People's Democratic Republic of Algeria

Ministry of higher education and scientific research

University of Kasdi Merbah Ouargla

Faculty of New Technologies of Information and Communication

Department of Computer Science and Information Technology

A thesis submitted in partial fulfillment of the requirements for the Master degree in Industrial computer science

BY:

BRAITHEL CHAHD & GHOULIA MALAK

THEME:

Pedestrian Trajectory Prediction using Deep Learning and Social Force Models

---

In front of the jury members

Dr. Saadi Wafaa	President	University of Ouargla
Dr. Maroua Meissa	Examiner	University of Ouargla
Dr. Mezati Messaoud	Supervisor	University of Ouargla

Academic year: 2024/2025

DEDICATION

First and foremost, all praise and thanks be to Allah, whose infinite mercy and guidance have sustained me throughout this journey. His blessings have been my source of strength, clarity, and perseverance at every step.

To myself,
For every moment of doubt I turned into strength, for every sleepless night I spent dreaming, working, and hoping… This achievement is a tribute to the resilience, passion, and perseverance I found within.

To my dearest parents,
Your unwavering love, sacrifices, and endless prayers have carried me through every obstacle. You are the roots of my success, and no words can express the depth of my gratitude.

To my beloved sisters, Randa and Salsabil,
Thank you for your constant encouragement, for being my comfort in hard times, and for believing in me when I doubted myself.

To my dear brothers, Abdelnour and Yahya,
You have been my strength, my pride, and my protectors. Your quiet support means more to me than you'll ever know.

To Zaid and Nada,
My little rays of sunshine your laughter, innocence, and joy remind me of the beauty in life and give me hope for the future.

To my cherished friends: Malak, Sara, Ikram, Manar, Rym, Houria, Siham, and Manal,
Thank you for the love, laughter, and understanding. Your presence in my life has been a gift, and your companionship made every step of this journey lighter.

And finally, to my respected supervisor, Dr. Mezati Messaoud,
Thank you for your guidance, patience, and encouragement. Your support has been a cornerstone of my academic journey, and I am deeply grateful for the knowledge and confidence you've helped me build.

Chahd Braithel

DEDICATION

With pride and gratitude, I dedicate this humble work…

To myself for enduring, persevering, and pushing forward despite all the hardships and challenges. This is the first step toward a dream long pursued.

To the one whose prayers were the secret of my blessings and success, to the source of love and comfort… my beloved mother.

To the one who taught me that ambition knows no limits, and who stood by me in every moment… my dear father.

To my dear brothers: Mohammed, Lakhdar, and Noureddine, who have always been my support and strength you hold a special place in my heart. To all the members of my cherished family, who never hesitated to encourage me and offer heartfelt prayers your love lit the path during my most difficult times.

To my dearest friends: Chahd, Sara, Ikram, Siham, Houria, Rym, Manar, and Manel, my companions and sisters in spirit thank you for your honesty, encouragement, and the laughter that made the burdens lighter.

To my esteemed supervisor, Dr. Mezati Messaoud, who guided me with patience and dedication I offer you my sincerest thanks and appreciation.

And to everyone who supported me with a word, a prayer, or even silent encouragement

I offer this achievement to you, for it is the fruit of shared effort and a new chapter in a story of hope that you helped me begin.

All praise and thanks be to Allah, first and always, for His grace and guidance that lit my way and strengthened my resolve.

Malak Ghoulia

الملخص

تُعدّ تنبؤات مسارات المشاة ضرورية لأنظمة النقل الذكية والأنظمة المستقلة المُدركة للبشر، إذ تضمن السلامة وتدعم التنقل الواعي بالسياق. وبينما تستطيع نماذج التعلم العميق القائمة على LSTM والمزودة بآليات الانتباه التقاط التبعيات الزمنية بشكل فعال، فإنها غالبًا ما تفتقر إلى القدرة على محاكاة ديناميكيات حركة المشاة والسلوك الجماعي والتفاعلات الاجتماعية في البيئات الحضرية المزدحمة. من جهة أخرى، تقدم نماذج القوى الاجتماعية (SFM) إطارًا مفسرًا مستوحى من الفيزياء قادرًا على نمذجة التفاعلات الاجتماعية والسلوك الجماعي للمشاة، لكنها محدودة في تعلم الأنماط الزمنية المعقدة من البيانات. يقدم هذا العمل إطارًا هجينًا يدمج السمات السلوكية المدركة اجتماعيًا المستخرجة عبر نموذج القوى الاجتماعية (SFM)مع نماذج التعلم العميق مثل LSTM و BiLSTMو GRU تجمع كل نسخة هجينة بين SFM وبنية متكررة (LSTM+SFM، BiLSTM+SFM، GRU+SFM)لالتقاط التفاعلات الاجتماعية وأنماط الحركة المتتابعة. باستخدام مجموعة بياناتJAAD ، تم اشتقاق سمات مثل قوى التفاعل الاجتماعي، والتحولات الزمنية، والخصائص المرتبطة بالحركة مثل السرعة والتسارع لدعم التنبؤ بالسياق. أظهرت التقييمات على بيانات نظيفة أن النماذج الهجينة المقترحة تتفوق على النماذج التقليدية، مقدمة تنبؤات دقيقة وغنية دلاليًا من خلال الجمع بين تفسيرية SFM وقدرات التعلم التسلسلي للشبكات العصبية المتكررة. وبرز نموذج BiLSTM+SFM كالأكثر فعالية، محققًا أدنى معدلات خطأ (ADE = 17.04 CFMSE=450.74,),MSE=213.3, CMSE=280.87=FDE=24.65,) متفوقًا بشكل ملحوظ على نموذج BiLSTM وحده(ADE = 31.81 وCMSE=803.87,CFMSE =1195.63,(FDE = 89.8 , MSE=576,82,). تسلط النتائج الضوء على أن النموذج الهجين يتفوق على النهج أحادية الطريقة) باستخدام إما BiLSTM أو SFM وحدها(، لا سيما في السيناريوهات المعقدة متعددة العوامل حيث يتطلب التنبؤ الدقيق نمذجة تأثير الحشد وأنماط التفاعل. يوضح هذا قدرة النموذج على تمثيل الديناميكيات الزمنية الدقيقة الحبيبات والاستجابات الاجتماعية للمشاة، مما يجعله مناسبا بشكل خاص للتطبيقات في الوقت الفعلي مثل أنظمة القيادة الذاتية والمراقبة الذكية وتخطيط المسار الروبوتي.

الكلمات المفتاحية:

تنبؤ مسار المشاة، الأنظمة المستقلة المدركة للبشر، نموذج القوى الاجتماعية(SFM) ، LSTM، BiLSTM، GRU، مجموعة بياناتJAAD,

.

Abstract

Pedestrian trajectory prediction is vital for intelligent transportation and human-aware autonomous systems, as it ensures safety and supports context-aware navigation. While LSTM-based deep learning models with attention mechanisms effectively capture temporal dependencies, they often lack the ability to simulate pedestrian motion dynamics, group behavior, and social interactions in crowded urban environments. On the other hand, Social Force Models (SFMs) provide an interpretable, physics-inspired framework capable of modeling social interactions and collective pedestrian behavior but are limited in learning complex temporal patterns from data. This work presents a hybrid framework that integrates socially-aware behavioral features extracted via a Social Force Model (SFM) with deep learning models such as LSTM, BiLSTM, and GRU. Each hybrid variant combines SFM with a recurrent architecture (LSTM+SFM, BiLSTM+SFM, GRU+SFM) to capture both social interactions and sequential motion patterns. Using the JAAD dataset, we derived features such as social interaction forces, temporal transitions, and motion-related attributes like velocity and acceleration to support context-aware trajectory prediction. Evaluation on clean data showed that the proposed hybrid models outperform traditional baselines, offering accurate and semantically rich predictions by combining the interpretability and social reasoning of SFM with the sequential learning capabilities of recurrent neural networks. Notably, the BiLSTM+SFM hybrid model demonstrated the highest effectiveness, achieving the lowest error metrics with (ADE=17.04, FDE=24.65, MSE=213.3, in 0.5s, CFMSE=450.74, CMSE=280.87), significantly outperforming the standalone BiLSTM model, which yielded (ADE =31.81, FDE = 89.89, MSE=576.82, in 0.5s, CFMSE=1195.63, CMSE=803.87). The results highlight that the hybrid model outperforms single-method approaches (using either BiLSTM or SFM alone), particularly in complex multi-agent scenarios where accurate prediction requires modeling crowd influence and interaction patterns. This demonstrates the model's ability to represent fine-grained temporal dynamics and social responses of pedestrians, making it especially suitable for real-time applications such as autonomous driving systems, intelligent surveillance, and robotic path planning.

Keywords

Pedestrian trajectory prediction, Human-aware autonomous systems, Social Force Model (SFM), LSTM, BiLSTM, GRU, JAAD dataset.

Résumé

La prédiction des trajectoires piétonnes est essentielle pour les systèmes de transport intelligents et les systèmes autonomes sensibles au contexte humain, car elle assure la sécurité et favorise une navigation contextuelle. Alors que les modèles d'apprentissage profond basés sur LSTM avec mécanismes d'attention capturent efficacement les dépendances temporelles, ils peinent souvent à simuler les dynamiques de mouvement piéton, les comportements de groupe et les interactions sociales dans des environnements urbains denses. En revanche, les modèles de forces sociales (SFM) offrent un cadre interprétable inspiré de la physique capable de modéliser les interactions sociales et le comportement collectif des piétons, mais sont limités dans l'apprentissage de schémas temporels complexes à partir des données. Ce travail propose un cadre hybride intégrant des caractéristiques comportementales sensibles au contexte social extraites via un modèle de forces sociales (SFM) avec des modèles d'apprentissage profond tels que LSTM, BiLSTM et GRU. Chaque variante hybride (LSTM+SFM, BiLSTM+SFM, GRU+SFM) permet de capturer à la fois les interactions sociales et les schémas de mouvement séquentiels. En utilisant le jeu de données JAAD, nous avons dérivé des caractéristiques telles que les forces d'interaction sociale, les transitions temporelles et des attributs liés au mouvement comme la vitesse et l'accélération pour renforcer la prédiction contextuelle des trajectoires. L'évaluation sur des données nettoyées a montré que les modèles hybrides proposés surpassent les modèles de base traditionnels, fournissant des prédictions précises et riches en sémantique, en combinant l'interprétabilité et la modélisation sociale du SFM avec les capacités d'apprentissage séquentiel des réseaux neuronaux récurrents. Le modèle hybride BiLSTM+SFM s'est avéré le plus performant, atteignant les meilleures métriques d'erreur (ADE = 17.04, FDE = 24.65, MSE =213.3 en 0,5s, CFMSE=450 .74 et CMSE=280.87), surpassant significativement le modèle BiLSTM seul (ADE = 31.81 et FDE = 89.89, MSE=576.82 en 0.5s, CFMSE=1195.63 et CMSE=803 .87). Les résultats mettent en évidence que le modèle hybride surpasse les approches unidimensionnelles (utilisant uniquement soit le BiLSTM, soit le SFM), en particulier dans les scénarios complexes et multi-agents, où une prédiction précise nécessite la modélisation de l'influence de la foule et des schémas d'interaction. Cela démontre la capacité du modèle à représenter des dynamiques temporelles fines ainsi que les réactions sociales des piétons, ce qui le rend particulièrement adapté aux applications en temps réel telles que les systèmes de conduite autonome, la surveillance intelligente et la planification de trajectoire pour les robots.

Mots-clés :
Prédiction de trajectoire piétonne, systèmes autonomes sensibles au contexte humain, modèle de forces sociales (SFM), LSTM, BiLSTM, GRU, jeu de données JAAD.

Table of Contents

List of figures

List of tables

List of Equations:

CHAPTER 1: General Introduction

CHAPTER 1: GENERAL INTRODUCTION

1.1 Background and Motivation:

Urban population expansion and the extensive use of autonomous systems have made pedestrian path prediction an area of critical research. Accurate prediction of pedestrian movement is critical for a wide range of applications, from urban mobility and autonomous driving to crowd analytics. In urban environments, understanding and predicting pedestrian behavior is crucial to the safety and effectiveness of autonomous vehicles, robotic systems, and smart infrastructure. For instance, autonomous cars must predict pedestrian paths in order to avoid accidents, and crowd management systems utilize trajectory forecasts to optimize pedestrian flow in public spaces.

Deep Learning (DL) approaches has been demonstrated as a powerful tool for trajectory prediction due to its capability to learn intricate spatial-temporal dependencies from large volumes of data. DL algorithms, such as recurrent neural networks (RNNs), Gated Recurrent Unit (GRU), long short-term memory networks (LSTMs), Bidirectional Long Short-Term Memory (BiLSTM), have experienced unparalleled effectiveness in modeling sequential patterns from pedestrian trajectories. These methods take enormous amounts of history data to predict future trajectories accurately. However, purely data-driven approaches have a hard time generalizing to new situations, particularly in highly interactive and dynamic environments where social interactions are paramount.

Social Force Models (SFMs) take a different perspective by modeling pedestrian interactions in terms of forces such as repulsion, attraction, and group cohesion. SFMs capture the social aspects of pedestrian dynamics very well and offer interpretable and intuitive explanations of how individuals navigate around dense spaces. But SFMs are rigid in their capacity to adapt to different and complex actual circumstances because they are pre-programmed according to rules and parameters that cannot potentially encapsulate all possible types of human activity.

The union of Social Force Models and Deep Learning presents a feasible solution for addressing the shortcomings of both approaches. By combining the data-driven strength of DL with the social compliance and explainability of SFMs, scientists can develop hybrid models that benefit from the virtues of each approach. This combination provides the ability to create models that are not only accurate but also robust and understandable.

1.2 Objectives of the Study

The overall objectives of this research are to design an innovative pedestrian trajectory prediction model that addresses key challenges in dynamic environments. Specifically, the study aims to:

- Develop a hybrid model: Combine Deep Learning and Social Force Models (SFM) to build a robust and scalable pedestrian trajectory prediction model. The proposed hybrid approach integrates socially-aware behavioral features from SFM with sequential learning capabilities of deep learning algorithms such as LSTM, BiLSTM and GRU.
- Improve prediction accuracy: Enhance the ability of the model to predict accurate future trajectories with lower displacement errors, particularly in interactive conditions.
- Verify performance: Evaluate the model on benchmark data and compare its performance against the state-of-the-art.

1.3 Contributions

This research makes several key contributions to the field of pedestrian trajectory prediction by addressing challenges. Key developments include:

- DL-SFM hybrid model: New formulation which integrates deep learning and social force models to improve pedestrian trajectory forecasting.
- Comparative analysis: Detailed comparison of the new model with existing state-of-the-art models, demonstrating its greater accuracy and robustness.
- Real-world validation: Large-scale evaluation on benchmark datasets, like JAAD, to validate the model's efficacy in diverse environments.
- Insights into interpretability: An understanding of the model's ability to encode social interactions and make interpretable predictions, providing meaningful findings for future research.

1.4 Outline of the Thesis

This thesis is structured as follows:

Chapter 2: Literature Review

This chapter provides a comprehensive background on pedestrian trajectory prediction, key behavioral elements and discussing critical applications.

Chapter 3: Related Works

This chapter reviews existing models for pedestrian prediction, categorizing them into physics-based, pattern-based, goal-based, and hybrid models.

Chapter 4: Problem Formulation

This chapter defines the research problem, outlines the research questions, and presents the methodology for evaluating the hybrid model on both ideal and real-world data, including its deployment in real-time settings.

Chapter 5: Analysis of Joint Attention in Autonomous Driving (JAAD)

This chapter analyzes the JAAD dataset, detailing pedestrian behavioral characteristics, annotation structures, and how it compares to traditional datasets.

Chapter 6: Prediction on Clean Data

This chapter explores the use of Social Force Models (SFM) to extract behavioral features, integrates them into a hybrid LSTM, BiLSTM, GRU framework, and evaluates its performance using standard prediction metrics.

Chapter 7: General Conclusion

This chapter provides a general summary of the thesis and the obtained results, and it outlines potential future directions.

.

CHAPTER 2: Literature Review

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

With the rapid advancement of artificial intelligence technologies and intelligent systems, the ability to accurately predict pedestrian trajectories has become a critical requirement in various applications such as autonomous vehicles, collaborative robots, and smart surveillance systems. Pedestrian trajectory prediction represents an advanced field within machine learning and human behavior analysis, where sequences of spatial coordinates of pedestrians over time are utilized to anticipate their future movements in dynamic, multi-factor environments. This chapter focuses on the theoretical and practical foundations of pedestrian trajectory prediction techniques, starting with defining the problem as a time-series forecasting task, and progressing through the analysis of key behavioral elements that influence pedestrian behavior such as intention, emotion, social interaction, and environmental context. The chapter also highlights the most prominent real-world applications that benefit from these techniques, emphasizing their role in enhancing safety and adaptive response in complex urban environments.

2.2 Pedestrian Trajectory Prediction

Pedestrian trajectory prediction is a field in artificial intelligence that aims to anticipate the future paths of individuals based on their past movements and the surrounding environment[1]. This is cast as a formally defined time-series forecasting problem in which an exemplar learning model is given a stream of history of pedestrian location information ($X_t = \{x_1, x_2, ..., x_{to6s}\}$[2], each x_i is a 2D/3D[3] location coordinates at time i) to predict future locations ($Y_t = \{x_{to6s+1}, ..., x_{tpre}\}$)[2] for some time horizon. The figure 1illustrates a pipeline for predicting pedestrian crossing behavior. It begins with capturing an observation sequence through vehicle-mounted cameras. From these observations, critical features such as vehicle speed, pedestrian bounding boxes. These features are then fed into advanced prediction models like 3DCNNs, hybrid 2DCNN+RNN models, or GCN+RNN networks. The system finally outputs a binary prediction determining whether the pedestrian intends to cross the road or not [4].

