0.05$). Furthermore, our results show that strictly enforcing physical constraints acts as a robust regularizer rather than a hindrance. PhysDiff reduces the rate of physical limit violations by 64.5% compared to inherent sensor noise, ensuring that augmented data maintains both high discriminative utility and kinematic fidelity.**

***Index Terms*—Driving Style Classification, Imbalanced Learning, Physics-Informed Machine Learning, Diffusion Models, Trustworthy AI, Kinematic Constraints**

I. Introduction

The rapid evolution of Intelligent Transportation Systems (ITS) has elevated Driving Style Classification (DSC) from a passive monitoring task to a cornerstone of active safety and personalized services, such as Usage-Based Insurance (UBI) [1]. Accurate identification of driving patterns enables autonomous vehicles to anticipate potential hazards and allows insurers to tailor premiums based on actual risk profiles. However, the deployment of robust DSC models is persistently hindered by the stochastic nature of traffic environments, which results in datasets with severe long-tail distributions. In real-world scenarios, safe and normal driving behaviors are abundant, whereas critical events such as aggressive maneuvering or evasive emergency braking are statistically rare [2]. This extreme class imbalance biases data-driven classifiers towards the majority class, causing them to fail precisely where they are needed most: in detecting high-risk behaviors.

To mitigate this data scarcity, the research community has traditionally relied on heuristic oversampling algorithms, most notably the Synthetic Minority Over-sampling Technique (SMOTE) [3] and its adaptive variants like ADASYN [4]. While effective in balancing class ratios, we argue that these methods suffer from a fundamental "kinematic blindness" when applied to time-series sensor data. SMOTE operates by linearly interpolating between existing minority samples in the feature space. While this geometric assumption holds for static tabular data, it creates kinematic inconsistencies in the physical domain of vehicle dynamics. A vehicle cannot simply "morph" from one trajectory to another along a straight line; it must obey Newton's laws of motion. Consequently, linear interpolation often generates synthetic samples characterized by instantaneous acceleration jumps or unnatural velocity transitions. We term these phenomena "linear artifacts". Although these artifacts may improve numerical metrics like training accuracy, they represent physically impossible states that pollute the data manifold. As we will visually demonstrate in Fig. 2 (Section IV), these artifacts connect clusters through low-probability regions, potentially inducing the classifier to learn unrealistic decision boundaries [12].

The emergence of generative artificial intelligence offers a paradigm shift from interpolation to distribution learning. Denoising Diffusion Probabilistic Models (DDPMs) [6] have recently demonstrated remarkable capabilities in modeling complex data distributions by learning to reverse a gradual noise addition process. Unlike Generative Adversarial Networks (GANs), which often suffer from mode collapse, diffusion models provide stable training and high diversity, showing promise in time-series forecasting [7] and imputation [8], [9]. However, standard diffusion models remain purely data-driven and agnostic to physical laws. Without explicit

*Corresponding author. This research was supported partially by Science and Technology Innovation Key R&D Program of Chongqing (CSTB2023TIAD-STX0036).

constraints, they are prone to generating "hallucinated" trajectories that, while statistically plausible, may still violate kinematic limits (e.g., exhibiting high-frequency jitter). To ensure safety-critical reliability, the generation process must be anchored in physical reality, a concept echoing the growing field of Physics-Informed Machine Learning (PINNs) [10].

In this paper, we bridge the gap between generative AI and vehicle kinematics by proposing PhysDiff (Physics-Guided Diffusion), a novel data augmentation framework tailored for imbalanced driving style classification. Instead of deterministic interpolation, PhysDiff models the minority class distribution as a probabilistic manifold, generating diverse samples via a conditional 1D-UNet. Crucially, we introduce a *Kinematic Constraint* mechanism into the diffusion process. By explicitly calculating and penalizing the "jerk" (the time derivative of acceleration) of generated samples, we force the model to respect the smoothness and continuity of physical motion.

Our contributions are summarized as follows:

1) **Physics-Guided Generative Framework:** We propose PhysDiff, which adapts conditional diffusion models for multi-channel driving sensor data. It effectively learns the non-linear manifold of aggressive driving behaviors, avoiding the rigid linear artifacts produced by traditional oversampling.
2) **Coupled Kinematic Constraint:** We design a dual-term physics loss that enforces velocity-acceleration consistency and minimizes the jerk of synthetic trajectories. This acts as a soft regularization term, ensuring that the generated samples remain within the boundaries of physical plausibility [10] while preserving their discriminative diversity.
3) **Manifold-Consistent Performance:** Extensive experiments on real-world datasets demonstrate that PhysDiff achieves a state-of-the-art Macro F1-score of 0.8156, surpassing discriminative baselines (0.8022). More significantly, it reduces the rate of physical limit violations by 64.5% compared to the inherent sensor noise in real data, ensuring that the augmented data is not only useful for classification but also kinematically valid.