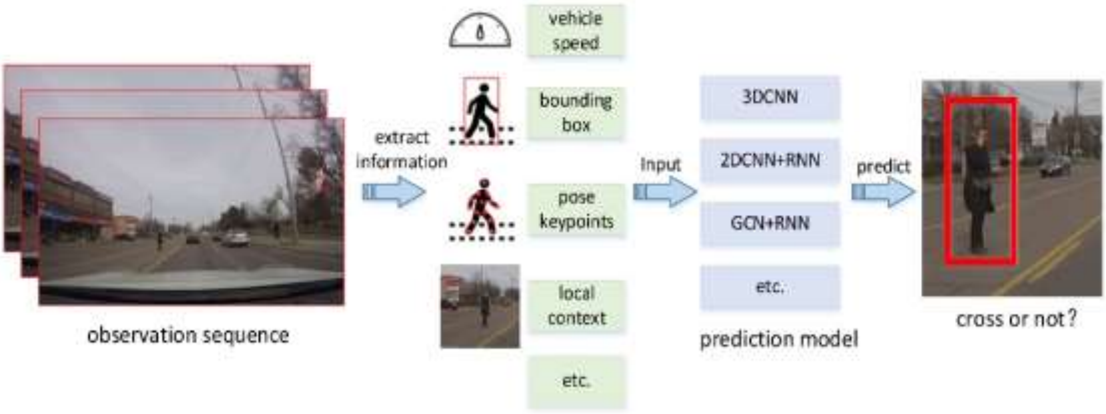

Figure 1:Overview of pedestrian crossing behavior prediction methods [4]

2.3 Key elements of pedestrian behavior

The key elements of pedestrian behavior, as illustrated in the image. First, Trajectory refers to the pedestrian's movement path, which is influenced by their intentions and decision-making[5]. Second, Intention involves planned actions inferred from visual cues and motion patterns[6]. Third, Emotion captures the psychological state that affects how pedestrians respond to their surroundings and make decisions. Fourth, Social Interactions highlight the influence of other people on a pedestrian’s timing and behavior[7]. Fifth, Environmental Context refers to the surrounding conditions that shape perception[5], risk assessment, and behavioral choices. Together, these elements form a comprehensive framework for analyzing and predicting pedestrian behavior in dynamic and complex urban settings[8]. As Figure 2 shows.

Figure 2:Pedestrian Behavior Components

2.4 Critical Applications

Accurate pedestrian trajectory prediction is vital for technologies like autonomous vehicles[9], robots[10], and surveillance systems[11], especially in smart cities[12]. Some applications:

2.4.1Autonomous Vehicles

Autonomous vehicles need advanced models to predict pedestrian trajectory in real time to ensure safety in urban environments. These models use features like body pose, speed, trajectory, and context (e.g., traffic lights) to estimate crossing behavior[7]. As shown in Figure3.

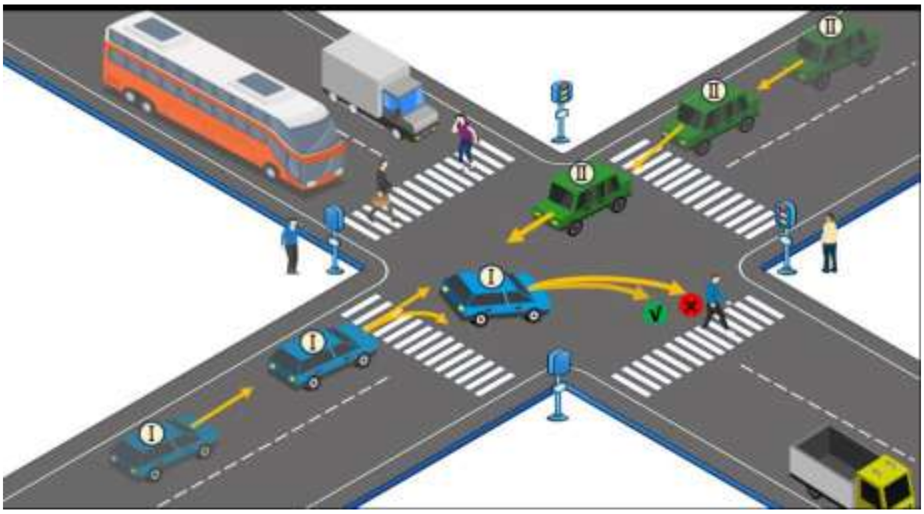

Figure 3:Intention-based and Risk-Aware Trajectory Prediction for Autonomous Driving in Complex Traffic Scenarios[13]

2.4.2 Robotics

Pedestrian behavior prediction is vital for effective pedestrian-robot interaction in shared spaces like hospitals and airports. Robots must detect subtle cues such as body orientation and gaze to navigate cooperatively and safely[14]. As shown in Figure 4.

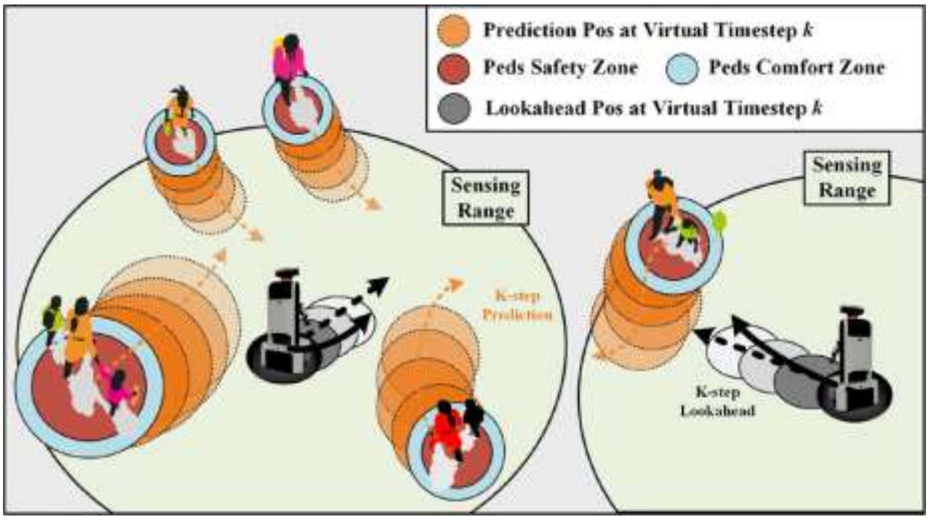

Figure 4:Robot-pedestrian interaction with the K-step proactive reward motivates robots to enhance their awareness of crowd comfort and safety.[15]

2.4.3 Smart Surveillance

Forecasts regarding individuals behavior feed into intelligent surveillance systems to deter crime, and enable crowd. With this ability to predict such behavior prior to occurrence, early intervention can be adopted by the authorities or through automation[11]. As shown in Figure 5.

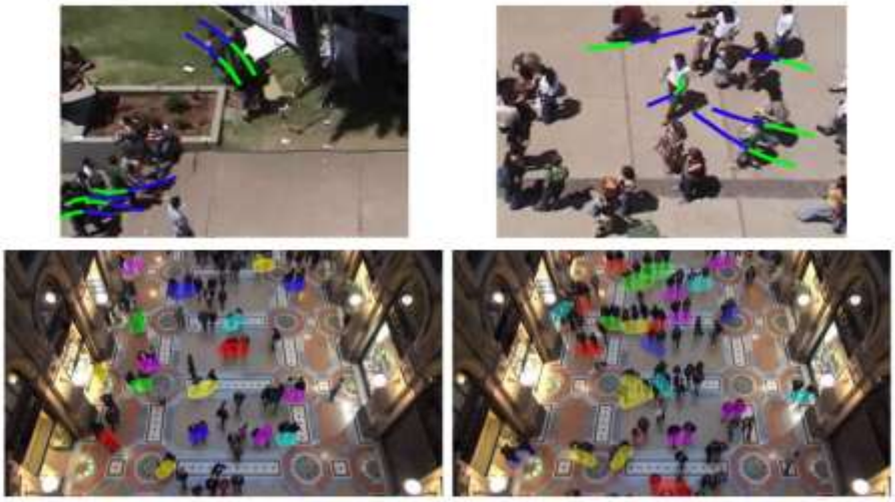

Figure 5:Predicting pedestrian behavior with smart surveillance systems[16]

2. 5 Conclusion

Pedestrian trajectory prediction remains a central and evolving field, bringing together the complexities of human behavior, social interactions, and environmental constraints. This chapter reviewed the theoretical and practical foundations of the field and highlighted the main motivations driving the growing interest in this topic. Despite the substantial progress achieved, fundamental challenges persist, such as model interpretability and generalizability across diverse environments. These challenges underscore the need for hybrid solutions that combine the strengths of deep learning with knowledge-based reasoning.

This chapter sets the stage for the following one, which introduces a novel hybrid prediction framework aimed at overcoming current limitations by enhancing accuracy and adaptability in real-world pedestrian interaction systems.

Chapter 3: Related Works

3.1 Introduction

In this chapter, we introduce previous research on the topic of pedestrian trajectory prediction. The first section categorizes existing models into the three main paradigms used in this field. Subsequently, we review the most relevant datasets that have been used to support pedestrian trajectory analysis and prediction.

3.2 Pedestrian Trajectory Prediction Models

Trajectory prediction seeks to forecast the future positions of pedestrians based on their prior movements, often formulated as a time series forecasting problem. With the rapid advancements in deep learning[17], models like RNNs[18], CNNs[19], GNNs[20], and Transformers[21] have emerged as powerful tools in this area, surpassing traditional statistical techniques such as Kalman Filters[22] and Hidden Markov Models[23], which struggle to handle complex behaviors in dynamic environments.

In contrast to knowledge-based models, which depend on predefined rules and manually engineered features, deep learning methods learn trajectory patterns directly from data, offering enhanced adaptability and scalability[8]. This shift has established deep learning[8] as the dominant approach in practical applications, ranging from autonomous vehicles[24] to intelligent surveillance systems[11].

Recent research has proposed various model architectures and design objectives, emphasizing key factors such as social interactions[24], motion intent[25], and environmental context[26]. These methods underscore the increasing demand for models that not only deliver high predictive accuracy but also maintain robustness and real-time performance, particularly in safety-critical scenarios. supported by Figure 6, which illustrates the three main model categories.

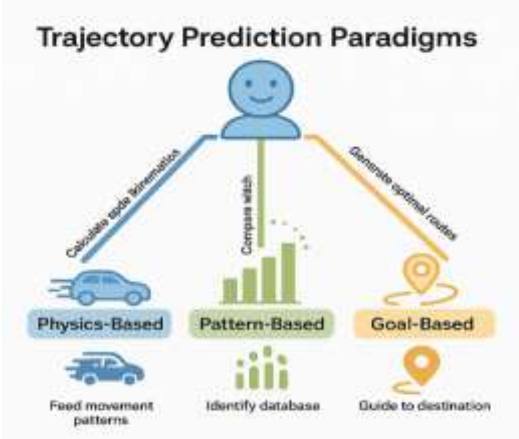

Figure 6:Trajectory Prediction Paradigms

3.2.1 Physics-Based Models

Physics-based models use deterministic rules or equations derived from physics or behavioral theories to model pedestrian trajectories. These models follow the Sense–Predict framework[25]. A simple example is the Constant Velocity (CV) model, which assumes pedestrians maintain their speed and direction over short horizons. Schöller et al[27]. highlighted the competitive performance of CV in benchmark datasets, recommending its use as a baseline in prediction tasks.

Extensions include the Constant Acceleration (CA) and Coordinated Turn (CT) models, which incorporate acceleration or directional change. Moreover, environmental constraints (such as sidewalks, obstacles, or intersections) can be integrated into physics-based models to enhance prediction realism[28].

A notable knowledge-based approach is the Social Force Model (SFM)[29], proposed by Helbing and Molnár, where pedestrians are modeled as particles influenced by attractive and repulsive social forces. This model captures group interactions and personal space constraints, making it effective for predicting trajectories in dense environments.

Despite their interpretability and low data requirements, these models may oversimplify human decision-making and struggle with complex or dynamic behaviors[30].

3.2.2 Pattern-Based Models

Pattern-based or data-driven models rely on learning from historical trajectory data, often using machine learning or deep learning techniques. These follow the Sense–Learn–

Predict paradigm, allowing models to generalize from observed patterns without requiring explicit rules[31].

Recurrent Neural Networks (RNNs)[32], particularly Long Short-Term Memory (LSTM) networks[18], have been widely used due to their ability to model temporal dependencies. Alahi et al. introduced the Social-LSTM model [33], which predicts pedestrian trajectories while accounting for nearby individuals' influence. Subsequent improvements integrated contextual cues, such as scene semantics or object interactions.

Poibrenski et al. combined Conditional Variational Autoencoders (CVAE) with RNNs to model[18] the multi-modal nature of pedestrian future positions. Other works, such as Yao et al. and Malla et al., incorporated pedestrian intention and behavior labels (e.g., crossing, waiting) to improve short-term prediction[34].

Although these models often achieve high accuracy, they demand extensive labeled datasets and computational power. Moreover, their interpretability remains limited, and they may suffer from overfitting in unseen scenarios[34].

3.2.3 Goal-Based Models

Goal-based models adopt the Sense Reason Act paradigm, aiming to infer pedestrians' long-term goals and use planning strategies to forecast their trajectories[25]. These models treat prediction as a decision-making task, often utilizing planning frameworks like Markov Decision Processes (MDP) or Inverse Reinforcement Learning (IRL)[35].

Such methods perform well when pedestrian goals can be inferred and a map of the environment is available. They outperform purely reactive models in structured or goal-directed scenarios, offering more plausible long-term forecasts[36]. However, their computational complexity can increase dramatically with the number of agents or environmental complexity, making them less practical in crowded or real-time applications[36].

3.2.4 Classification Criteria

To systematically classify deep learning (DL) models for pedestrian trajectory prediction, we define the following evaluation criteria based on the reviewed literature:

•Type of Algorithms: Identifies whether the approach uses deep learning models such as LSTM for temporal dependencies, CNN for spatial feature extraction, GAN for generating realistic trajectories, or other approaches like GNNs for interaction modeling and RL-based models for adaptive decision-making.

•Network Structures: Identifies whether the deep learning model follows a sequential structure (e.g., LSTM, RNN) for time-series data or a non-sequential structure (e.g., CNN, GAN, Transformers) for parallel and non-linear data processing.

•Prediction Tasks: Categorizes tasks such as interaction modeling, trajectory planning, and intention prediction.

•Use of Advanced AI Techniques: Identifies whether deep learning models integrate Transformer Networks, Variational Autoencoder (VAE), or other techniques like Reinforcement Learning (RL) and Inverse Reinforcement Learning (IRL).

Articl	Type of algorithms				Network structures		predection tasks			the use of advanced AI		
	LSTM	CNN	GAN	Other	séquential	Non-séquential	interaction	trajectory planning	intention	Transformer Networks	Variational Autoencoder (VAE)	Other
Hongjia Zhang-2020-[37]	x				x		PIS	EMP	APC			Attention Mechanism
Amir Rasouli-2020-[38]	x	x			x	x	PIS	IPV	CFT			Attention Mechanism
Jie Yang -2020-[39]		x				x	APC	EMF	ECM			Multi-Scale Detection

Joseph Gesnouin-2020-[40]		x		Auto-encoder		x	PIS	EMP	APC			Autoencoder-Based Feature Learning
J. Lorenzo-2020-[41]	x	x			x		APB	PP	EPU			Autoencoder-Based Feature Learning
Jun Hayakawa, Behzad Dariush-2020-[42]		x		TRN			CFT	PP	EMP			Temporal Relation Network

Dan Xiong -2021-[43]	x	x	x		x		AHI	PMP	**PAC**	x		
Arash Kalatian-2021-[44]	x				x		APB	IPV	IAV			
J. Lorenzo-2021-[45]		x		Transformer	x	x	APB	PMP	**EMP**	x		
Chi Zhang -2021-[46]		x		TCN ,GCN		x	APB	EPU	**APC**			GNN
Raphael Rozen	x			Bi-RNN	x		EMSI	PP	EMP			Asymmetrical Bi-RNN

berg-2021-[47]												
Lina Achaji-2021-[48]				Transformer Networks		x	APB	PP	EPU	x		
Joseph Gesnouin-2021-[49]		x		GRU, Atrous Convolutions	x	x	SFT	PMP	EPU			Temporal, Self-Attention
Chi Zhang-2022-[50]	x	x	x		x	x	PMP	EMP	EMSI	x		

Riddhi Joshi -2022-[51]	x	x			x	x	EMF	PP	EMF		x	
Eric Liang -2022-[19]		x		yolo		x	CFT	PP	APB	x		
Amir Rasouli-2022-[52]	x	x		SAIM	x	x	PIS	PP	IAV	x		
Dong xu Guo -2022-[53]	x	x			x	x	PIS	PMP	IAV	x		
Jibran Ali		x			x		PP	EPU	APB			X(Attention Mechanism)

Abbasi,2023 [54]													
Zhengming Zhang,2023[55]				Transformer		x	EMSI	PP	EBT	x			
Esteban Moreno-2023[6]				FNN	x		PP	PP	IPV				
Jia Huang-2023-[26]	x			Transformer	x		PP	PMP	IAV	x			
Ting Fu,2023[56]		x		DeepSORT		x	PMP	PP	APB				

Chen 2024{ 42]				GCN	x		PIS	EMP				Attention Mechanism
Li 2024[43]				GCN	x		PIS					Heterogeneous Graph Learning
Liang 2024[44]				TCN	x			EMP				Visual Information Integration
Ling 2024[45]				GCN	x				IAV			Spatial–Temporal Attention Mechanism

Table 1:Comparative Analysis of Pedestrian Trajectory Prediction Models: Structures, Applications, and Interpretability

(PIS): Predicts pedestrian interactions with surroundings; **(EMP): Estimates** movement paths to enhance safety and navigation; **(APC):** Analyzes pedestrian movement in crowded environments; **(IPV):** Integrating pedestrian intention with vehicle dynamics; **(CFT):** Considers factors like traffic signals, road boundaries, and pedestrian density; **(ECM):** Examines crowd flow in metro stations; **(EMF):**Extracting multi-scale features using convolutional neural networks; **(APB):** analyze pedestrian behavior and interactions; **(PP):** Predicts pedestrian; **(EPU):** Examines pedestrian behavior in urban environments; **(AHI):** Analyzes human-human and human-space interactions; **(PMP):** predict movement paths in dynamic scenes; **(PAC):** prediction accuracy in crowded areas; **(IAV):** improve AV sensing systems; **(EMSI):** Encodes pedestrian motion with social interactions.

3.3 Hybrid Models

To leverage the strengths of multiple paradigms, hybrid models combine physics-based and data-driven or goal-driven components[1]. Korbmacher and Tordeux argue that while data-driven models outperform traditional methods in accuracy, physics-based components still contribute to model interpretability and robustness.