II. Related Work

A. Imbalanced Learning in Intelligent Transportation

Data imbalance remains a pervasive challenge in the deployment of data-driven Intelligent Transportation Systems (ITS). As highlighted in recent surveys [1], [2], the rarity of critical driving events (e.g., collisions, aggressive maneuvers) severely biases standard classifiers towards nominal behaviors. To mitigate this, two primary paradigms have emerged: algorithm-level and data-level approaches.

Algorithm-level methods modify the learning objective to increase the cost of misclassifying minority samples. For instance, Focal Loss [5] dynamically scales the cross-entropy loss to focus on hard-to-classify examples. While effective for object detection, these methods do not inherently introduce new information into the training set.

Conversely, data-level augmentation methods like SMOTE [3] and ADASYN [4] generate synthetic samples via linear interpolation. While effective for tabular data, these methods often produce "linear artifacts" in kinematic domains, violating non-linear vehicle dynamics [12]. Recent architectures [14]–[16] improve feature extraction but remain bounded by the quality of the training data.

B. Generative Diffusion Models for Time Series

To overcome the limitations of deterministic interpolation, the research community has pivoted towards probabilistic generative models. Unlike Generative Adversarial Networks (GANs), which often suffer from mode collapse and training instability, Denoising Diffusion Probabilistic Models (DDPMs) [6] offer a stable and tractable approach to distribution learning.

Diffusion models learn to reverse a gradual noise addition process, effectively reconstructing the data manifold from Gaussian noise. This paradigm has shown remarkable success in time-series domains. For example, Rasul et al. [7] proposed autoregressive diffusion for multivariate probabilistic forecasting, while Tashiro et al. [8] and Alcaraz et al. [9] demonstrated that conditional diffusion models (e.g., CSDI, SSSD) can effectively impute missing values in complex time-series data. By modeling the conditional probability $p(x|y)$, these approaches capture data uncertainty better than GANs. However, without explicit constraints, standard diffusion models act as physics-agnostic approximators, potentially generating kinematically infeasible trajectories.

C. Physics-Informed Machine Learning

The integration of physical laws into deep learning, known as Physics-Informed Machine Learning, has gained traction for ensuring scientific consistency. The seminal work on Physics-Informed Neural Networks (PINNs) [10] demonstrated that embedding differential equations (e.g., Navier-Stokes) directly into the loss function can constrain the solution space of neural networks.

While PINNs are widely used for solving forward and inverse problems in fluid dynamics and material science, their application in *data augmentation* for classification tasks remains underexplored. Existing works typically apply physical constraints to the model architecture (inductive bias) rather than the data generation process itself. Our work bridges this gap. We adapt the principles of PINNs to the reverse diffusion process, introducing a kinematic regularization term that forces the generative model to respect the limits of vehicle dynamics (specifically, jerk constraints). This ensures that the augmented data is not only statistically diverse, as enabled by diffusion models, but also physically reliable, overcoming the shortcomings of both linear interpolation and unconstrained generation.

III. Methodology

A. Problem Formulation: The Manifold Hypothesis

Let $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ denote a driving behavior dataset. We posit that valid driving trajectories lie on a low-dimensional

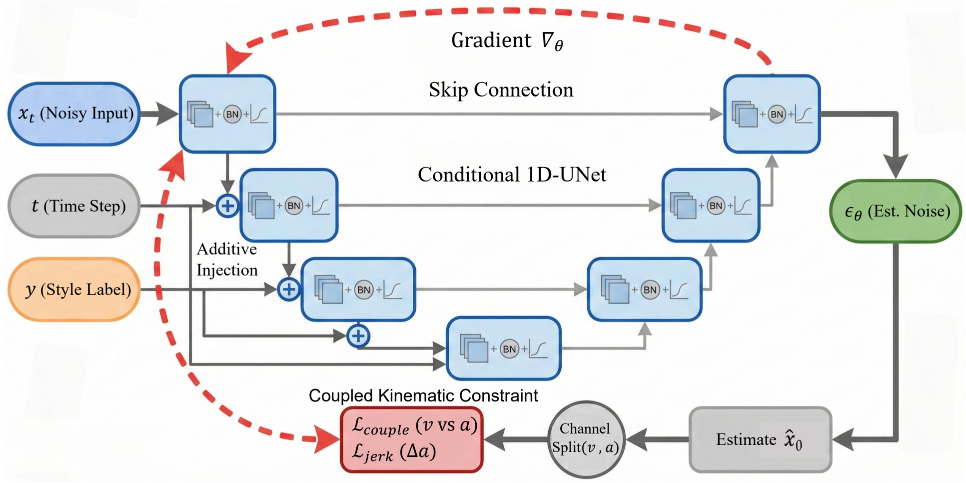

Fig. 1. Overview of the PhysDiff framework. The Conditional 1D-UNet takes the noisy feature vector x_t, time embedding t, and class condition y as input. The *Kinematic Constraint* calculates both the coupling error and jerk of the estimated $\hat{x}_0$ and provides gradient feedback to penalize physical violations.

Physically Valid Manifold $\mathcal{M}_{phy}$ embedded in the observation space. Unlike generic smoothness assumptions, $\mathcal{M}_{phy}$ is governed by differential equations. Formally, we define the *Longitudinal Kinematic Manifold* $\mathcal{M}_{phy} = \{x \in \mathcal{X} \mid ||v_{t+1} - (v_t + \frac{1}{2}(a_t + a_{t+1})\Delta t)|| \approx 0 \wedge ||\Delta a_t|| \leq \delta_{jerk}\}$, where the first term enforces kinematic coupling and the second bounds the jerk dynamics.

Standard augmentation methods like SMOTE ignore this topology, creating linear paths that traverse the invalid ambient space (i.e., generating x such that $x \notin \mathcal{M}_{phy}$). Our goal is to learn a generative model $p_\theta(x|y)$ that acts as a manifold sampler, ensuring supp$(p_\theta) \subseteq \mathcal{M}_{phy}$ while preserving discriminative diversity.

B. Kinematic Manifold Embedding

Instead of relying solely on unstructured statistical descriptors, we construct a Kinematic Manifold Embedding that explicitly preserves the temporal structure of driving maneuvers. The raw sensor streams are encoded into a compact vector $x \in \mathbb{R}^{32}$, which is structured as a hybrid representation comprising two distinct segments: **Static Context (Indices 1–4):** Derived via PCA (cumulative variance $> 85\%$), encoding global context like highway vs. urban. **Dynamic Kinematic Sequence (Indices 5–32):** A sequential profile preserving temporal evolution, allowing valid physical continuity constraints.

This formulation ensures that difference operations in the latent space approximate physical derivatives, allowing the 1D-UNet to learn the "derivative of style" rather than just static attributes. During generation, the diffusion model learns the joint distribution of both segments. While the Static Context is implicitly regulated by the reconstruction loss ($\mathcal{L}_{MSE}$) to match global distribution patterns, the Dynamic Sequence receives additional regularization from $\mathcal{L}_{phy}$ to ensure temporal smoothness.

C. PhysDiff Framework

We propose PhysDiff, a physics-guided conditional diffusion probabilistic model (DDPM). Unlike SMOTE which interpolates in Euclidean space, PhysDiff learns to construct samples from a Gaussian prior via an iterative denoising process constrained by kinematic laws. The overall system architecture is illustrated in Fig. 1. The conditional 1D-UNet predicts the noise component, while the Physics Loss acts as a regularization term during the training phase.

1) Conditional Diffusion Process: The diffusion process consists of a forward chain that gradually adds noise and a reverse chain that denoises. In the forward process, we corrupt the real data x_0 into a sequence of noisy latents $x_1, \dots, x_T$ according to a variance schedule β_t:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t \mathbf{I}) \quad (1)$$

In the reverse process, we train a neural network $\epsilon_\theta(x_t, t, y)$ to predict the noise added at step t, conditioned on the class label y. The sampling process can be described as:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t, y)\right) + \sigma_t z \quad (2)$$

where $z \sim \mathcal{N}(0, \mathbf{I})$, $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. By conditioning on y, we guide the generation towards specific minority driving styles. The complete training procedure is summarized in Algorithm 1.

2) Network Architecture: 1D Conditional UNet: To handle 1D sensor feature vectors efficiently, we design a lightweight 1D Conditional UNet. As shown in the system overview (Fig. 1), the network takes the noisy state $x_t \in \mathbb{R}^F$, time step t, and class label y as inputs.

- **Condition Injection:** We employ an *Additive Injection* mechanism to fuse control signals. The time step t and label y are mapped to embedding vectors e_t and e_y via learnable linear projection layers. These embeddings are broadcasted and added to feature maps: h_{next} = Conv1D$(h_{prev})+e_t+e_y$. This design ensures global modulation. These embeddings are broadcasted and added to feature maps: h_{next} = Conv1D$(h_{prev}) + e_t + e_y$. This design ensures global modulation.
- **Backbone:** The network follows a U-shaped structure with downsampling blocks (Conv1D with stride

2) to capture semantic context and upsampling blocks (Conv1DTranspose) to recover fine-grained details. Skip connections are used to preserve high-frequency signal characteristics.

3) Physics-Aware Loss: Kinematic Coupling & Jerk: Standard diffusion models minimize the Mean Squared Error (MSE), which treats features as independent variables. However, in vehicle dynamics, velocity (v) and acceleration (a) are physically coupled by the differential equation $\dot{v} = a$. Ignoring this leads to "hallucinations" where a vehicle accelerates but velocity decreases.