Silvestri et al. demonstrated that integrating domain knowledge into neural networks can enhance learning with limited data. Antonucci et al[57]. proposed embedding the Social Force Model into an LSTM network structure[43], producing physically-consistent predictions. Similarly, Kothari et al[58]. used discrete choice models to output interpretable trajectory predictions aligned with human intention[58].

Hybrid approaches represent a promising direction for future research, offering a balanced trade-off between accuracy, generalizability[59], and interpretability[58].

3.4 Datasets

To train and evaluate pedestrian trajectory prediction models[60], a variety of datasets have been collected, capturing pedestrian trajectories and behavioral attributes in different real-world scenarios. Based on the nature and purpose of the collected data, these datasets are commonly categorized into three main types:

3.4.1 Pedestrian Motion Datasets

Among the most widely used datasets in the literature are the ETH BIWI Walking Pedestrians Dataset ETH[61] and UCY[62]. These datasets consist of video recordings and annotated pedestrian trajectories captured from a top-down perspective in public urban spaces like sidewalks and plazas, with minimal interaction between pedestrians and vehicles[24].

Despite their limited size and relatively simple settings, these datasets have become standard benchmarks for pedestrian motion prediction[7]. Even the most recent state-of-the-art models often rely on them to evaluate performance, due to their accessibility and the availability of established evaluation protocols[63]. As shown in Figure 7.

Figure 7:pedestrian prediction motion with ETH Dataset[61]

3.4.2 Pedestrian Intention Datasets

More recent datasets such as JAAD (Joint Attention in Autonomous Driving)[64] and PIE (Pedestrian Intention Estimation)[65] provide video sequences enriched with annotations related to pedestrian behaviors, attributes, and spatial positions. A distinguishing feature of these datasets is their emphasis on pedestrian intention, with labels manually provided by researchers[66].

These detailed annotations enable the development of more sophisticated models that incorporate intention estimation as an additional feature. For example, a model might first determine whether a pedestrian intends to cross the street and then use that information to predict their future trajectory[51].

While this two-stage approach shows promise in enhancing prediction accuracy, it also introduces potential challenges: incorrect intention classification may negatively influence trajectory predictions, and the added processing time can become a bottleneck particularly in real-time systems such as autonomous driving applications[24], where prediction speed is critical. Fig. 8 represents an annotated frame from PIE[65]. As visible, each pedestrian was labeled with a bounding box and the action they are performing. For example, the pedestrian on the right is labeled as looking and standing.

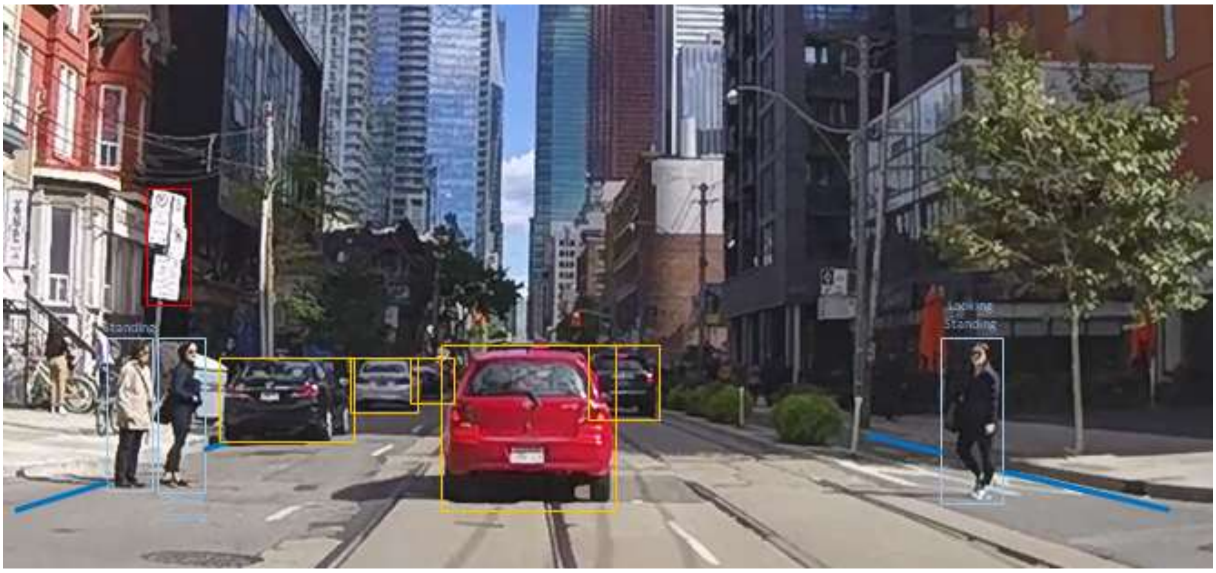

Figure 8:An annotated frame from the PIE Dataset [64]

3.4.3 Pedestrian behavior Dataset

The JAAD (Joint Attention in Autonomous Driving) dataset[64] holds a unique position among pedestrian-related datasets, as it captures both trajectory data and rich contextual annotations from an egocentric viewpoint[66]. Designed to support research in joint attention and human behavior understanding in traffic, JAAD provides high-resolution video sequences recorded from a moving vehicle in diverse urban settings, covering various weather, lighting, and traffic conditions[64].

In addition to tracking pedestrian positions over time, JAAD includes annotations such as pedestrian actions[64] (e.g., walking[67], standing[68], looking[69]), environmental context (e.g., presence of crosswalks[67], traffic signals[51]), and vehicle-pedestrian interaction cues[24]. This makes JAAD suitable not only for behavioral analysis but also for direct application in pedestrian trajectory prediction, especially in real-world autonomous driving scenarios where such context plays a vital role in forecasting pedestrian movement[66].

By combining motion trajectories with scene semantics and behavioral labels, JAAD enables hybrid models that leverage both data-driven learning and knowledge-based reasoning[64]. This integration enhances the robustness of trajectory predictions in complex environments, making JAAD an essential benchmark for developing and evaluating context-aware prediction systems[44].

3.5 Conclusions

Pedestrian trajectory prediction models rely heavily on the quality and diversity of datasets. While traditional datasets like ETH and UCY offer simple motion data, newer datasets such as PIE and JAAD provide richer behavioral and contextual information. In particular, JAAD combines trajectory data with real-world interaction cues, making it highly suitable for training context-aware models in autonomous driving. Choosing the right dataset is essential for achieving accurate.

Building on the challenges identified in the previous chapters, this chapter outlines key research questions related to pedestrian trajectory prediction in real-world scenarios and proposes a comprehensive methodology to address them. It also sets the stage for developing a hybrid modeling.

Chapter 4 :Problem Formulation

CHAPTER 4: PROBLEM FORMULATION

4.1 Introduction

Based on the analysis of the current state of the art in pedestrian trajectory prediction, this chapter discusses the identified challenges in the field, presents the research questions, and finally proposes a methodology to address them.

4.1 Problem Statement

As discussed in the previous chapters, pedestrian trajectory prediction models often rely on datasets collected in urban areas from static, elevated cameras. While these datasets provide valuable insights into pedestrian motion patterns in controlled environments, they fail to capture the complexity of interactions between pedestrians and vehicles an essential aspect in real-world applications such as autonomous driving.

Moreover, different modeling approaches namely physics-based models such as the Social Force Model (SFM)[70], and data-driven models like Long Short-Term Memory networks (LSTM) offer distinct advantages and limitations[43]. While SFM provides interpretable and socially-aware movement modeling, it struggles in capturing non-standard or unexpected pedestrian behaviors. On the other hand, deep learning models[71] achieve high accuracy by learning from data, yet they lack interpretability[58] and generalization in unseen scenarios[59].

Recent literature suggests that combining both approaches into hybrid models could enhance prediction accuracy and contextual understanding. However, questions remain about how such integration should be implemented, under what conditions hybrid models outperform individual approaches, and how these models behave in realistic, noisy environments[1].

Another issue lies in the evaluation of model performance. Most research evaluates models on clean, well-tracked pedestrian trajectories, rarely accounting for sensor occlusions, tracking failures, or missing data common in real-world sensor environments. Additionally, there is limited research exploring how these models can be effectively deployed on autonomous vehicles in real time[24], where challenges such as generalization to new environments[59], robustness to hardware failures, and latency in prediction outputs can significantly affect performance. As figure 9 show.

Figure 9:Dynamic change in pedestrian trajectory[72]

4.2 Research Questions

RQ1a: What are the quantitative motion properties of pedestrian trajectories in autonomous driving datasets?

RQ1b: How do these trajectories differ from those found in traditional pedestrian datasets?

RQ2a: What are the strengths and limitations of physics-based and data-driven models for pedestrian trajectory prediction?

RQ2b: What are the potential advantages of hybrid models over individual modeling approaches?

RQ3a: What are the main challenges associated with trajectory prediction using realistic, noisy data?

RQ3b: To what extent does model performance degrade when using imperfect data as opposed to ideal input?

4.3 Methodology

In this section it is discussed how each of the RQs is answered with an experimental analysis.

4.3.1 RQ1 – Analysis of Autonomous Driving vs. Traditional Datasets

To address RQ1, datasets such as JAAD[64], Waymo, and Argoverse will be utilized as examples of real-world, on-vehicle sensor data capturing complex pedestrian-vehicle interactions. These will be compared with traditional benchmark datasets like ETH[61] and

UCY[62], which primarily involve pedestrian-only scenes. Statistical motion analysis tools such as OpenTraj[73] will be used to extract and compare motion properties like velocity[27], acceleration[29], interaction range, and trajectory curvature.

4.3.2 RQ2 – Model Evaluation on Ideal Data

For RQ2, both a physics-based model (SFM)[29] and a data-driven model (LSTM) [18]will be implemented and tested on clean, unaltered trajectories. SFM is selected for its interpretability and relevance to social interaction modeling, while LSTM is chosen for its strength in learning temporal dependencies. The models will be evaluated using standard metrics including Average Displacement Error (ADE)[74] and Final Displacement Error (FDE)[74] across multiple prediction horizons. Should either approach prove insufficient in accuracy, a hybrid model will be designed to integrate the strengths of both.

4.3.3 RQ3 – Model Evaluation on Realistic Data

To simulate realistic conditions, selected trajectories will be altered to reflect sensor noise, occlusions, and incomplete tracking mimicking real-world deployment scenarios. The same evaluation metrics (ADE, FDE) will be used to assess how model performance is affected[74]. This experiment will help identify the robustness of each model type, and determine if hybrid models provide better resilience under non-ideal conditions.

4.4 Conclusion

This chapter addressed the main challenges currently facing the field of pedestrian trajectory prediction and introduced a set of research questions aimed at bridging the gaps present in traditional modeling approaches particularly with regard to real-world data and the integration of knowledge-based and deep learning-based models. A detailed methodology was presented to answer these questions, involving a comparative analysis of different types of datasets, performance evaluation under both ideal and realistic conditions, and the design of a deployment framework for real-world implementation on autonomous vehicles.

The next chapter will focus specifically on the analysis of the JAAD dataset, which represents a realistic environment rich in pedestrian-vehicle interactions, making it a valuable reference for building and evaluating the proposed model.

CHAPTER 5: Analysis of Joint Attention in Autonomous Driving (JAAD)

5.1 Introduction

In recent years, we have witnessed phenomenal advancements in autonomous driving technologies, which must be capable of accurately perceiving what is going on around them and making instantaneous, context-aware decisions. One of the most complex challenges in this area is dealing with pedestrians' behavior, which entails some randomness and lack of responsiveness. This is where the JAAD dataset (Joint Attention in Autonomous Driving) becomes useful. It was specifically designed to study the phenomenon of joint attention between pedestrians and cars. Joint attention refers to the ability of different drivers and pedestrians to comprehend each other's signals such as eye contact or body language and coordinate safe decisions.

5.2 Definition of JAAD Dataset:

The JAAD (Joint Attention in Autonomous Driving) dataset is a novel, high-quality, annotated collection of real-world video recordings designed to facilitate the analysis of pedestrian and driver behavior in urban traffic environments[64]. Created to address the complexity of joint attention the social interaction process by which pedestrians and drivers convey information through non-verbal cues[75]. JAAD provides dense annotations for pedestrian behavior, attributes (such as age, gender, and intention to cross), environmental conditions (such as weather and traffic density[76]), and scene elements (such as traffic lights and crosswalks) [66]. Composed of 346 video clips captured in various geographical locations and under varying conditions[64]. It includes annotations of subtle actions such as looking at oncoming vehicles, waving hands, decelerating while crossing[75]. The data was collected using vehicle-mounted high-resolution dash cameras in a variety of urban environments in North America and Europe [66], capturing real-world interactions from the driver's perspective As Figure 10 shows the dataset enables critical research in pedestrian detection, behavior prediction, and perception system development in autonomous driving [66].

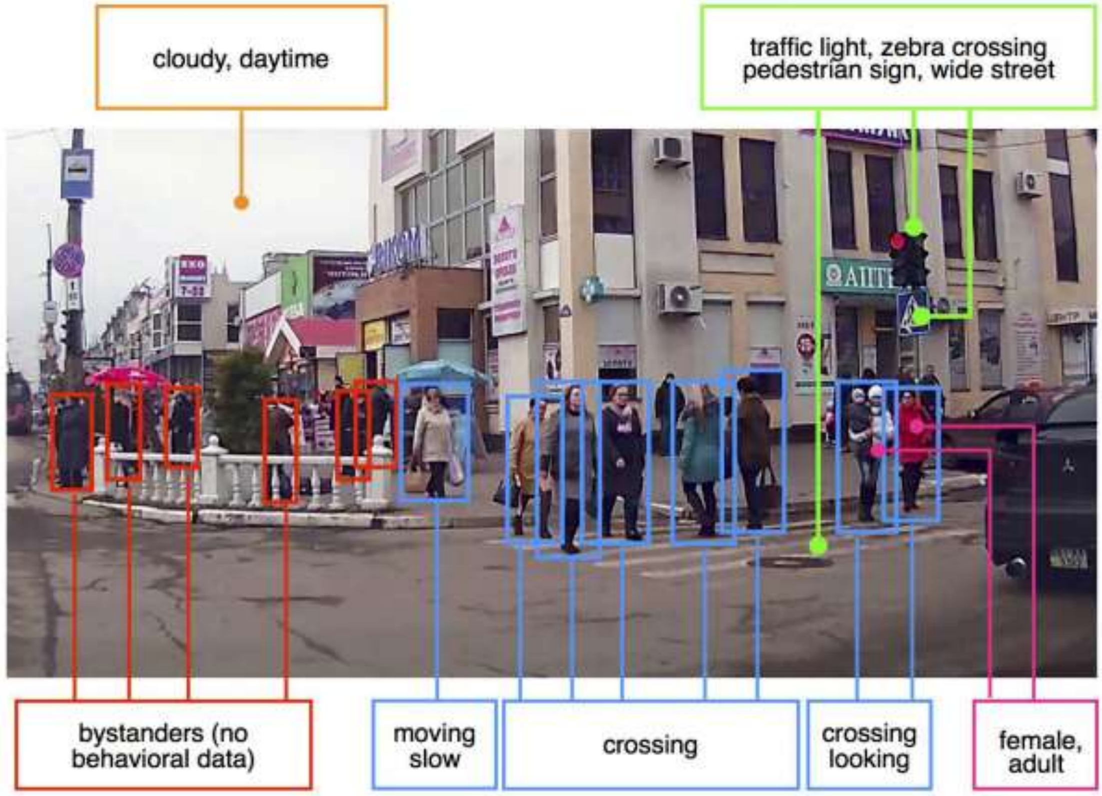

Figure 10:Example of annotations provided in the dataset: bounding boxes for all pedestrians, behavioral labels, gender and age for pedestrians crossing or intending to cross, contextual tags [74]

5.3 Pedestrian Behavioral Properties in the JAAD Dataset

Pedestrian behavior attributes are explored as presented in the JAAD dataset[66]. Various behavioral cues, including vehicle gaze, hand gestures, curb stops. The findings show that the vast majority of pedestrians perform communicative actions prior to crossing[75]. Temporary stops and glances in the direction of the approaching vehicle are the most frequent behaviors. The interactions were labeled in the JAAD dataset[66] with consistent annotation, allowing precise extraction and statistical analysis of interaction patterns for various traffic scenarios and across ages[75].

Other than analyzing communicative actions, age is also a key indicator of pedestrian behavioral characteristics. Research has proven that walking speed varies greatly across various age groups, with young individuals walking at an average speed of around 1.05 m/s, while older individuals walk slower at around 0.85 m/s[67]. These kinematic differences affect the way and the time pedestrians decide to cross[75], particularly in dynamic urban environments[67]. For example, older pedestrians tend to take longer to cross after hesitation and are less likely to change direction or speed of movement abruptly. Understanding such distinctions is necessary in order to develop pedestrian behavior models capable of capturing age-related differences. As Figure 11 show.

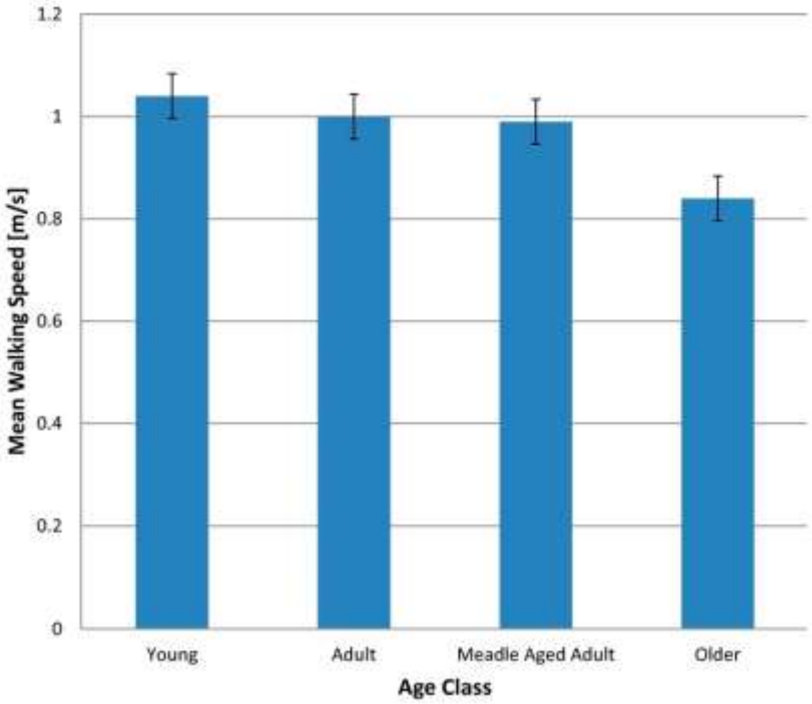

Figure 11:Effect of the age class on mean speed of individual pedestrians.[76]

5.3.1 Interaction and Attention Efficiency

In traditional data sets, path efficiency is defined as a proportion of the straight-line distance over the actual path length. In JAAD, this is made general for the calculation of interaction efficiency, i.e., the manner in which a pedestrian expresses and executes the intention to cross[75]. The majority of JAAD cases illustrate pedestrians making eye contact with the driver or providing a hand signal before crossing and hastily executing the movement. In contrast, inefficient interactions are characterized by delay or blocked crossings[75]. This measure is an indicator of the influence of social cues on movement efficiency in public spaces[66].