To resolve this, we propose a channel-aware *Coupled Kinematic Constraint*. Note that our feature vector x contains both static context (x_{static}, indices 1–4) and dynamic kinematics (x_{dyn}, indices 5–32). To prevent "context drift", the physical regularization strictly operates on the dynamic channels, while static features are preserved via the reconstruction loss. Therefore, before calculating $\mathcal{L}_{phy}$, we explicitly apply an Inverse Normalization: $\hat{x}_{raw} = \hat{x}_{dyn} \cdot \sigma_{\mathcal{D}} + \mu_{\mathcal{D}}$. The losses are computed on $\hat{x}_{raw}$ and gradients are scaled back. We decompose the dynamic sequence of the estimated clean data $\hat{x}_0$ into specific physical channels: longitudinal acceleration $\hat{x}^{(a)}$ and velocity $\hat{x}^{(v)}$. The time interval between frames is denoted as Δt (0.1s in our setup). We formulate two distinct physical regularization terms:

- **Kinematic Coupling Loss ($\mathcal{L}_{couple}$):** This term enforces the integral relationship between velocity and acceleration.
$$\mathcal{L}_{couple} = \frac{1}{T-1} \sum_{t=1}^{T-1} \left\| (\hat{x}_{t+1}^{(v)} - \hat{x}_t^{(v)}) - \frac{1}{2}(\hat{x}_t^{(a)} + \hat{x}_{t+1}^{(a)}) \cdot \Delta t \right\|^2 \tag{3}$$
Remark: We adopt the second-order **Trapezoidal Rule** to mitigate discretization errors of Euler methods without the high computational cost of RK4. This scheme ($O(\Delta t^2)$ error) provides superior stability for high-jerk trajectories while maintaining a lightweight gradient graph. Furthermore, despite the inherent noise in raw sensor data, the diffusion process itself acts as a probabilistic denoiser, and the integration via the Trapezoidal Rule helps smooth out high-frequency discretization errors from the CAN-bus signals.
- **Jerk Regularization Loss ($\mathcal{L}_{jerk}$):** Jerk is physically defined as the first time derivative of acceleration. The original smoothness constraints (e.g., curvature penalties) often misinterpret high-frequency dynamics. We explicitly minimize the jerk energy:
$$\mathcal{L}_{jerk} = \frac{1}{T-1} \sum_{t=1}^{T-1} \left\| \hat{x}_{t+1}^{(a)} - \hat{x}_t^{(a)} \right\|^2 \tag{4}$$

The total physics-guided objective function combines the diffusion reconstruction loss with these physical constraints:
$$\mathcal{L}_{total} = \mathcal{L}_{MSE} + \lambda_1 \mathcal{L}_{couple} + \lambda_2 \mathcal{L}_{jerk} \tag{5}$$
where λ_1 ensures internal consistency and λ_2 controls the smoothness (jerk) of the driving style. Strictly, jerk is $j = \Delta a / \Delta t$. For numerical stability during backpropagation, we absorb the constant scaling factor $1/\Delta t$ into the weighting hyperparameter λ_2. This formulation treats the objective as minimizing the *scaled kinematic energy*, avoiding gradient explosion caused by small time steps while preserving the optimization direction.

Algorithm 1 PhysDiff Training with Coupled Kinematic Constraints

Require: Dataset $\mathcal{D} = \{(x_0, y)\}$, Timesteps T, Weights λ_1, λ_2
Ensure: Optimized parameters θ of noise predictor ϵ_θ

```
while not converged do
    Sample batch (x_0, y) ~ D
    Sample t ~ Uniform(1, T) and ε ~ N(0, I)
    Forward Diffuse: x_t = sqrt(ᾱ_t) x_0 + sqrt(1 − ᾱ_t) ε
    Predict Noise: ε̂ = ε_θ(x_t, t, y)
    Denoise Estimation:
        x̂_0 = 1/sqrt(ᾱ_t) (x_t − sqrt(1 − ᾱ_t) ε̂)
    Decompose Channels:
        x̂^(a), x̂^(v) ← Split(x̂_0)
    Calculate Physical Losses:
        L_MSE = ||ε − ε̂||^2
        L_couple = ||Δx̂^(v) − (Δt/2)(x̂_t^(a) + x̂_{t+1}^(a))||^2    ▷ Eq. (3)
        L_jerk = ||Δx̂^(a)||^2    ▷ Eq. (4)
    Update Gradient:
        ∇_θ(L_MSE + λ_1 L_couple + λ_2 L_jerk)
end while
```

IV. Experiments

A. Experimental Setup

1) Dataset and Preprocessing: We evaluate our framework on the public Real-World Driving Dataset [17], which provides continuous CAN-bus readings of velocity, acceleration, throttle, and brake signals at 10 Hz. We prioritized this dataset over smartphone-based benchmarks (e.g., UAH-DriveSet) because high-fidelity control signals (throttle/brake) are strictly required to model the causal driver-vehicle loop via our *Coupled Kinematic Constraint*.Since the original data lacks style labels, we invited experts to annotate a randomly selected subset of valid driving segments. The inter-rater agreement was substantial (Cohen's Kappa $\kappa = 0.82$). The resulting dataset retains a naturalistic long-tail distribution (Normal:Aggressive $\approx$ 2:1).