5.3.2 Entropy and Predictability of Intention

To quantify the variability and predictability of pedestrian behavior in JAAD, a Gaussian Mixture Model was trained on the annotated behaviors (e.g., walking, looking, stopping)[69]. The entropy of the distribution was estimated using kernel density methods.

Results show that JAAD has much higher entropy than usual datasets, due to the diverse behaviors and interaction types. Furthermore, conditional entropy analysis was conducted, measuring uncertainty[77] in future behavior within a 2.5s window of observation. High conditional entropy quantifies the presence of challenging-to-model behavior with simple motion predictors, further substantiating the need for multi-modal prediction models with visual features and context[52].

5.3.3 Behavioral Deviations and Cues

A very notable feature of JAAD is the introduction of non-bodily motion cues such as behavioral cues. Curvatures deviating from straight motion, here, are not only given geometrical connotations (rate of change over time of angle of heading), but also weighed considering intent to move[75]. People generally change direction or halt according to their eye contact visibility by the car or responsiveness to surrounding stimuli[75]. The deviation of angles calculated at important time points (e.g., 1.5s and 3.0s) reflects that these deviations are not always smooth or uniform, which suggests cognitive processing such as risk assessment and decision-making[69].

5.4 Overview of JAAD Dataset Structure

When you download the JAAD dataset, you will receive a folder structure which will include folders of videos, annotations and metadata files[64]. Here is a detailed description of each file type:

5.4.1 Video Files

The JAAD dataset is composed of a series of short video recordings from a dashboard-mounted camera in a moving vehicle, approximating the perception of an autonomous vehicle[9] navigating real urban environments[66]. The clips give rich contextual detail to the study of human-vehicle interactions, capturing subtle pedestrian behaviors such as hesitation, beginning of crossing, and direction of gaze towards encroaching traffic[54].

5.4.2 Annotations Files

These are the core of the dataset. Usually in **XML**, **CSV**, or **TXT** format depending on the version. They contain frame-level and object-level metadata for each video[64].

Typical fields in annotation files:

Field	**Description**
frame	The frame number within the video
id	Unique object identifier (e.g., a pedestrian ID)
bbox	Bounding box coordinates: [x, y, width, height]
behavior	Observed behavior: walking, crossing, standing, looking, etc.
Weather	Weather conditions (e.g., sunny, rainy)
scene_type	Scene context: urban, suburban, etc.
time_of_day	Time category: day, night, morning, etc.
location	City or country where the video was recorded

Table 2:explain the components of the annotations file

5.4.3 JAAD _clips

This directory contains raw video clips that have been used in the JAAD (Joint Attention in Autonomous Driving) dataset. Each is a short actual urban driving scenario (typically 5 to 15 seconds) captured by a camera mounted on a dashboard[64]. They are the primary source for analyzing pedestrian activity and vehicle-to-pedestrian interaction across multiple environmental contexts[66].

Video Characteristics:

- Formats: Videos are provided in MP4 format[64].
- Total clips**:** 346 MP4 videos, each lasting between 5 and 10 seconds, and totaling over 82,032 frames[64].
- Resolution: Most of the videos contain 1920×1080pixels, but a few are slightly different depending on the equipment used for recording[66].
- Frame rate**:** Typically, 30 frames per second (fps)[64].
- These clips are saved in a folder named JAAD_clips and are labeled with sequential numbers (video_0001.mp4, video_0002.mp4, etc.). Frame-by-frame annotations for these videos are provided separately as CSV or TXT, and not embedded as part of the video files[66].

5 .4.4. attributes.txt / attributes.csv

The attributes.txt and attributes.csv are core components of the JAAD dataset, consisting of static pedestrian descriptions that occur in the video shots[64]. The data is used in pedestrian movement classification and analysis across different situations and is very helpful in training models to be capable of distinguishing patterns of interaction based on a person's characteristics. These files include a variety of features such as gender, age, and so forth in numeric or text forms suitable for computation[64]. Typical fields include:

Field	**Description**
Id	Pedestrian ID
Gender	Male or Female
Age_group	Young, Adult, Meadle aged adult, Older
Group_size	Number of people walking together in the group
Motion_direction	Direction of movement (e.g., LONG for longitudinal or LAT for lateral)
Looking	Whether the pedestrian is looking toward the vehicle
Crossing	Is the pedestrian crossing the road?
Walking	Is the pedestrian walking?
traffic_direction / num_lanes	The direction of traffic flow and number of road lanes present in the scene

Table 3:Description of the content of the file attributes.txt/ attributes.csv

5.4.5 Vehicle Motion Annotations

The file `video_x_vehicle.xml` contains vehicle motion annotations at the frame level for a specific video clip in the JAAD dataset[64]. It is structured in XML format and records how the vehicle (from which the video was recorded) behaves over time[64].The file consists of a series of frame elements, each with:

- Id: the frame number.
- Action: the vehicle's motion status at that frame.

Types of Actions

Each frame is labeled with one of the following values:

- moving slow: the vehicle is moving at a low speed.
- decelerating: the vehicle is slowing down.
- stopped: the vehicle is stationary.
- accelerating: the vehicle is speeding up.

5.4.6 Pedestrian Appearance file

The video_x_appearance.xml file in dataset JAAD contains detailed annotations describing the visual appearance of pedestrians in each video frame[64]. The data is structured as a sequence of <box> elements within a <track> block, which represents the tracking of an individual pedestrian throughout the clip[75]. Each box corresponds to a single frame and records a variety of appearance-related attributes, including:

- Clothing: color of upper and lower garments (light/dark)[64].
- Accessories: such as backpacks, handbags, shoulder bags, strollers, umbrellas, caps, or sunglasses[64].
- Object interaction: such as carrying a phone or having a baby[64].
- Pose/direction: the visible body orientation (front, back, left, right) based on the pedestrian's position and movement in the scene[64].

5.4.7 traffic file content

The file contains annotated information related to a traffic scene from the JAAD dataset, providing a detailed description of each frame in the corresponding video. with each frame having an ID and several attributes indicating the presence of traffic-related elements:

- **id:** The frame number within the video.
- **ped_crossing:** Indicates whether a pedestrian crossing is visible in the frame (`0` = not present).
- **ped_sign:** Specifies whether there is a pedestrian traffic sign (`0` = not present).
- **stop_sign**: Indicates whether a stop sign is visible (`0` = not present).
- **traffic_light:** Shows the state of the traffic light; usually set to `n/a` (not applicable) in this scene.

5.5 Comparison with Traditional Datasets

Given the importance of selecting an appropriate dataset for studying pedestrian behavior and their interactions in urban environments, it is beneficial to compare some of the most widely used datasets in this domain. Such a comparison provides a clearer understanding of the strengths and limitations of each dataset (JAAD , ETH , UCY , SDD), helping researchers make informed choices based on their specific objectives. Below is a comparison table:

Feature / Dataset	**jaad**[64]	**ETH**[61] / **UCY**[62]	**Stanford Drone (SDD)**[78]
Data Type	Video + Behavioral Annotations	Trajectory Coordinates (2D)	Drone Video + Trajectories
Camera Angle	From inside a vehicle (dashboard view)	Top-down / Fixed camera	Top-down (drone)
Pedestrian Behavior	Yes (looking, stopping, crossing, etc.)	No	Partially
Interaction with Vehicles	Yes	No	Yes
Annotations	Behavior + Demographics + Context	Trajectories only	Object locations only
Primary Purpose	Pedestrian intention and social behavior	Trajectory prediction	Trajectory prediction & group interaction

Objects in Scene	Pedestrians and vehicles	Pedestrians only	Pedestrians + Vehicles + Bicycles
Availability	Free for research	Free	Free

Table 4:Comparison between traditionel datasets

5.7Advantages of the JAAD Dataset

the JAAD dataset offers a valuable resource for research on pedestrian intent and joint attention in urban settings, thanks to its detailed behavioral annotations and realistic environments[66]. Here are some of its advantages:

5.7.1 New Focus on Human Intent and Joint Attention

JAAD is among the earliest datasets to directly record and annotate pedestrians' intent and joint attention behaviors such as glancing at the car or pointing. This focus makes it particularly well-suited for study of the social dynamics of drivers and pedestrians[66].

5.7.2 Naturalistic Urban Scenarios

The videos were recorded on real-world city streets under variable weather and lighting conditions (e.g., sunny, rainy, cloudy), which makes the dataset more applicable to real-world autonomous driving scenarios[75].

5.7.3 Fine-Grained Behavioral Annotations

Every pedestrian is annotated with fine-grained behavior labels such as walking, standing, crossing, looking, and hand gestures, which facilitate intention prediction and social behavior modeling research[79].

5.7.4 Well-Structured and Accessible

The data is available for research, provided in structured TXT and CSV formats which are easy to process using common data science tools[66].

5.8 Limitations of the JAAD Dataset

Though advanced, its limited scope, lack of 3D or vehicle data, and narrow cultural diversity make generalization and wider use in large-scale autonomous driving systems difficult. Some of its limitations include:

5.8.1 Constraints of Dataset Scale and Sensory Coverage

The JAAD dataset is limited in scale and sensory coverage, with only 346 clips, 2D annotations, and no 3D or vehicle motion data restricting its usefulness for complex behavior modeling and contextual driver-pedestrian interaction.[79].

5.8.2 Limitations in Behavioral Diversity and Cultural Representativeness

Most JAAD scenes feature only one or two pedestrians, limiting the modeling of group or crowd behaviors. Additionally, the dataset lacks cultural diversity, with recordings mainly from Canada and Russia, which may reduce the generalizability of learned behavior models[80].

5.9 Conclusion

This chapter has provided a detailed examination of the JAAD dataset, ranging from motion properties to behavioral deviations, interaction efficiency, and intention predictability of pedestrians. The dataset is unique compared to conventional datasets due to the availability of behavioral annotations and context-rich scenes. Its complexity demands state-of-the-art models that go beyond trajectory extrapolation to include intention recognition and attention modeling. In its current form, JAAD is an excellent dataset for developing and testing behavior-aware autonomous driving systems.

The following chapter explores the Social Force Model (SFM) as one of the most effective approaches in this area, highlighting its integration with deep learning techniques to enhance trajectory prediction and evaluate outcomes using metrics.

Chapter 6: Prediction on clean data

6.1 Introduction

Understanding pedestrians' behavior in urban areas is an essential element of the development of intelligent mobility systems, like autonomous cars, crowd management systems, and city planning tools. Among the theoretical models that have proved to be effective in this area, the Social Force Model (SFMs) is one of the strongest frameworks for explaining human interactions using both physical and behavioral laws. This chapter is a detailed examination of the SFMs, describing its key elements self-driving, social, and environmental forces, and how these are mathematically represented to achieve realistic pedestrian motion. Furthermore, it describes the approach utilized to extract motion and context features that underlie the model and discusses the manner in which features are fused within a hybrid framework that incorporates SFM and deep learning models with a view to enhancing the predictive accuracy of pedestrian trajectories. The chapter concludes by describing evaluation metrics and discussing performance outcomes from empirical experiments.

6.2 Social Force Models (SFMs):

Social Force Models (SFMs) are computational models used to simulate pedestrian movement in crowded environments based on the laws of classical mechanics[70]. These models represent the motion of individuals as driven by virtual social forces that encapsulate both psychological motivations and physical constraints[29]. The fundamental forces in SFMs include: (1) self-driving forces, which represent an individual's intention to move toward a goal at an optimal speed and direction; (2) social-interactive forces, which account for repulsive interactions with other pedestrians to avoid collisions and simulate group behavior, and (3) environmental forces[29], which arise from obstacles and physical boundaries such as walls or vehicles. The combined effect of these forces is used in differential equations to compute the acceleration and subsequent motion of each pedestrian. This allows SFMs to simulate collective phenomena such as lane formation, crowd bottlenecks, and dynamic path adjustments[29].

Despite their simplicity, SFMs provide a mathematically grounded and computationally efficient approach to modeling pedestrian dynamics[29].

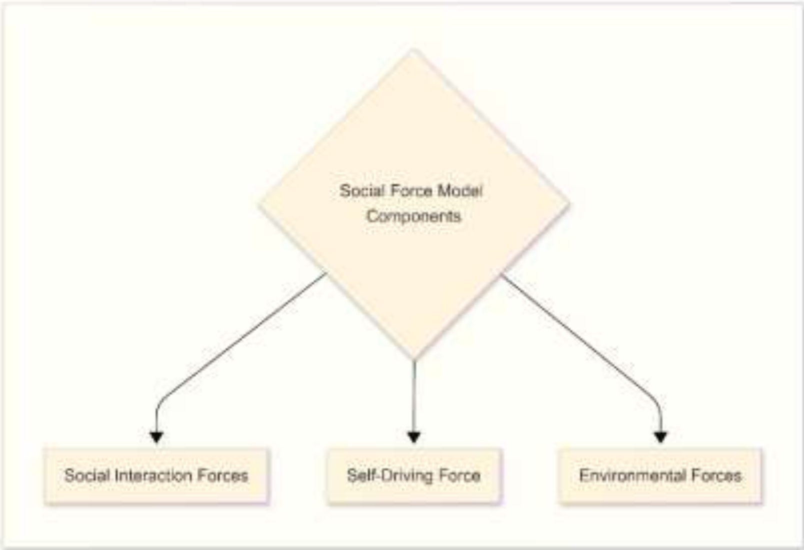

Figure 12:Components of the Social Force Model

6.3 Social Force-Based Pedestrian Modeling and it's feature Extraction

The effectiveness of pedestrian trajectory prediction using Social Force Models (SFM) largely depends on the accurate extraction and interpretation of features that characterize human motion[70], interactions[24], and environmental context[26]. These features can be categorized into three main types: **self-driven forces**, **social-interactive forces**, and **environmental forces**, each playing a specific role in the decision-making and movement patterns of pedestrians[29].

$$TotaL\ Force = Fself + Fsocial + F_{env} \quad [29]$$

(1): Social Force

6.3.1 Self-Driven forces Features

The **self-driving forces** represents the internal motivation of a pedestrian to move toward a specific destination with a desired speed and direction. It is goal-directed and dominant in unconstrained scenarios[81]. Formally, it is defined as:

$$Fself = m.\frac{v_{des} - v(t)}{T} \quad [81]$$

(2): self-driving forces

Where:

- m is the pedestrian's mass (usually normalized)[81],
- v_{des} is the desired velocity vector (i.e., speed and direction toward the goal)[81],
- $v(t)$ is the current velocity[81],
- T is a relaxation time parameter (e.g., 0.5–1.5 s) representing how quickly a pedestrian adjusts to the desired velocity[81].

This force seeks to minimize the discrepancy between the current velocity and the target one, enabling gradual and realistic acceleration toward a goal. For example, a pedestrian aiming to cross the street will align their heading and speed to reach the sidewalk on the opposite side while adjusting for real-time conditions[82].

Self-driven forces originate from the pedestrian's personal goal, such as moving toward a destination at a desired speed. The fundamental features include:

a) Velocity:

Represents the actual speed of the pedestrian, calculated from the x and y motion components. It indicates how fast the person is moving:

$$v_x(t) = \frac{x_t - x_{t-1}}{\Delta_t}, v_y(t) = \frac{y_t - y_{t-1}}{\Delta_t}, where\ \Delta_t = \frac{1}{30} seconds(for\ 30\ fps)$$

$$v_t = \sqrt{v_{x(t)}^2 + v_y^2}(t)$$

(3): Velocity

b) Acceleration:

It measures the change in a pedestrian's speed over time and is used to estimate self-propulsion or the need to adjust velocity:

$$a_x(t) = \frac{v_{x,t} - v_{x,t-1}}{\Delta_t} \cdot a_y(t) = \frac{v_{y,t} - v_{y,t-1}}{\Delta_t} \qquad a_t = \sqrt{a_{x,t}^2 + a_{y,t}^2}$$

(4): Acceleration

It indicates whether a pedestrian is speeding up or slowing down, which can be essential when anticipating stops or starts.

6.3.2 Social Interaction Features

The **social force** models the repulsive effect that other pedestrians and moving entities (e.g., vehicles) have on an individual to avoid collisions and preserve personal space. Each surrounding agent j exerts a force inversely proportional to the distance between them and the pedestrian. One popular formulation is:

$$Fsocial^{(j)} = A \cdot exp\left(\frac{r_{ij} - d_{ij}}{B}\right) \cdot n_{ij} \quad [29]$$

(5): Social Interaction

Where:

- A is the interaction strength,
- B is the interaction range,
- d_{ij} is the current distance between pedestrian iii and agent j,
- r_{ij} is the sum of their radii (modeling physical body sizes),
- $n_{ij} = \frac{P_i - P_j}{\|P_i - P_j\|}$ is the normalized direction vector pointing from j to i.

The exponential formulation reflects that pedestrians react more strongly to close threats than distant ones. This force causes people to veer away from each other or slow down to avoid overlap, especially in dense or crowded environments. When extended to include vehicles, a modified force can account for their higher speed and risk level[29].

The social interaction force simulates repulsive behaviors that prevent collisions with dynamic agents, especially vehicles.

a) Nearest Neighbor Distance:

The distance to the closest nearby pedestrian. This feature helps model social interaction and collision avoidance. Its mathematical expression is as follows:

$$d_{nn}(t) = min\sqrt{\left(x_t^j - x_t^i\right)^2 + \left(y_t^j - y_t^i\right)^2} \quad ,j \neq i$$

(6): Nearest Neighbor Distance

b) Nearest Neighbor Velocity X/Y:

These features capture the x and y components of the velocity of the pedestrian who is spatially closest to the target pedestrian at a given time step:

$$these\ are\ the\ values\ \left(v_{x,t}^j, v_{y\cdot t}^j\right)\ of\ the\ nearest\ neighbor\ j\ found\ usind\ the\ distance\ above$$

6.3.3 Environmental Features

Environmental forces arise from **static obstacles** in the environment, such as walls, curbs, buildings, parked cars, or street furniture. These obstacles induce a **repulsive potential** that prevents pedestrians from colliding with or walking through them. A typical formulation is:

$$F_{env}^{(k)} = A' \cdot exp\left(\frac{r_i - d_{ik}}{B'}\right) \cdot n_{ik} \quad [83]$$

(7): Environmental

Where:

- A' and B' are environment-specific constants,
- d_{ik} is the distance between pedestrian iii and obstacle k,
- n_{ik} is the normal vector from the obstacle surface pointing toward the pedestrian.

This force helps ensure that pedestrian trajectories remain physically feasible, such as walking around walls instead of through them. It also guides behavior on sidewalks or crossings, where environmental boundaries constrain motion into navigable zones[83].

In more advanced models, environmental forces can include **semantic features** such as designated walkways, road signs, or traffic light status. For example, pedestrians may wait before crossing when detecting a red light or may be attracted to crosswalks[83].

Environmental factors influence the decision-making process indirectly, as pedestrians adapt to physical and semantic surroundings.

a) Goal Distance:

The current distance from the pedestrian to their target location. It provides a measure of how far they are from their destination. Its mathematical expression is as follows:

$$d_{goaL} = \sqrt{\left(x_{gaoL} - x\right)^2 + \left(y_{goaL} - y\right)^2}$$

(8): Goal Distance

b) Goal Angle:

The angle between the pedestrian's current position and the direction toward their goal (last frame). It helps determine the desired walking direction. Its mathematical expression is as follows:

$$\theta_{g_{oaL}} = arctan\left(\frac{y_{g_{oaL}} - y}{x_{g_{oaL}} - x}\right)$$

(9): Goal Angle

This feature set provides a comprehensive foundation for modeling pedestrian behavior under the SFM framework. Each variable contributes to a better understanding of how pedestrians interact with their environment, adapt to moving agents, and pursue personal goals[84].

6.4 Classification Criteria:

The methodology section classifies hybrid knowledge-based models with deep learning for pedestrian trajectory prediction, evaluating model type, network structure, prediction task, and incorporation of deep learning. The classification determines the way hybrid models combine knowledge-based reasoning and deep learning to make improved predictions.

To evaluate pedestrian trajectory prediction models in an organized manner, we detail the following evaluation criteria from literature considered:

- **Model Type:** Checks if the approach is knowledge-based models such as the Social Force Model (SFM) or Optimal Reciprocal Collision Avoidance (ORCA) or other rule-based and graph-based approaches.
- **Network Structure:** Checks if models are rule-based, agent-based, or deep learning-based structures.
- **Prediction Tasks:** Categorizes tasks such as interaction modeling, trajectory planning, and collective behavior analysis.
- **Use of Advanced AI Techniques:** Investigates whether Reinforcement Learning (RL), Inverse Reinforcement Learning (IRL), or other AI-enforced methods are employed.
- **Model Applications:** Investigates model uses such as high-density crowd simulation, metro stations, stadiums, and mass events.
- **Interpretability:** Investigates how well the model explains its predictions and supports decision-making in real-world scenarios.

Aricle	**Model Type**			**Model Structure**	**predection tasks**			**the use of advanced AI**			**Model Applications**			**Interpretability**
	SFM	**ORCA**	**OTHER**		**interaction**	**trajectory planning**	**collective behavior analysis**	**RL**	**IRL**	**OTHER**	**high-density crowd simulations**			
											metro stations	**stadiums**	**large events**	
Ningbo Cao -2020- [85]	**X**		FSFM	SFM, enhanced with FL	SPT	OCE	IDE			Integrates FL with the SFM	PGF	OCE	OES	Highly Interpretable
Manon Prédhumeau 2020- [86]	**X**			SFM, enhanced with SRD	SPT	PMC	SRC			Integrates SRM with SFM	PGF	SMG	OES.	Highly Interpretable
Dongfang Yang 2020- [87]	**X**		VPI	SFM	CMA	PMD	PGA			**MPC**	PGF	OCE		Highly Interpretable

Maria Kabtou l- 2020- [88]			PVC	Behavioral model	CMA	PMC	PGA				CMA	OTM	OES	Highly Interpretable
Manon Prédhumeau -2021- [89]	**X**		ABM	**SFM, enhanced with ABM**	CMA	PMD	**PGA**			ABM,decision models, and robotic simulation tools	SAP	CMA	EBT	Highly Interpretable
Dongfan Yang -2021- [90]			SG-SFM	SFM, enhanced with SG-SFM	EBT	SPV				SG-SFM	SPS	OTM	EBT	Highly Interpretable
Sven Kreiss -2021- [91]			DSFM	SFM, enhanced with DNN	EPA	PMC	OTM			DNN	PGF	OCE	OES	Moderately interpretable
Peng Wang Xiaoda Wang -2021- [92]	**X**			SFM modified with psychological factors for Newton’s laws.	SPT	PMC	**IDE**			**psychological theories (Yerkes –**	PGF	OES	IDE	Highly interpretable

										Dodson Law)				
Jinrui Liu -2021-[93]			SCM+AFF	Cellular Automata enhanced with subdivided cells and anticipation field.	OES	OCE	ELE				PGF	OCE	OES	Highly interpretable
Senlin Mu -2022-[94]		**X**		DRL-based hierarchical model	SPT	PMD	**IDE**	**X**			OME	OCE	OES	Highly interpretable
Gaining Han-2022-[95]	**X**			Human-Vehicle -SFM	CMA	PMD	PRV			Combines behavior, social forces, and virtual forces for shared spaces.	SAP	OTM	EBT	Highly interpretable

Predhumeau-2022-[96]	**X**			ABM-SFM	CMA	PMC	PGA			Expert Model	SAP	CMA	EBT	Highly interpretable
Yongqing Guo-2022-[97]	**X**			SFM enhanced with signal countdown and pedestrian interaction.	CMA	EBT	SPM				OME	OCE	EBT	Highly interpretable
Balint Varga,2023[98]	**X**		MDP	White-box framework for pedestrian-AV interactions. .	CMA	PMC	PGA			Integrates DNN, MDP and PSO	SAP	CMA	PGA	High Interpretability
Iuliia Kotseruba,2023 [99]			Intend-Wait-Perceive-Cross Model	ABM with perceptual constraints like FoV, memory, and scanning.	SPIO	PMC	SPIO			Integrates FoV modeling, memory, .	**SPT**	PGF		High Interpretability
Haifeng Sang, 2024[42]	**X**			Physics-Constrained Model	CMA	PMC	PGA	**X**		*		OCE	OES	High Interpretability

Dongkun Zhang, 2024[43]		**X**		Policy-Embedded Planner integrating rule-based & neural policy	CMA	PMC	PGA	**X**				OCE, OTM		High Interpretability

Table 5:Comparative Summary of Pedestrian Trajectory Prediction Studies Based on Scene Type, Dataset, Prediction Task, and Evaluation Metrics.

(FL): Fuzzy Logic; **(SRD):** Social Relationship Dynamics; **(PSO):** Particle Swarm Optimization; **(MDP):** Markov Decision Process; **(SRM):** Social Relationship Modeling; **(VPI):** Vehicle-Pedestrian Interaction; **(MPC):** Model Predictive Control; **(PVC):** Pedestrian-Vehicle Cooperation Behavioral Model; **(SG-SFM):** Sub-Goal Social Force Model; **(SFM):** Social Force Modeling; **(DSFM):** Deep Social Force Model; **(DNN):** Deep Neural Networks; **(ABM):** Agent-Based Modeling; **(SCM):** Subdivided Cellular Automata Model; **(AFF):** Anticipation Floor Field. **(SPT):** Pedestrian interaction by social ties; **(SPS):** Pedestrian-AV interaction in shared spaces; **(EBT):** Pedestrian behavior in mixed traffic; **(OCE):** Crowd evacuation in events; **(PMC):** Pedestrian movement with cooperation; **(PMD):** Pedestrian movement by density & AVs; **(PGA):** Group dynamics with AVs; **(PGF):** Passenger flow by social ties; **(OTM):** Traffic management by crowd motion; **(CMA):** Crowd movement around AVs; **(OES):** Event space & evacuation optimization. **(*):** Integrates a differential constraint module to enforce motion law.

6.5 Hybrid architecture:

The diagram below represents a code-level abstraction of the proposed hybrid model architecture for pedestrian trajectory prediction. This hybrid approach combines explicit physical modeling through the Social Force Model (SFM) with the temporal learning power of Deep Learning (DL). Each block in the figure corresponds to an actual component or function within the implementation pipeline, highlighting how the system is organized in code and how data flows through each stage.

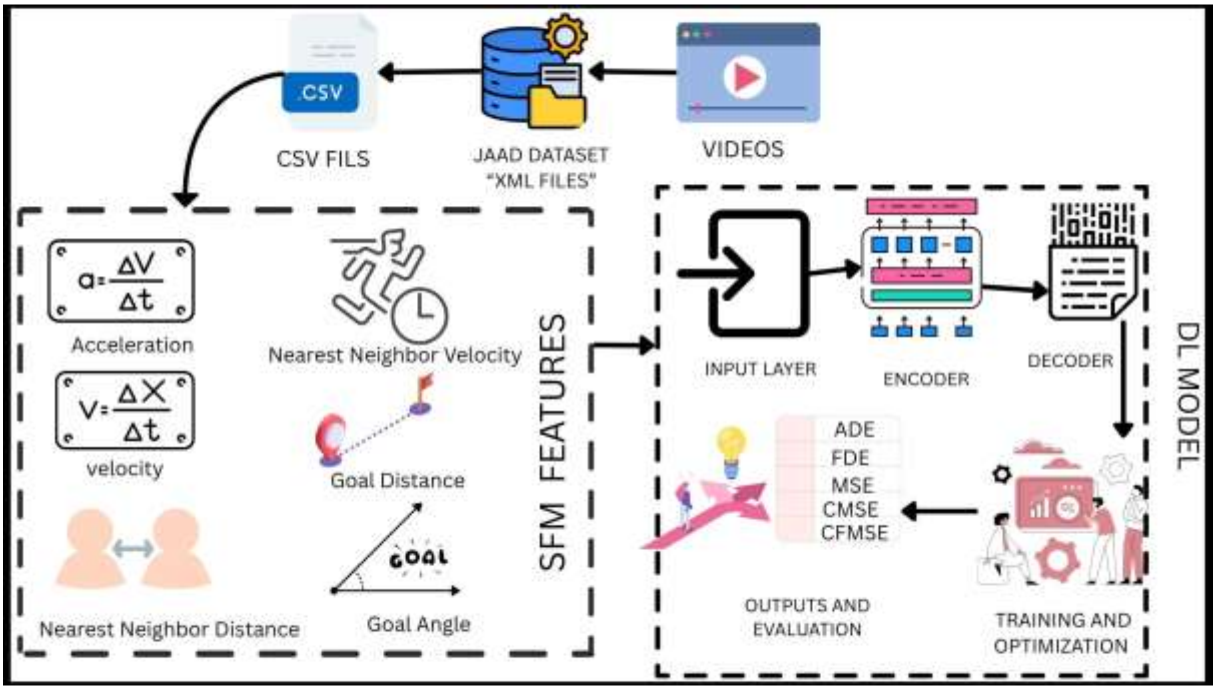

Figure 13:hybride model architecture

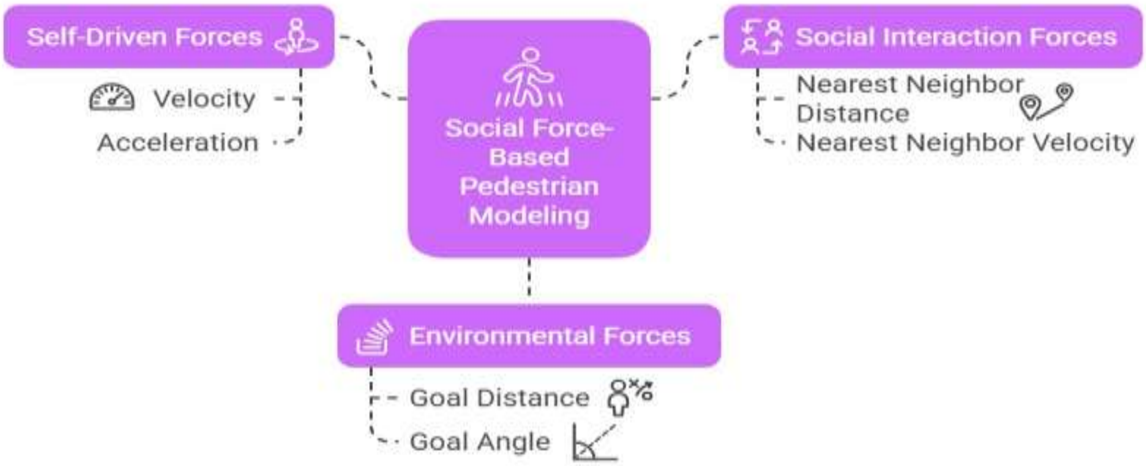

Figure 14:Categorization of Extracted Features for Pedestrian Trajectory Prediction

The hybrid model architecture is designed to combine the strengths of knowledge-driven and data-driven approaches for pedestrian trajectory prediction. It leverages structured physical and contextual insights from the Social Force Model (SFM)[29] alongside the temporal modeling capabilities of deep learning, specifically LSTM ,BiLSTM and GRU enhanced with attention and multi-stage decoding[18]. This integration enables the system to capture both low-level kinematic patterns and high-level social interactions, resulting in more accurate and context-aware trajectory forecasts.

a) Data Extraction and Preprocessing(JAAD XML and CSV Integration)

- The codebase begins by loading pedestrian trajectory data from the **JAAD dataset**, specifically from .xml and .csv files.
- XML files provide frame-by-frame annotations, including pedestrian locations, traffic light states, and vehicle actions.
- The resulting structured data includes time-series sequences of (x,y) coordinates, IDs, and contextual tags, and is saved in intermediate CSV files for modular access.

b) Knowledge-Based Module (KB Model using SFM)

- **SFM Integration**: The **Social Force Model** computes dynamic interaction-aware features to represent the influence of social and physical forces on pedestrian behavior.
- **Computed Features**:
    - **Kinematic Attributes**: Velocity, acceleration, and derived motion properties.
    - **Social Metrics**: Nearest neighbor distance and velocity.
    - **Goal-Oriented Descriptors**: Goal distance and goal angle.

c) Deep Learning Model (DL Model)

- **Input Layer:** The raw position data and the extracted SFM features are concatenated to form a dense input vector, which is passed into the deep learning pipeline.
- **LSTM Encoder :**
    - Learns sequential dependencies from historical motion data.
    - Captures short-term and long-term behavioral patterns of pedestrians.
- **Decoder :**
    - Refines encoded representations to produce future trajectory predictions.
    - Incorporates outputs from attention mechanisms, enabling focus on significant input segments.
- **Outputs and Evaluation :**
    - Model performance is assessed using common trajectory prediction metrics such as ADE (Average Displacement Error), FDE (Final Displacement Error), MSE, CMSE, and CFMSE.

6.6 Evaluation Metrics for Pedestrian Trajectory Prediction

To quantitatively assess the accuracy of trajectory prediction models, several error metrics are commonly used to compare predicted trajectories against the ground truth. These metrics provide insight into spatial deviations, temporal alignment, and overall prediction quality. In this section, we describe the most widely adopted metrics: Average Displacement Error (ADE), Final Displacement Error (FDE), Dynamic Time Warping (DTW), Mean Squared Error (MSE), Cumulative Mean Squared Error (CMSE), Cumulative Frame Mean Squared Error (CFMSE).

6.6.1 Average Displacement Error:

The Average Displacement Error (ADE) measures the mean Euclidean distance between the predicted and actual positions over the entire trajectory. It reflects the overall accuracy of the model across all predicted time steps.

$$ADE=\frac{1}{n}\sum_{i=1}^{n}\sqrt{(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2} \quad [100]$$

(10): Average Displacement Error

Where:

- n is the number of predicted time steps,
- (x_i, y_i)) are the ground truth coordinates at time step iii,
- $(\hat{x}_i, \hat{y}_i)$are the predicted coordinates at the same step.

This metric captures the **average spatial deviation** over time and is sensitive to consistent drift in the trajectory.

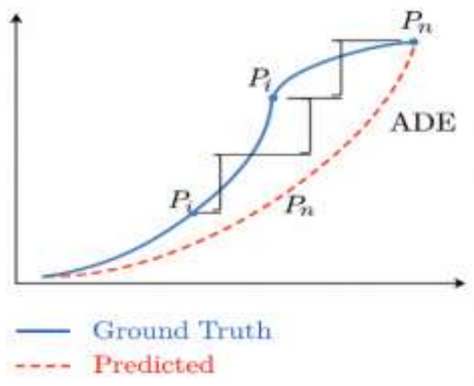

Figure 15:Average Displacement Error (ADE)

6.6.2 Final Displacement Error:

The Final Displacement Error (FDE) measures the Euclidean distance between the predicted final point and the ground truth final point. It focuses on long-term prediction accuracy.

$$FDE=\sqrt{(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2} \quad [100]$$

(11): Final Displacement Error

FDE is particularly important in applications where the final destination is critical, such as in autonomous navigation or human intention prediction.

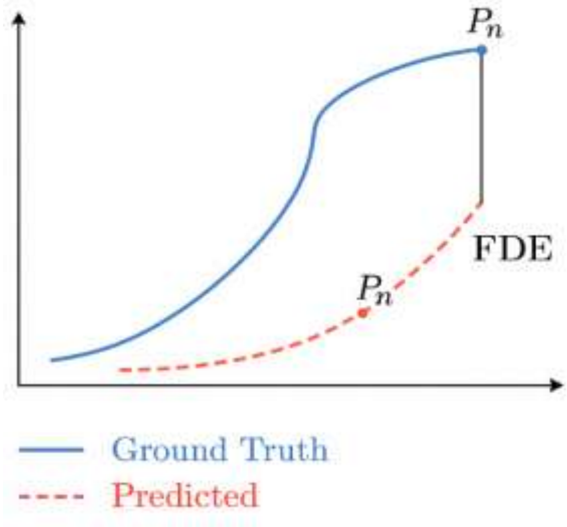

Figure 16:Final Displacement Error (FDE)

6.6.3 Dynamic Time Warping:

Dynamic Time Warping (DTW) is a temporal alignment metric that measures the similarity between two sequences (predicted and actual trajectories) by finding the optimal nonlinear mapping between them. Unlike ADE or FDE, DTW can handle trajectories that differ in time or speed but follow similar spatial paths.

$$DTW(T_1, T_2) = \min_{\pi} \sum_{(i,j)\in\pi} d(T1[i], T_2[j]) \qquad [101]$$

(12): Dynamic Time Warping

Where:

- π is a warping path.
- $d(\cdot,\cdot)$ the Euclidean distance.