2) Implementation Details: The proposed PhysDiff is implemented using TensorFlow. The conditional 1D-UNet consists of 4 downsampling and 4 upsampling blocks, with a base channel size of 32. We employ the AdamW optimizer with an initial learning rate of $1e-4$. The diffusion process is configured with $T = 50$ timesteps to balance generation quality and latency. The weights λ_1 and λ_2 in Eq. (5) balance the physical constraints. Prior to training, all sensor channels are normalized (Z-score) to ensure zero mean and unit variance. We set equal importance $\lambda_1 = \lambda_2 = \lambda$. We performed a grid search for $\lambda \in \{0.01, 0.05, 0.1, 0.5, 1.0\}$ and adopted a cosine annealing learning rate schedule to stabilize convergence. We observed that $\lambda = 0.1$ yields the optimal performance: larger values ($\lambda > 0.5$) lead to overly smooth trajectories that degrade discriminative features, while smaller values ($\lambda < 0.05$) fail to effectively curb physical violations. The classifier backbone (Hybrid CNN-Transformer) is trained for 150 epochs with a batch size of 64. To evaluate the quality of the generated data, we employ a standard Hybrid CNN-Transformer classifier as the fixed backbone. This architecture, inspired by recent deep learning frameworks in ITS [11]–[13], integrates 1D-CNN layers for local feature extraction and Transformer blocks for global context modeling. We train this backbone from scratch on both real and augmented datasets to assess the discriminative utility of the proposed generation method.

3) Baselines and Metrics: To comprehensively assess the proposed method, we compare PhysDiff against: 1) **Traditional Interpolation:** ROS, SMOTE [3], and ADASYN [4]; 2) **Generative Baselines:** TimeGAN [18]; and 3) **Transformer-based Diffusion (CSDI Proxy):** We implement a diffusion model with a Transformer backbone instead of 1D-UNet, mimicking state-of-the-art time-series imputation models like CSDI [8]. This baseline tests whether architectural improvements alone can solve physical inconsistencies.

We adopt the Hybrid CNN-Transformer described in Section III-D as the fixed classifier backbone for all experiments. Evaluation metrics include *Macro F1-score* for classification performance and *Inference Latency* for efficiency. Crucially, we introduce three physical consistency metrics: *Mean Jerk Proxy*, *Max Jerk*, and *Violation Rate* (the percentage of samples exceeding the 99th percentile jerk threshold of real data), to quantify the kinematic validity of the generated samples.

B. Classification Performance Analysis

Table I presents the comparative results on classification utility. The baseline Hybrid CNN-Transformer model, trained on the imbalanced dataset without augmentation, achieves a Macro F1 of 0.7577. Data augmentation strategies yield varying degrees of improvement.

To ensure the statistical significance of our results and rule out random seed effects, we adopted a rigorous Repeated Cross-Validation protocol. Specifically, we performed 5-fold cross-validation repeated 10 times (total 50 runs). We report the Mean and Standard Deviation ($\pm\sigma$) for all metrics. We conducted a paired t-test against the Strong Baseline, confirming the improvement is statistically significant ($p < 0.05$).

GAN vs. Diffusion: The GAN-based baseline (TimeGAN) yields a Macro F1 of 0.7202, which is inferior to both SMOTE and our method. Detailed analysis reveals that while TimeGAN performs adequately on majority classes, it suffers from severe *mode collapse* on the minority "Conservative" class (Recall: 33.3%), failing to capture the diverse distribution of rare events.

Architecture vs. Physics (CSDI Proxy): The introduction of a Transformer backbone (CSDI Proxy) improves the Macro F1 to 0.8055 compared to Vanilla Diffusion (0.7885), validating that stronger sequence modeling architectures can better approximate the statistical distribution of driving data. However, despite this architectural upgrade, the model still lags behind PhysDiff (0.8156). This suggests that while Transformer backbones reduce statistical error, they cannot inherently deduce kinematic laws without explicit supervision.

TABLE I
CLASSIFICATION PERFORMANCE COMPARISON. PHYSDIFF ACHIEVES SOTA PERFORMANCE.