Lower DTW values indicate more similar trajectories in both space and progression over time.

6.6.4 Mean Squared Error:

MSE is a standard statistical measure used to quantify the average squared difference between the predicted and actual positions:

$$MSE= \frac{1}{n}\sum_{i=1}^{n}\left\|P_i - \hat{P}_i\right\|^2 \qquad [102]$$

(13): Mean Squared Error

Where P_i and $\hat{P}_i$ represent the true and predicted points at timestep iii. MSE is widely used as a loss function during training due to its differentiability and penalization of larger errors.

6.6.5 Cumulative Frame Mean Squared Error:

CFMSE calculates the mean of the squared errors frame-by-frame (i.e., timestep by timestep) over the full trajectory. It reflects the accumulated error across time:

$$CFMSE=\frac{1}{n}\sum_{i=1}^{n}[(x_\lambda - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \qquad [103]$$

(14): Cumulative Frame Mean Squared Error

This metric provides insight into how errors evolve throughout the predicted sequence.

6.6.6 Cumulative Mean Squared Error:

CMSE is a time-dependent variation of MSE, calculated cumulatively up to a given timestep k. It helps track the model's performance over time:

$$CMSE_k = \frac{1}{k}\sum_{i=1}^{k}\left\|P_i - \hat{P}_i\right\|^2 \quad [103]$$

(15): Cumulative Mean Squared Error

By plotting CMSE over time, one can observe whether the model's prediction error grows or decreases along the trajectory.

6.7 Our Results of Metrics:

The results shown in the comparative table illustrate that integrating the Social Force Model (SFM) into deep learning architectures (LSTM, GRU, and BiLSTM) leads to consistent and significant performance improvements in pedestrian trajectory prediction. Across all temporal metrics including MSE at 0.5s, 1.0s, and 1.

Moreover, in terms of spatial error metrics, the SFM_BiLSTM model again outperformed all others, achieving the lowest Average Displacement Error (ADE = 17.0444) and Final Displacement Error (FDE = 24.6534). These improvements indicate that the hybrid model not only predicts more accurate positions over time but also better aligns with the true final destination of pedestrians. The SFM-enhanced LSTM and GRU models also showed strong improvements across most metrics, though to a slightly lesser extent than the BiLSTM configuration. This suggests that the bidirectional architecture of BiLSTM, when combined with socially-aware modeling, offers the best predictive capabilities among the tested variants.As figure 17 shows.

Overall, the experimental findings strongly support the hypothesis that embedding social interaction models such as SFM into data-driven frameworks significantly enhances their ability to anticipate realistic pedestrian trajectories. By explicitly accounting for inter-agent influences, collision avoidance, and group dynamics, these hybrid models better reflect the complexities of human movement.

Metric	LSTM	GRU	BiLSTM	SFM_LSTM	SFM_GRU	SFM_BiLSTM
MSE @ 0.5s	226.6002	595.8923	576.8247	365.5771	313.5808	**213.3009**
MSE @ 1.0s	482.4144	530.8178	711.3059	555.9243	388.8210	**287.0850**
MSE @ 1.5s	711.5703	2213.4063	1195.6312	983.8686	766.6855	**450.7435**
CMSE over 1.5s	473.3442	786.3802	803.4028	520.3348	392.8069	**280.8758**
CFMSE @ 1.5s	711.5703	2213.4063	1195.6312	983.8686	766.6855	**450.7435**
ADE	23.5094	33.1052	31.8195	22.8255	20.6705	**17.0444**
FDE	34.5808	65.6454	89.8911	35.7499	32.9701	**24.6534**

Table 6:Performance Comparison of Pedestrian Trajectory Prediction Methods

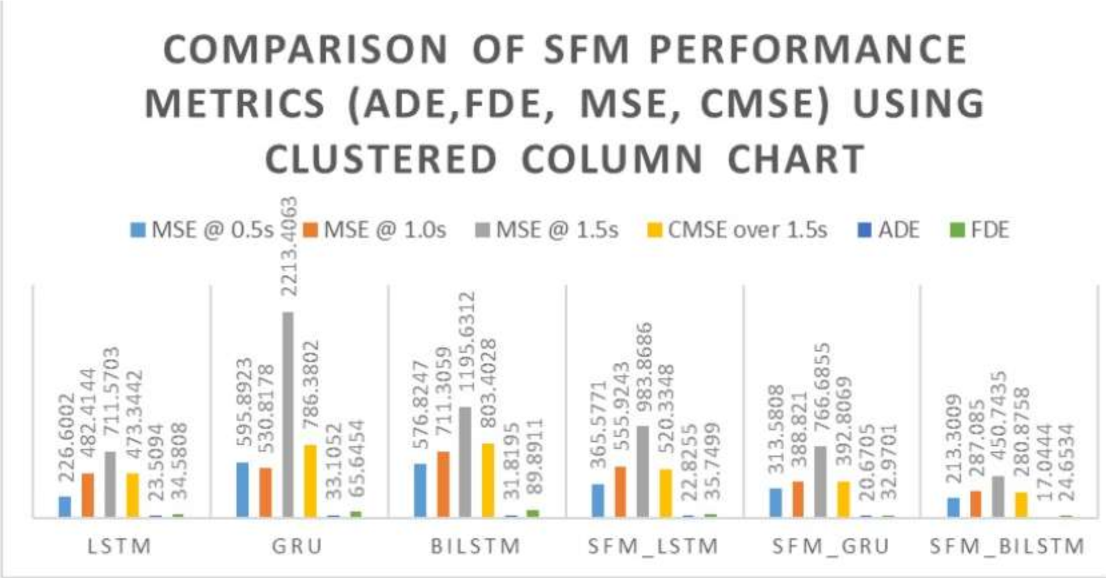

Figure 17:comparison of sfm performance metrics using clustered column chart

6.8 Visual Analysis of Prediction Quality:

In addition to the quantitative evaluation, this section presents a qualitative analysis aimed at gaining deeper insights into the predictive behavior of the proposed models. Through visual inspection of training curves, predicted trajectories, and positional distributions, we assess the models' ability to capture spatial and temporal dynamics in pedestrian motion. These visualizations serve not only to validate numerical findings but also to highlight strengths and limitations that may not be evident from metrics alone, particularly in socially complex or ambiguous scenarios.

6.8.1 DL Model Analysis

To complement the quantitative evaluation, we present qualitative insights into the performance of the deep learning (DL) model using BiLSTM. The analysis draws upon three visual components: the training curve, predicted trajectories, and coordinate distributions, which together provide a comprehensive perspective on the model's learning dynamics, spatial accuracy, and statistical fidelity.

a) Training Curve:

The BiLSTM-only model demonstrates a consistent decline in Mean Squared Error (MSE) across epochs. The training loss initiates at approximately 0.06 and steadily decreases to around 0.016, while the validation loss starts near 0.017 and reaches a stable value close to 0.004 by the final epochs. The small and stable gap between training and validation losses indicates good generalization, though a slightly slower convergence rate suggests limited model

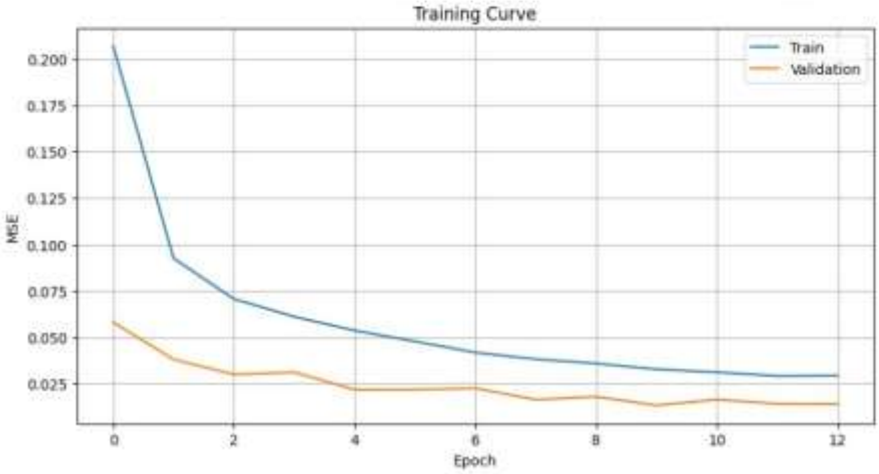

Figure 18:Training and validation loss curves for the BiLSTM model

capacity to fully capture complex temporal dependencies.

b) Trajectory Visualization:

The trajectory plots reveal that the BiLSTM model partially reconstructs the ground truth paths. While some predicted sequences closely follow the target (e.g., Trajectory 3 and Trajectory 5), others like Trajectory 1 and Trajectory 4 show moderate deviations, especially in curved or dynamic motion segments. This indicates that while the model captures general trends, it struggles to model fine-grained temporal transitions or sudden directional changes.

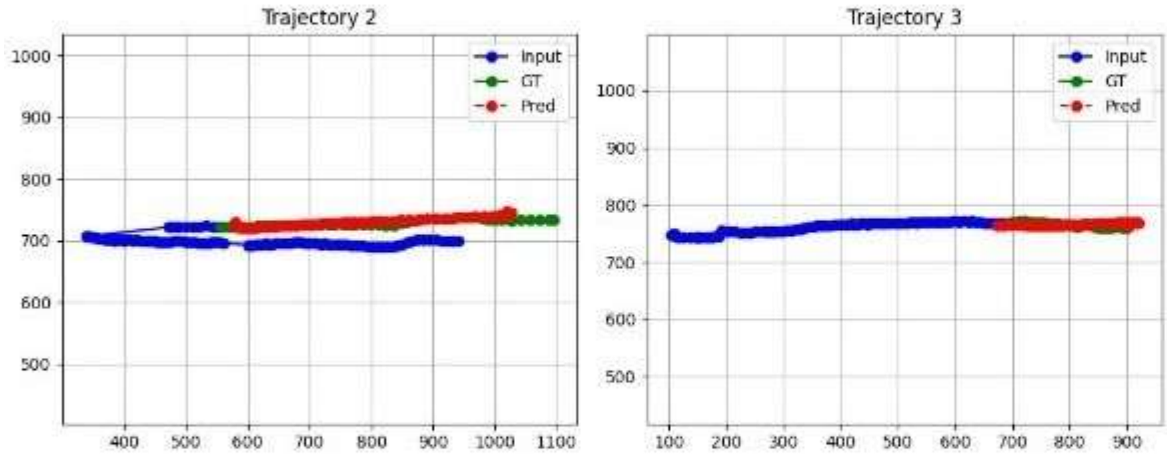

Figure 19:Predicted vs. ground truth trajectories using the BiLSTM model

c) Coordinate Distribution:

Histograms comparing predicted and ground truth coordinates show an overall alignment, particularly in the Y-axis distribution. However, discrepancies exist in peak heights and spread, especially in the X-axis, where some predicted values deviate from the actual statistical profile. This suggests that although the model approximates spatial distributions reasonably well, it lacks precision in replicating detailed spatial characteristics.

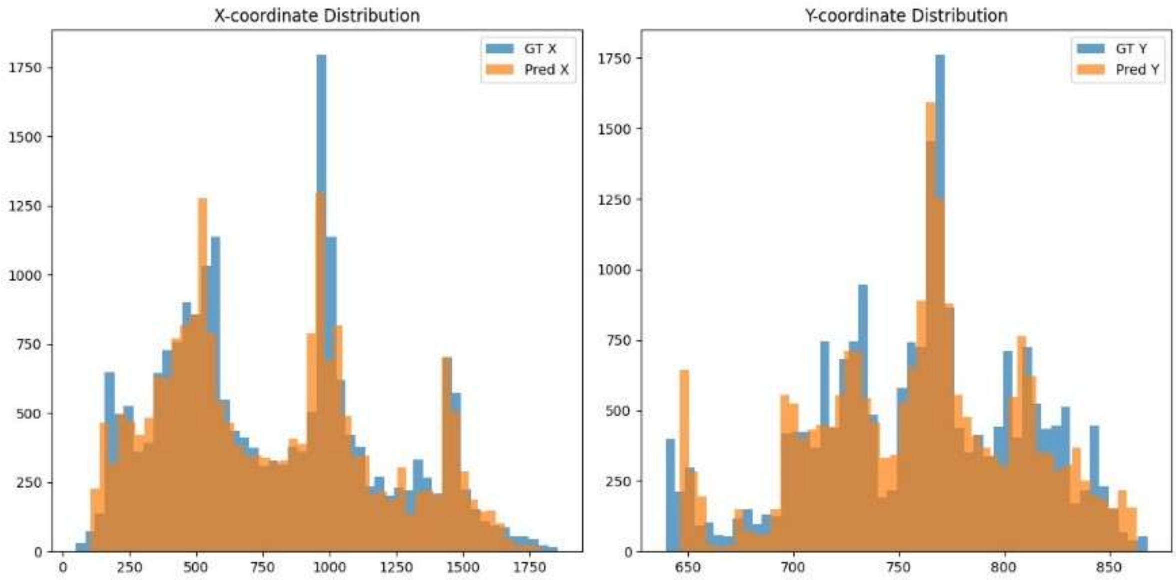

Figure 20:Distributions of predicted and ground truth X and Y coordinates for the BiLSTM model

6.8.2 Hybrid Model Analysis

This section provides qualitative insights into the hybrid architecture, which combines a Social Force Model (SFM) with BiLSTM. We use visual evidence of training dynamics, trajectory accuracy, and spatial distribution fidelity to assess the model's enhanced capabilities.

a) Training Curve

The hybrid model demonstrates rapid and stable convergence. The training loss decreases from approximately 0.06 to 0.016, while the validation loss sharply drops from about 0.017 to a stable minimum near 0.004. The lower and smoother validation curve indicates

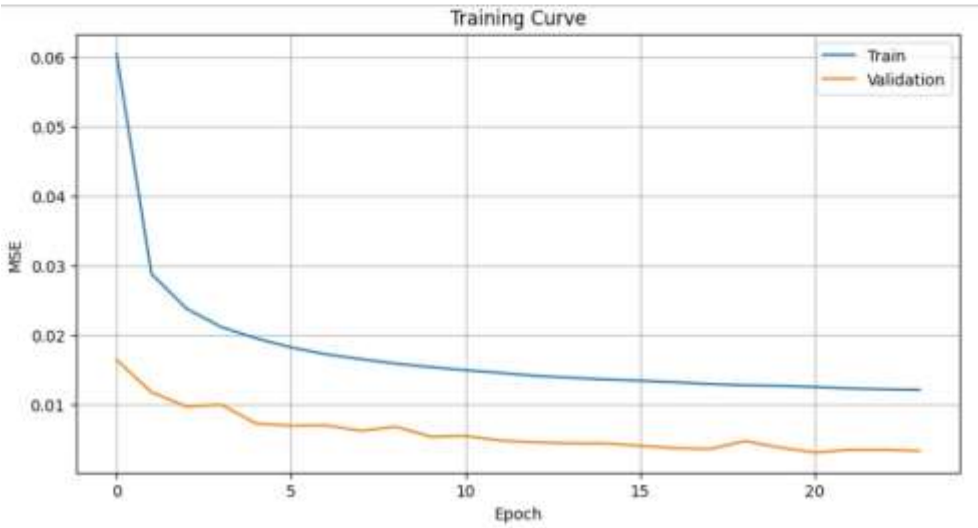

Figure 21:Training and validation loss curves for the Hybrid model (SFM + BiLSTM)

strong generalization with minimal overfitting. Compared to the BiLSTM-only model, this architecture learns more efficiently and effectively, thanks to the added physics-informed priors from the SFM component.

b) Trajectory Visualization

In this model, the predicted trajectories are highly accurate and almost indistinguishable from the ground truth across all tested samples. The hybrid architecture successfully captures the complex interactions and movement dynamics, particularly in challenging trajectories like

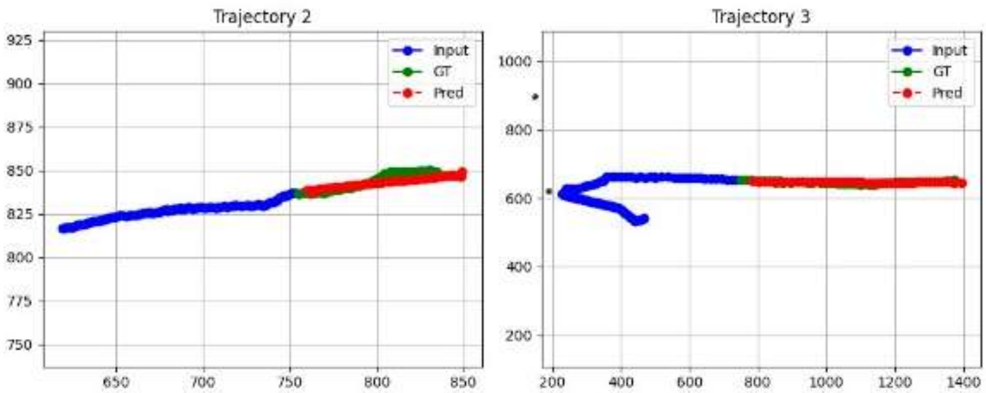

Figure 22:Predicted vs. ground truth trajectories using the Hybrid model

Trajectory 1 and Trajectory 3. This performance is attributed to the inclusion of the SFM, which enables the model to account for inter-agent interactions and social behaviors.

c) Coordinate Distribution

Predicted trajectories generated by the hybrid model show exceptional alignment with the ground truth, even in complex cases such as Trajectory 1 and Trajectory 3. The predictions follow the correct curvature and motion patterns with high fidelity. The inclusion of SFM provides the model with contextual knowledge about inter-agent dynamics, enabling it to make socially

Figure 23:Predicted vs. ground truth trajectories using the Hybrid model

aware

and temporally coherent predictions.

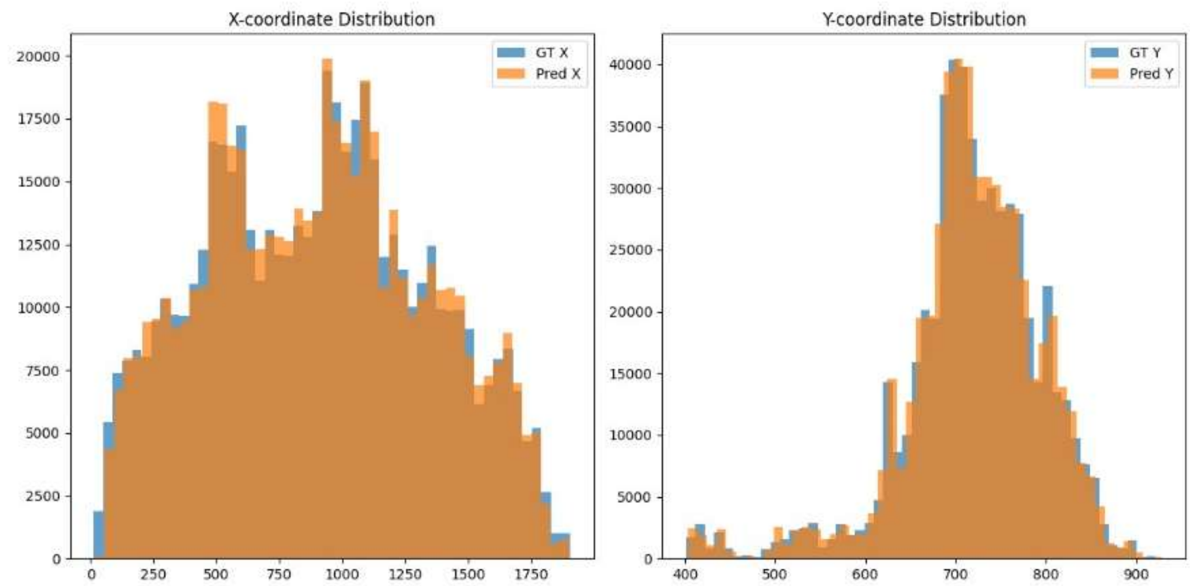

Figure 24:Distributions of predicted and ground truth X and Y coordinates for the Hybrid model

6.8.3 Predictive Performance Comparison between DL and Hybrid Architectures

This comparative analysis evaluates two trajectory prediction approaches: a standalone BiLSTM-based deep learning model and a hybrid model integrating BiLSTM with a Social Force Model (SFM). The BiLSTM-only architecture performs adequately, achieving consistent reductions in training and validation loss, and capturing coarse trajectory patterns. However, it faces difficulties in accurately modeling complex or socially-influenced movements, and in fully preserving the statistical characteristics of the spatial data.

In contrast, the hybrid model (SFM + BiLSTM) demonstrates significantly improved performance across all evaluation dimensions. It achieves lower prediction error, greater alignment between predicted and ground truth trajectories, and more accurate coordinate distributions. The integration of SFM enhances the model's contextual understanding, enabling more robust generalization and socially plausible predictions. Consequently, the hybrid architecture proves more effective for applications that require high-fidelity trajectory forecasting in environments where interactions and crowd behavior play a crucial role.

6.9 Conclusion:

This chapter has highlighted the Social Force Model as a solid theoretical framework for understanding and simulating pedestrian movement in complex environments. By integrating SFM with deep learning approaches, the predictive capabilities were significantly improved through the incorporation of dynamic, social, and contextual features extracted from real-world data. The experimental findings validated the efficacy of the hybrid model in minimizing trajectory prediction errors in both spatial and temporal contexts, thereby justifying its feasibility in practical applications. Overall, this study is a significant step towards the development of more accurate and context-sensitive systems that are able to interact with human-oriented environments.

CHAPTER 7: General Conclusion and Future directions

This thesis aimed to address the challenges of pedestrian trajectory prediction in complex urban environments by designing a hybrid model that combines various deep learning architectures Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (BiLSTM)with the Social Force Model (SFM). The proposed hybrid approach leverages the predictive power of these recurrent neural networks alongside the interpretability and physical reasoning provided by SFM, enabling a more accurate and socially-aware understanding of human movement behavior.

The JAAD dataset was utilized as the primary source for extracting rich behavioral and social features, due to its realistic scenes and interactive scenarios that showcase pedestrian behaviors and their interactions with vehicles and the environment.

Several standard evaluation metrics were adopted to assess the prediction performance of the hybrid models that integrate deep learning architectures (LSTM, GRU, BiLSTM) with the Social Force Model (SFM). Among the evaluated models, the BiLSTM with SFM hybrid model achieved the best performance across all this metrics (Average Displacement Error (ADE=17,04), Final Displacement Error (FDE=24,65), Mean Squared Error (MSE=213.30 in 0,5s), Cumulative Frame Mean Squared Error (CFMSE=450,74), and Cumulative Mean Squared Error (CMSE=280,87)), demonstrating its effectiveness in capturing both temporal dependencies and pedestrian trajectory.

The results demonstrated that the hybrid model outperformed the BiLSTM baseline in terms of prediction accuracy and quality across all metrics. This confirms the effectiveness of incorporating social and environmental dynamics into pedestrian modeling. However, the model still faces several limitations, such as:

- Limited cultural and behavioral diversity within the JAAD dataset.
- Constraints related to dataset scale and sensory coverage.

The hybrid approach incorporating Bidirectional LSTM (BiLSTM) networks and Social Force Model (SFM) has been successful in predicting pedestrian paths by uniting temporal dynamics with social interaction forces. Still, there are a number of promising future developments to further expand its predictive power, flexibility, and application scope:

- **From Trajectory Prediction to Intention Recognition:**

A meaningful extension of the current hybrid model is to shift from predicting precise future positions to also inferring pedestrian intent, such as crossing a street, yielding, or stopping. Recognizing intent can significantly improve the robustness of autonomous systems in complex environments.

- **Exploring Alternative Architectures:**

Replacing the LSTM component with alternative architectures like Convolutional Neural Networks (CNNs) or Transformer-based models can lead to better spatial-temporal understanding, especially when handling large-scale video or sensor data. Transformer-based models such as Social-STGCNN have shown state-of-the-art results in recent studies.

- **Exploring Alternative Datasets such as PIE**

The generalization capability of the hybrid model (BiLSTM + SFM) can be improved by training and evaluating it on diverse datasets that capture various pedestrian behaviors and environments. Shifting from the JAAD to PIE (which includes pedestrian intention annotations) offers new challenges and opportunities to develop a more comprehensive and robust model.

- **Utilizing the Model in Decision-Making Systems**

The results demonstrated that the model is suitable for real-time applications. Therefore, its usage can be extended to include:

  - Integration with path planning systems in autonomous vehicles to interact effectively with pedestrians.
  - Deployment in collaborative robots operating in crowded environments that require accurate prediction of human movement.

Bibliography

[1] R. Korbmacher and A. Tordeux, "Review of Pedestrian Trajectory Prediction Methods : Comparing Deep Learning and Knowledge-based Approaches," vol. 1, pp. 1–24.

[2] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 2255–2264, 2018, doi: 10.1109/CVPR.2018.00240.

[3] C. H. Chen and D. Ramanan, "3D human pose estimation = 2D pose estimation + matching," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 5759–5767, 2017, doi: 10.1109/CVPR.2017.610.

[4] X. Zhou, H. Ren, T. Zhang, X. Mou, Y. He, and C. Y. Chan, "Prediction of Pedestrian Crossing Behavior Based on Surveillance Video," *Sensors*, vol. 22, no. 4, pp. 1–18, 2022, doi: 10.3390/s22041467.

[5] M. A. Rezaei, F. Ayar, E. Javanmardi, M. Tsukada, and M. Javanmardi, "Where Do You Go? Pedestrian Trajectory Prediction using Scene Features," 2025, [Online]. Available: http://arxiv.org/abs/2501.13848

[6] E. Moreno *et al.*, "Pedestrian Crossing Intention Forecasting at Unsignalized Intersections Using Naturalistic Trajectories," *Sensors*, vol. 23, no. 5, 2023, doi: 10.3390/s23052773.

[7] M. Gulzar, Y. Muhammad, and N. Muhammad, "A Survey on Motion Prediction of Pedestrians and Vehicles for Autonomous Driving," *IEEE Access*, vol. 9, pp. 137957–137969, 2021, doi: 10.1109/ACCESS.2021.3118224.

[8] C. Zhang and C. Berger, "Pedestrian Behavior Prediction Using Deep Learning Methods for Urban Scenarios: A Review," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 10, pp. 10279–10301, 2023, doi: 10.1109/TITS.2023.3281393.

[9] F. Camara *et al.*, "Pedestrian Models for Autonomous Driving Part I: Low-Level Models, from Sensing to Tracking," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6131–6151, 2021, doi: 10.1109/TITS.2020.3006768.

[10]N. Ostermann-myrau and H. C. Mar, "Pedestrians and Robots : A Novel Dataset," no. iii.

[11]M. Qaraqe *et al.*, "PublicVision : A Secure Smart Surveillance System for Crowd Behavior Recognition," *IEEE Access*, vol. 12, no. January, pp. 26474–26491, 2024, doi: 10.1109/ACCESS.2024.3366693.

[12]R. A. Gonzalez, R. E. Ferro, and D. Liberona, "Government and governance in intelligent cities, smart transportation study case in Bogotá Colombia," *Ain Shams Eng. J.*, vol. 11, no. 1, pp. 25–34, 2020, doi: 10.1016/j.asej.2019.05.002.

[13]"AIModels.fyi." Accessed: May 30, 2025. [Online]. Available: https://www.aimodels.fyi/author-profile/wen-wei-03e08047-fa2a-4091-a28f-9bc103b8512a

[14]S. Agrawal, N. Dengler, and M. Bennewitz, "Evaluating Robot Influence on Pedestrian

Behavior Models for Crowd Simulation and Benchmarking," pp. 1–12.

[15] L. Dong, Z. He, C. Song, X. Yuan, and H. Zhang, "Multi-robot social-aware cooperative planning in pedestrian environments using attention-based actor-critic," *Artif. Intell. Rev.*, vol. 57, no. 4, pp. 1–22, 2024, doi: 10.1007/s10462-024-10739-w.

[16] "Trajectory Based Human Behaviour Understanding - Signal Processing, Artificial Intelligence and Vision Technologies." Accessed: May 30, 2025. [Online]. Available: https://research.qut.edu.au/saivt/research-projects/trajectory-based-human-behaviour-understanding/

[17] B. I. Sighencea, R. I. Stanciu, and C. D. Căleanu, "A review of deep learning-based methods for pedestrian trajectory prediction," *Sensors*, vol. 21, no. 22, 2021, doi: 10.3390/s21227543.

[18] B. Ghojogh and A. Ghodsi, "Recurrent Neural Networks and Long Short-Term Memory Networks: Tutorial and Survey," 2023, [Online]. Available: http://arxiv.org/abs/2304.11461

[19] E. Liang and M. Stamp, "Predicting pedestrian crosswalk behavior using Convolutional Neural Networks," *Traffic Inj. Prev.*, vol. 24, no. 4, pp. 338–343, 2023, doi: 10.1080/15389588.2023.2186734.

[20] J. Zhou *et al.*, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, no. September 2020, pp. 57–81, 2020, doi: 10.1016/j.aiopen.2021.01.001.

[21] Y. Yuan, X. Weng, Y. Ou, and K. Kitani, "AgentFormer: Agent-Aware Transformers for Socio-Temporal Multi-Agent Forecasting," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 1, no. 2, pp. 9793–9803, 2021, doi: 10.1109/ICCV48922.2021.00967.

[22] E. Benhamou, "Kalman filter demystified : from intuition to probabilistic graphical model to real case in financial markets," 2018.

[23] A. Lincoln, "Hidden Markov Model," no. January 2015, 2018, doi: 10.1007/978-1-4302-5990-9.

[24] S. Rezwana and N. Lownes, "Interactions and Behaviors of Pedestrians with Autonomous Vehicles: A Synthesis," *Futur. Transp.*, vol. 4, no. 3, pp. 722–745, 2024, doi: 10.3390/futuretransp4030034.

[25] A. Rudenko, L. Palmieri, M. Herman, K. M. Kitani, D. M. Gavrila, and R. O. Dec, "Human Motion Trajectory Prediction : A," 2019, doi: 10.1177/ToBeAssigned.

[26] M. Azarmi, M. Rezaei, T. Hussain, and C. Qian, "Local and Global Contextual Features Fusion for Pedestrian Intention Prediction," *Commun. Comput. Inf. Sci.*, vol. 1883 CCIS, pp. 1–13, 2023, doi: 10.1007/978-3-031-43763-2_1.

[27] C. Scholler, V. Aravantinos, F. Lay, and A. Knoll, "What the constant velocity model can teach us about pedestrian motion prediction," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1696–1703, 2020, doi: 10.1109/LRA.2020.2969925.

[28] K. Jayaraman, D. M. Tilbury, X. Jessie Yang, A. K. Pradhan, and L. P. Robert, "Analysis and Prediction of Pedestrian Crosswalk Behavior during Automated Vehicle Interactions," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 6426–6432, 2020, doi: 10.1109/ICRA40945.2020.9197347.

[29] D. Helbing and P. Molnár, "Social force model for pedestrian dynamics," *Phys. Rev. E*, vol. 51, no. 5, pp. 4282–4286, 1995, doi: 10.1103/PhysRevE.51.4282.

[30] Q. Liu, X. Shi, R. Jiang, H. Zhang, L. Lu, and R. Shibasaki, "Modeling interpretable social interactions for pedestrian trajectory," *Transp. Res. Part C Emerg. Technol.*, vol. 162, 2024, doi: 10.1016/j.trc.2024.104617.

[31] P. Kothari, S. Kreiss, and A. Alahi, "Human Trajectory Forecasting in Crowds: A Deep Learning Perspective," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 7386–7400, 2022, doi: 10.1109/TITS.2021.3069362.

[32] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications," *Information*, vol. 15, no. 9, p. 517, 2024, doi: 10.3390/info15090517.

[33] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-fei, and S. Savarese, "Social LSTM : Human Trajectory Prediction in Crowded Spaces".

[34] Y. Yao, E. Atkins, M. Johnson-Roberson, R. Vasudevan, and X. Du, "Coupling Intent and Action for Pedestrian Crossing Behavior Prediction," *IJCAI Int. Jt. Conf. Artif. Intell.*, pp. 1238–1244, 2021, doi: 10.24963/ijcai.2021/171.

[35] R. Chandra, H. Karnan, N. Mehr, P. Stone, and J. Biswas, "Multi-Agent Inverse Reinforcement Learning in Real World Unstructured Pedestrian Crowds," pp. 1–8, 2024, [Online]. Available: http://arxiv.org/abs/2405.16439

[36] Y. Chen, B. Ivanovic, and M. Pavone, "ScePT: Scene-consistent, Policy-based Trajectory Predictions for Planning," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 17082–17091, 2022, doi: 10.1109/CVPR52688.2022.01659.

[37] O. Data, "Research on a Pedestrian Crossing Intention," 2020.

[38] A. Rasouli, I. Kotseruba, T. Kunic, and J. Tsotsos, "PIE: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction," *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-October, pp. 6261–6270, Oct. 2019, doi: 10.1109/ICCV.2019.00636.

[39] J. Yang, W. Y. He, T. L. Zhang, C. L. Zhang, L. Zeng, and B. F. Nan, "Research on subway pedestrian detection algorithms based on SSD model," *IET Intell. Transp. Syst.*, vol. 14, no. 11, pp. 1491–1496, 2020, doi: 10.1049/iet-its.2019.0806.

[40] J. Gesnouin, S. Pechberti, G. Bresson, B. Stanciulescu, and F. Moutarde, "Predicting intentions of pedestrians from 2d skeletal pose sequences with a representation-focused multi-branch deep learning network," *Algorithms*, vol. 13, no. 12, 2020, doi: 10.3390/a13120331.

[41] J. Lorenzo, I. Parra, F. Wirth, C. Stiller, D. F. Llorca, and M. A. Sotelo, "RNN-based Pedestrian Crossing Prediction using Activity and Pose-related Features," *IEEE Intell. Veh. Symp. Proc.*, pp. 1801–1806, 2020, doi: 10.1109/IV47402.2020.9304652.

[42] J. Hayakawa and B. Dariush, "Recognition and 3D Localization of Pedestrian Actions from Monocular Video," *2020 IEEE 23rd Int. Conf. Intell. Transp. Syst. ITSC 2020*, 2020, doi: 10.1109/ITSC45102.2020.9294551.

[43] X. Dan, “Spatial-Temporal Block and LSTM Network for Pedestrian Trajectories Prediction,” 2020.

[44] A. Kalatian and B. Farooq, “A context-aware pedestrian trajectory prediction framework for automated vehicles,” *Transp. Res. Part C Emerg. Technol.*, vol. 134, 2022, doi: 10.1016/j.trc.2021.103453.

[45] J. Lorenzo, I. Parra, and M. A. Sotelo, “IntFormer: Predicting pedestrian intention with the aid of the Transformer architecture,” 2021.

[46] C. Zhang, C. Berger, and M. Dozza, “Social-IWSTCNN: A social interaction-weighted spatio- temporal convolutional neural network for pedestrian trajectory prediction in urban traffic scenarios,” *IEEE Intell. Veh. Symp. Proc.*, vol. 2021-July, pp. 1515–1522, 2021, doi: 10.1109/IV48863.2021.9575958.

[47] R. Rozenberg, J. Gesnouin, and F. Moutarde, “Asymmetrical Bi-RNN for pedestrian trajectory encoding,” 2021.

[48] L. Achaji, J. Moreau, T. Fouqueray, F. Aioun, and F. Charpillet, “Is attention to bounding boxes all you need for pedestrian action prediction?,” *IEEE Intell. Veh. Symp. Proc.*, vol. 2022-June, pp. 895–902, 2022, doi: 10.1109/IV51971.2022.9827084.

[49] J. Gesnouin, S. Pechberti, B. Stanciulcscu, and F. Moutarde, “TrouSPI-Net: Spatio-temporal attention on parallel atrous convolutions and U-GRUs for skeletal pedestrian crossing prediction,” *Proc. - 2021 16th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2021*, 2021, doi: 10.1109/FG52635.2021.9666989.