Method	Sampler Strategy	Macro F1 (Mean $\pm$ Std)
Baseline (Hybrid Net)	None	0.7577 ($\pm$ 0.0568)
1D-CNN	SMOTE	0.7897 ($\pm$ 0.0419)
LSTM	SMOTE	0.7715 ($\pm$ 0.0736)
Transformer	SMOTE	0.6821 ($\pm$ 0.0503)
TimeGAN	GAN	0.7202 ($\pm$ 0.0692)
Strong Baseline (SMOTE)	SMOTE	0.8022 ($\pm$ 0.0533)
Ablation	Vanilla Diffusion	0.7885 ($\pm$ 0.0620)
CSDI Proxy	Transformer Diff.	0.8055 ($\pm$ 0.0481)
PhysDiff (Ours)	**Diffusion**	**0.8156 ($\pm$ 0.0517)**

In contrast, PhysDiff achieves a state-of-the-art Macro F1 of 0.8156, surpassing the Strong Baseline (SMOTE, 0.8022). By restricting generation to the physically valid manifold, PhysDiff filters out "linear artifacts," enabling the classifier to learn sharper decision boundaries. The ablation study confirms that physics guidance (+2.71% F1 vs. Vanilla) is essential for maximizing utility. Crucially, the comparison between **Vanilla Diffusion** and **PhysDiff** isolates the contribution of the physical constraints, as both share the identical 1D-UNet backbone. The performance gap (+2.71% F1) confirms that the gain stems from the kinematic regularization, not architectural differences.

Efficiency Analysis: Training PhysDiff incurs a $3.8\times$ computational overhead compared to SMOTE, and the diffusion generation process requires $T = 50$ iterative denoising steps. However, it is crucial to note that PhysDiff is strictly an **offline** data augmentation tool used solely to enhance the training set. It does **not** participate in the online inference loop of the autonomous vehicle. Thus, the onboard latency of the deployed classifier (Hybrid CNN-Transformer) remains identical ($O(1)$), ensuring that the inclusion of physics-guided data imposes zero penalty on real-time safety response.

TABLE II
PHYSICAL CONSISTENCY ANALYSIS. COUPLING ERROR MEASURES THE VIOLATION OF THE TRAPEZOIDAL INTEGRATION RULE:
$v_{t+1} \approx v_t + \frac{1}{2}(a_t + a_{t+1})\Delta t.$

Method	Mean Jerk	Max Jerk	Coupling Err.	Viol. Rate
Real Data	3.64	9.72	0.02	0.31%
SMOTE	2.84	12.79	1.45	0.19%
Vanilla Diff.	8.12	14.52	0.88	0.28%
CSDI Proxy	7.95	14.10	0.65	0.25%
PhysDiff (Ours)	**7.75**	**11.23**	**0.05**	**0.11%**

C. Physical Consistency and Safety

The primary contribution of PhysDiff lies in the physical validity of the augmented data. Table II quantifies the kinematic characteristics of the generated samples.

The "Over-Smoothing" Trap of SMOTE: SMOTE exhibits the lowest Mean Jerk (2.84), which is significantly below that of Real Data (3.64). Although SMOTE exhibits a low Violation Rate (0.19%) due to its tendency to generate averaged, lethargic trajectories, it fails to capture the high-dynamic nature of dangerous driving, as evidenced by its