[50] C. Zhang, “Predicting Pedestrian Behavior in Urban Traffic Scenarios Using Deep Learning Methods,” 2022, [Online]. Available: https://research.chalmers.se/en/publication/530314%0Ahttps://research.chalmers.se/publication/530314/file/530314_Fulltext.pdf

[51] R. Joshi and D. L. Silver, “Pedestrian Traffic Prediction using Deep Learning,” *Proc. Int. Florida Artif. Intell. Res. Soc. Conf. FLAIRS*, vol. 35, p. 2021, 2022, doi: 10.32473/flairs.v35i.130731.

[52] A. Rasouli and I. Kotseruba, “PedFormer: Pedestrian Behavior Prediction via Cross-Modal Attention Modulation and Gated Multitask Learning,” *Proc. - IEEE Int. Conf. Robot. Autom.*, vol. 2023-May, pp. 9844–9851, 2023, doi: 10.1109/ICRA48891.2023.10161318.

[53] D. Guo, T. Mordan, and A. Alahi, “Pedestrian Stop and Go Forecasting with Hybrid Feature Fusion,” *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 940–947, 2022, doi: 10.1109/ICRA46639.2022.9811664.

[54] J. A. Abbasi, N. M. Imran, L. C. Das, and M. Won, “WatchPed : Pedestrian Crossing Intention Prediction Using Embedded Sensors of Smartwatch,” no. 3.

[55] Z. Zhang, R. Tian, and Z. Ding, “TrEP: Transformer-Based Evidential Prediction for Pedestrian Intention with Uncertainty,” *Proc. 37th AAAI Conf. Artif. Intell. AAAI 2023*, vol. 37, pp. 3534–3542, 2023, doi: 10.1609/aaai.v37i3.25463.

[56] J. Huang, A. Gautam, and S. Saripalli, “Learning Pedestrian Actions to Ensure Safe Autonomous Driving,” *IEEE Intell. Veh. Symp. Proc.*, vol. 2023-June, 2023, doi: 10.1109/IV55152.2023.10186530.

[57] A. Antonucci, G. P. R. Papini, L. Palopoli, and D. Fontanelli, “Generating Reliable and Efficient Predictions of Human Motion: A Promising Encounter between Physics and Neural Networks,” no. June, 2020, doi: 10.1109/ACCESS.2021.3138614.

[58] P. Kothari, B. Sifringer, and A. Alahi, “Interpretable social anchors for human trajectory forecasting in crowds,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 15551–15561, 2021, doi: 10.1109/CVPR46437.2021.01530.

[59] J. Gesnouin, S. Pechberti, B. Stanciulescu, and F. Moutarde, “Assessing Cross-dataset Generalization of Pedestrian Crossing Predictors”.

[60] P. S. Chib and P. Singh, “Trajectory Prediction with Missing Data: Datasets, ImpPedestrianutation, and Benchmarking,” *Adv. Neural Inf. Process. Syst.*, vol. 37, no. NeurIPS, pp. 1–3, 2024.

[61] “ETH Dataset | Papers With Code.” Accessed: May 12, 2025. [Online]. Available: https://paperswithcode.com/dataset/eth

[62] “UCY Dataset | Papers With Code.” Accessed: May 12, 2025. [Online]. Available: https://paperswithcode.com/dataset/ucy

[63] V. Balachandran, J. Chen, N. Joshi, B. Nushi, and J. W. S. Yousefi, “EUREKA: Evaluating and Understanding Large Foundation Models,” *Msr ’24*, pp. 1–55, 2024.

[64] “JAAD dataset.” Accessed: May 12, 2025. [Online]. Available: https://data.nvision2.eecs.yorku.ca/JAAD_dataset/

[65] A. Rasouli, I. Kotseruba, T. Kunic, and J. Tsotsos, “PIE: A large-scale dataset and models for pedestrian intention estimation and trajectory prediction,” *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2019-Octob, pp. 6261–6270, 2019, doi: 10.1109/ICCV.2019.00636.

[66] I. Kotseruba, A. Rasouli, and J. K. Tsotsos, “Joint Attention in Autonomous Driving (JAAD),” pp. 1–10, 2016, [Online]. Available: http://arxiv.org/abs/1609.04741

[67] A. Forde and J. Daniel, “Pedestrian walking speed at un-signalized midblock crosswalk and its impact on urban street segment performance,” *J. Traffic Transp. Eng. (English Ed.*, vol. 8, no. 1, pp. 57–69, 2021, doi: 10.1016/j.jtte.2019.03.007.

[68] A. Hayat, M. D. Fernando, B. P. Bhuyan, and R. Tomar, “Human Activity Recognition for Elderly People Using Machine and Deep Learning Approaches,” *Inf.*, vol. 13, no. 6, 2022, doi: 10.3390/info13060275.

[69] M. Chaabane, A. Trabelsi, N. Blanchard, and R. Beveridge, “Looking ahead: Anticipating pedestrians crossing with future frames prediction,” *Proc. - 2020 IEEE Winter Conf. Appl. Comput. Vision, WACV 2020*, pp. 2286–2295, 2020, doi: 10.1109/WACV45572.2020.9093426.

[70] Ó. Gil and A. Sanfeliu, “Human Motion Trajectory Prediction Using the Social Force Model for Real-Time and Low Computational Cost Applications,” *Lect. Notes Networks Syst.*, vol. 976

LNNS, pp. 235–247, 2024, doi: 10.1007/978-3-031-58676-7_19.

[71] T. Talaei Khoei, H. Ould Slimane, and N. Kaabouch, “Deep learning: systematic review, models, challenges, and research directions,” *Neural Comput. Appl.*, vol. 35, no. 31, pp. 23103–23124, 2023, doi: 10.1007/s00521-023-08957-4.

[72] S. Liu, J. Sun, P. Yao, Y. Zhu, T. Mao, and Z. Wang, “DTDNet: Dynamic Target Driven Network for pedestrian trajectory prediction,” *Front. Neurosci.*, vol. 18, p. 1346374, Apr. 2024, doi: 10.3389/FNINS.2024.1346374/BIBTEX.

[73] B. Zhang, J. Amirian, H. Eberle, J. Pettré, C. Holloway, and T. Carlson, “From HRI to CRI: Crowd Robot Interaction—Understanding the Effect of Robots on Crowd Motion: Empirical Study of Pedestrian Dynamics with a Wheelchair and a Pepper Robot,” *Int. J. Soc. Robot.*, vol. 14, no. 3, pp. 631–643, 2022, doi: 10.1007/s12369-021-00812-7.

[74] N. Uhlemann, F. Fent, and M. Lienkamp, “Evaluating Pedestrian Trajectory Prediction Methods With Respect to Autonomous Driving,” *IEEE Trans. Intell. Transp. Syst.*, pp. 1–11, 2024, doi: 10.1109/TITS.2024.3386195.

[75] A. Rasouli, I. Kotseruba, and J. K. Tsotsos, “Are They Going to Cross? A Benchmark Dataset and Baseline for Pedestrian Crosswalk Behavior,” *Proc. - 2017 IEEE Int. Conf. Comput. Vis. Work. ICCVW 2017*, vol. 2018-Janua, pp. 206–213, 2017, doi: 10.1109/ICCVW.2017.33.

[76] N. Carolina *et al.*, “Enhancing AV Traffic Safety through Pedestrian Detection, Classification, and Communication [Project],” *Enhancing AV Traffic Saf. through Pedestr. Detect. Classif. Commun.*, no. 2019, 2022, [Online]. Available: http://connect.ncdot.gov/projects/research/Pages/ProjectSearch.aspx%0Ahttp://connect.ncdot.gov/projects/research/Pages/ProjDetails.aspx?ProjectID=2019-28%0Ahttps://trid.trb.org/view/1723547

[77] Y. Liu, Z. Ye, R. Wang, B. Li, Q. Z. Sheng, and L. Yao, “Uncertainty-Aware Pedestrian Trajectory Prediction via Distributional Di ff usion,” 2024.

[78] “Computational Vision and Geometry Lab.” Accessed: May 12, 2025. [Online]. Available: https://cvgl.stanford.edu/projects/uav_data/

[79] I. Kotseruba, A. Rasouli, and J. K. Tsotsos, “Benchmark for evaluating pedestrian action prediction,” *Proc. - 2021 IEEE Winter Conf. Appl. Comput. Vision, WACV 2021*, pp. 1257–1267, 2021, doi: 10.1109/WACV48630.2021.00130.

[80] A. Rasouli, I. Kotseruba, and J. K. Tsotsos, “It’s not all about size: On the role of data properties in pedestrian detection,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11129 LNCS, pp. 210–225, 2019, doi: 10.1007/978-3-030-11009-3_12.

[81] F. Farina, D. Fontanelli, A. Garulli, A. Giannitrapani, and D. Prattichizzo, “Walking ahead: The headed social force model,” *PLoS One*, vol. 12, no. 1, pp. 1–23, 2017, doi: 10.1371/journal.pone.0169734.

[82] W. Wu, M. Chen, J. Li, B. Liu, and X. Zheng, “An Extended Social Force Model via Pedestrian Heterogeneity Affecting the Self-Driven Force,” *IEEE Trans. Intell. Transp. Syst.*, vol.

23, no. 7, pp. 7974–7986, 2022, doi: 10.1109/TITS.2021.3074914.

[83] E. Cristiani and D. Peri, "Handling obstacles in pedestrian simulations: Models and optimization," *Appl. Math. Model.*, vol. 45, no. October, pp. 285–302, 2017, doi: 10.1016/j.apm.2016.12.020.

[84] U. DigitalCommons, U. All Graduate Theses, and S. Boyer, "Pedestrian Behavior at Signalized Intersections Throughout Utah," 2022, [Online]. Available: https://digitalcommons.usu.edu/etd/8561

[85] N. Cao, L. Zhao, M. Chen, and R. Luo, "Fuzzy Social Force Model for Pedestrian Evacuation under View-Limited Condition," *Math. Probl. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/2879802.

[86] M. Prédhumeau, J. Dugdale, and A. Spalanzani, "Modeling and Simulating Pedestrian Social Group Behavior with Heterogeneous Social Relationships," *Proc. 2020 Spring Simul. Conf. SpringSim 2020*, 2020, doi: 10.22360/SpringSim.2020.HSAA.004.

[87] D. Yang, K. Redmill, and U. Ozguner, "A Multi-State Social Force Based Framework for Vehicle-Pedestrian Interaction in Uncontrolled Pedestrian Crossing Scenarios," *IEEE Intell. Veh. Symp. Proc.*, pp. 1807–1812, 2020, doi: 10.1109/IV47402.2020.9304561.

[88] M. Kabtoul, A. Spalanzani, and P. Martinet, "Towards Proactive Navigation: A Pedestrian-Vehicle Cooperation Based Behavioral Model," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 6958–6964, 2020, doi: 10.1109/ICRA40945.2020.9196669.

[89] M. Prédhumeau, "Simulating realistic pedestrian behaviors in the context of autonomous vehicles in shared spaces," *Proc. Int. Jt. Conf. Auton. Agents Multiagent Syst. AAMAS*, vol. 3, pp. 1817–1819, 2021.

[90] D. Yang, F. T. Johora, K. A. Redmill, Ü. Özgüner, and J. P. Müller, "Sub-Goal Social Force Model for Collective Pedestrian Motion Under Vehicle Influence," vol. XX, no. X, pp. 1–14, 2021.

[91] S. Kreiss, "Deep Social Force," 2021.

[92] P. Wang and X. Wang, "Social Force in Pedestrian Crowd," pp. 1–13, 2021.

[93] J. Liu, M. Li, and P. Shu, "Subdivided cellular automata model considering anticipation floor field and analysis of pedestrian detour behavior," *Sustain.*, vol. 13, no. 19, 2021, doi: 10.3390/su131910621.

[94] S. Mu, X. Huang, M. Wang, D. Zhang, D. Xu, and X. Li, "Optimizing pedestrian simulation based on expert trajectory guidance and deep reinforcement learning," *Geoinformatica*, vol. 27, no. 4, pp. 709–736, 2023, doi: 10.1007/s10707-023-00486-5.

[95] G. Han, Z. Wu, W. Zhang, and W. Wang, "Research on the Behavioral Dynamics Motion Planning Method of the Human-Vehicle Social Force Model," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3154532.

[96] M. Predhumeau, L. Mancheva, J. Dugdale, and A. Spalanzani, "Agent-Based Modeling for Predicting Pedestrian Trajectories Around an Autonomous Vehicle," *J. Artif. Intell. Res.*, vol. 73,

pp. 1385–1433, 2022, doi: 10.1613/jair.1.13425.

[97] Y. Guo, S. Ma, F. Wei, L. Lu, F. Sun, and J. Wang, “Analysis of Behavior Characteristics for Pedestrian Twice-Crossing at Signalized Intersections Based on an Improved Social Force Model,” *Sustain.*, vol. 14, no. 4, 2022, doi: 10.3390/su14042003.

[98] B. Varga, D. Yang, and S. Hohmann, “Intention-Aware Decision-Making for Mixed Intersection Scenarios,” *SACI 2023 - IEEE 17th Int. Symp. Appl. Comput. Intell. Informatics, Proc.*, no. 19, pp. 369–374, 2023, doi: 10.1109/SACI58269.2023.10158550.

[99] I. Kotseruba and A. Rasouli, “Intend-Wait-Perceive-Cross: Exploring the Effects of Perceptual Limitations on Pedestrian Decision-Making,” *IEEE Intell. Veh. Symp. Proc.*, vol. 2023-June, 2023, doi: 10.1109/IV55152.2023.10186749.

[100] A. Mohamed, D. Zhu, W. Vu, M. Elhoseiny, and C. Claudel, “Social-Implicit: Rethinking Trajectory Prediction Evaluation and The Effectiveness of Implicit Maximum Likelihood Estimation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13682 LNCS, pp. 463–479, 2022, doi: 10.1007/978-3-031-20047-2_27.

[101] “An introduction to Dynamic Time Warping.” Accessed: May 22, 2025. [Online]. Available: https://rtavenar.github.io/blog/dtw.html?utm_source=chatgpt.com

[102] “Understanding Mean Squared Error (MSE) in Machine Learning.” Accessed: May 22, 2025. [Online]. Available: https://futureagi.com/blogs/mean-squared-error-2025

[103] Y. He, Y. Sun, Y. Cai, C. Yuan, J. Shen, and L. Tian, “Multi-modal Pedestrian Trajectory Prediction based on Pedestrian Intention for Intelligent Vehicle,” *KSII Trans. Internet Inf. Syst.*, vol. 18, no. 6, pp. 1562–1582, 2024, doi: 10.3837/tiis.2024.06.008.

ANNNEX

PRML 2025 THE 6TH INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION AND MACHINE LEARNING

Notification of Acceptance of the PRML 2025

Chongqing, China, June 13-16, 2025

http://www.prml.org

Engineering Village

Paper ID : CU0033

Paper Title : From Rules to Learning: Hybridizing Knowledge-Based Models for Pedestrian Trajectory Prediction

Dear Messaoud Mezati, Siham Beggaa, Houria Benboubkeur, Chahd Braithel, and Malak Ghouila,

First of all, thank you for your submission to PRML 2025. The review procedure for your paper has been finished. We are delighted to inform you that your manuscript has been accepted for presentation at the 6th International Conference on Pattern Recognition and Machine Learning (PRML 2025) in Chongqing, China during June 13 to 16, 2025. Your paper was double blind-reviewed, and based on the evaluations, your paper has successfully passed the final review. The reviewers' comments are enclosed.

The conference has received submissions from about 12 different countries and regions during the submission period. According to the recommendations from reviewers and technical program committees, we are glad to inform you that your paper has been selected for oral presentation and publication. You are invited to present your academic achievements in PRML 2025.

PRML 2025 is co-sponsored by Chongqing University of Posts and Telecommunications and Sichuan University, supported by Xinjiang University, University of Electronic Science and Technology of China, Xi'an Jiaotong University, and Tibet University, Wenzhou Medical University, University of Malaysia, etc.

This paper of PRML 2025 will be published in the **PRML 2025 Conference Proceedings**, which will be indexed by **Ei Compendex** and **Scopus.**

(Important Steps for your registration): Please do finish all the 4 steps on time to guarantee the paper published in the proceeding successfully:

1. Revise your paper according to the Review Comments in the attachment carefully. (Eight authors at most each paper)

2. Format your paper according to the Template carefully. (Choose word or latex)
http://prml.org/template.docx (word)

3. Register and pay Registration fee through the online system.

http://confsys.iconf.org/register/prml2025 (An account is needed for online registration.)

4. Send your final papers (both .doc (word)/.zip (latex) and .pdf format) and payment (in .jpg format) to us via email address: icprml@163.com. (Before May 05, 2025) (Very important)

*Address all reviewers' comments and revise your paper accordingly before submitting the final camera-ready version of your manuscript. Please make sure that all Figures and Tables are of high quality and their content is easily readable. Once submitted, no revisions will be accepted.

*Conference secretary will contact you later about the copyright, please pay attention to your e-mail inbox.

NOTE: Cancellation and Refund Policy:

All cancellations must be sent in writing to the conference email box: **icprml@163.com** and are subject to the following cancellation charges:

* Conference registration cancelled before March 10, 2025 with reasonable grounds: 100% of the payment will be refunded, however an administration fee of 30USD will be charged for the service.

* Conference registration cancelled before April 10, 2025 with reasonable grounds: 50% of the payment will be refunded.

* Conference registration cancelled before April 30, 2025 with reasonable grounds: 30% of the payment will be refunded.

* Conference registration cancelled after April 30, 2025 with reasonable grounds: the payment will not be refunded.

** 1. The above item is based on deduction for any bank transfer fees (30USD). 2. In case of visa rejection, no-show or early departure, no refund of registration fees will be made. 3. All refunds will be processed after the conference. 4. Due to force majeure including but not limited to earthquake, natural disaster, war and country policy, organizer reserves the rights to change the conference dates or venue with immediate effect and does not assume responsibility for any additional costs, charges, or expenses; to include, charges made for travel and lodging.

After the registration, we will send all qualified papers to the publish house and index organization for publishing directly.

Lastly, please strictly adhere to the format specified in the conference template while preparing your final paper. If you have any problem, please feel free to contact us via **icprml@163.com**. For the most updated information on the conference, please check the conference website at **http://www.prml.org/**. The Conference Program will be available at the website in mid-June, 2025.

Again, congratulations. We are looking forward to seeing you in Chongqing, China.

Yours sincerely,

PRML 2025
www.prml.org
icprml@163.com

International Conference on Pattern Recognition and Machine Learning
PRML

- 2 -