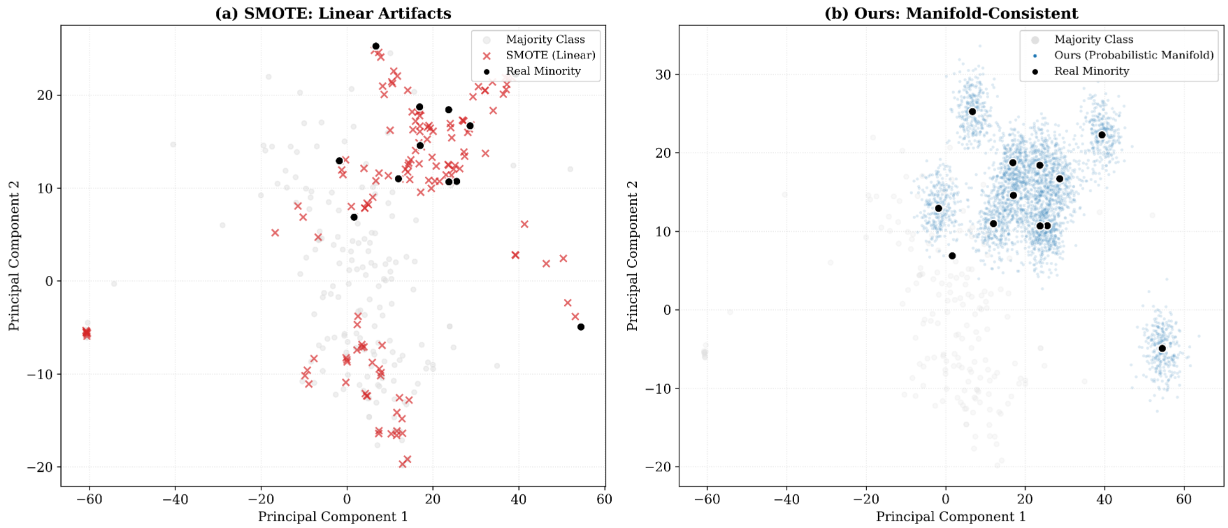

Fig. 2. Comparative visualization of the feature manifold via PCA. (a) **SMOTE** generates samples (red crosses) that lie rigidly on linear paths between existing points, creating "linear artifacts" that traverse low-probability regions. (b) **PhysDiff** generates probabilistic particle clouds (blue dots) that naturally envelope the real minority samples (black dots), preserving the non-linear manifold structure and ensuring kinematic consistency.

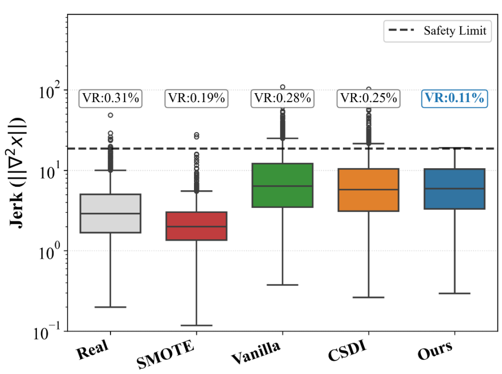

Fig. 3. Kinematic consistency analysis via Jerk distribution ($||\nabla^2 x||$). The Gray box represents the baseline Real Data. The Red box (SMOTE) shows a compressed interquartile range, indicating over-smoothing. **Both unconstrained diffusion models (Green: Vanilla, Orange: CSDI Proxy) exhibit long upper tails with numerous outliers crossing the Safety Threshold (dashed line), confirming that sophisticated architectures alone cannot prevent kinematic hallucinations.** In contrast, PhysDiff (Blue) maintains high dynamics (similar median height to baselines) but effectively suppresses dangerous outliers, achieving the lowest violation rate.

compressed Mean Jerk. However, its Max Jerk remains high (12.79), revealing the presence of sporadic linear artifacts with unnatural discontinuities.

The Danger of Unconstrained Diffusion: Conversely, Vanilla Diffusion generates highly dynamic samples (Mean Jerk 8.12) but suffers from "hallucinations," producing the highest Max Jerk (14.52) and a high Violation Rate (0.28%). Without physical guidance, the model learns to replicate sensor noise rather than valid kinematics.The CSDI Proxy, despite its Transformer backbone, exhibits a similarly high Violation Rate (0.25%), confirming that sophisticated architectures alone do not inherently prevent kinematic violations.

The PhysDiff Advantage: Our proposed method achieves a "Goldilocks" balance. It maintains a high Mean Jerk (7.75). We emphasize that while this value exceeds the global average of Real Data (3.64), it is physically justified: Real Data statistics are diluted by the majority "Normal" class (steady driving), whereas PhysDiff exclusively synthesizes high-dynamic minority scenarios. Thus, this elevated Jerk indicates successful retrieval of the minority distribution rather than noise. Crucially, strictly enforcing the physics loss reduces the Violation Rate to 0.11%, which is 64.5% lower than Real Data noise and 42.1% lower than SMOTE. Furthermore, the significantly reduced Coupling Error (0.05 vs. 1.45 for SMOTE) validates that PhysDiff correctly learns the second-order differential relationship between velocity and acceleration channels, significantly mitigating the kinematic decoupling often observed in linear interpolation. Note that while post-hoc smoothing (e.g., Kalman filters) could reduce Jerk, it cannot recover the lost coupling between velocity and acceleration channels destroyed by linear interpolation.

Visual Confirmation via Boxplot: Figure 3 provides a visual corroboration of these quantitative results. As illustrated, the Red distribution (SMOTE) is notably compressed compared to the Gray baseline (Real Data), reflecting a loss of dynamic diversity. The Green distribution (Vanilla Diffusion) extends significantly beyond the *Safety Threshold* (dashed line), confirming that unconstrained generation leads to severe kinematic violations. Crucially, the Blue distribution (PhysDiff) demonstrates a "safe dynamic" profile: it retains the necessary variance to represent aggressive driving (comparable to Vanilla) but successfully suppresses the extreme outliers, keeping the generated samples strictly within physical limits.

D. *Manifold Visualization*

To visually inspect the quality of the generated distributions, we project the high-dimensional feature space into 2D using PCA, as illustrated in Fig. 2.

Fig. 2(a) reveals the structural flaw of SMOTE: the synthetic samples (red crosses) form rigid, artificial lines connecting distinct clusters. These lines cut through the empty space of the manifold, representing "phantom" states that are kinematically impossible for a vehicle to traverse. In contrast, Fig.

2(b) demonstrates the superiority of PhysDiff. The generated samples (blue dots) form natural, diffusive clouds that strictly follow the manifold of the real minority data (black dots). This "Probabilistic Manifold Expansion" ensures that the augmented data respects the underlying physics of the driving domain, avoiding the geometric violations inherent in deterministic interpolation.

E. Discussion: Synergy of Safety and Utility

Contrary to the intuition that constraints hurt utility, our results show that PhysDiff ($F1 = 0.8156$) surpasses unconstrained baselines (SMOTE: 0.8022).

We attribute this synergy to the role of kinematic constraints as a domain-specific regularizer. SMOTE and unconstrained diffusion tend to overfit the noise in the minority class, generating "out-of-distribution" samples that lie in the low-probability regions of the manifold (e.g., impossible acceleration jumps). These artifacts act as adversarial noise, confusing the classifier's decision boundary. By strictly enforcing physical laws, PhysDiff prunes these invalid regions, ensuring that the augmented data strictly adheres to the underlying manifold of real driving dynamics. This cleaner, physics-compliant supervision signal enables the classifier to learn more robust and generalizable features, thereby improving performance on the test set.

V. Conclusion

This work addresses the critical challenge of ensuring both discriminative utility and kinematic fidelity in data-driven Intelligent Transportation Systems. While traditional oversampling techniques like SMOTE effectively mitigate class imbalance, we identify that their reliance on deterministic linear interpolation introduces severe physical violations. These "linear artifacts" distort the true data manifold, creating synthetic driving trajectories that defy the laws of vehicle dynamics.

To resolve this dilemma, we proposed PhysDiff, a physics-guided conditional diffusion framework. By modeling the minority class distribution as a probabilistic manifold and enforcing a novel *Kinematic Constraint* on the generation process, PhysDiff successfully synthesizes diverse, high-fidelity samples. Our empirical results demonstrate a significant advancement: PhysDiff achieves a classification performance superior to state-of-the-art baselines (Macro F1: 0.8156), while fundamentally outperforming them in physical validity. Specifically, it reduces the rate of kinematic limit violations by 64.5% compared to the inherent sensor noise in real data, effectively eliminating unrealistic acceleration jumps.

This study advocates for a paradigm shift in safety-critical applications: transitioning from purely accuracy-driven metrics to Trustworthy AI. Our results demonstrate that this transition does not require a compromise; the incorporation of physical laws yields a simultaneous gain in F1 score (+0.0134) and physical safety (+64.5% compliance), effectively preventing the model from learning dangerous, physically impossible decision boundaries.

References

[1] G. A. M. Meiring and H. C. Myburgh, "A review of intelligent driving style analysis systems and related artificial intelligence algorithms," *Sensors*, vol. 15, no. 12, pp. 30653–30682, 2015.

[2] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, no. 1, pp. 1–54, 2019.

[3] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[4] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE International Joint Conference on Neural Networks (IJCNN)*, 2008, pp. 1322–1328.

[5] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.

[6] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 6840–6851.

[7] K. Rasul, C. Seward, I. Schuster, and R. Vollgraf, "Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting," in *Proc. International Conference on Machine Learning (ICML)*, 2021, pp. 8857–8868.

[8] Y. Tashiro, J. Song, Y. Song, and S. Ermon, "CSDI: Conditional score-based diffusion models for probabilistic time series imputation," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 34, 2021, pp. 24804–24816.

[9] J. L. Alcaraz and N. Strodthoff, "Diffusion-based time series imputation and forecasting with structured state space models," *arXiv preprint arXiv:2208.09399*, 2022.

[10] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019.

[11] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[12] H. Ismail Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: a review," *Data Mining and Knowledge Discovery*, vol. 33, no. 4, pp. 917–963, 2019.

[13] Q. Wen *et al.*, "Transformers in time series: A survey," *arXiv preprint arXiv:2202.07125*, 2022.

[14] S. Kiranyaz *et al.*, "1D convolutional neural networks and applications: A survey," *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, 2021.

[15] Y. Jiao, M. Miao, Z. Yin, and B. Tao, "A hierarchical hybrid learning framework for multi-agent trajectory prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 8, pp. 10344–10354, 2024.

[16] H. Liao *et al.*, "Chain-of-thought guided multimodal large language models for scene-aware accident anticipation in autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, no. 11, pp. 19371–19380, 2025.

[17] M. Steinstraeter, J. Buberger, and D. Trifonov, "Battery and heating data in real driving cycles," *IEEE Dataport*, 2020. [Online]. Available: https://dx.doi.org/10.21227/6jr9-5235.

[18] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, 2019.